Understanding and Supporting Student Problem Solving in Mathematics Exams with Artificial Intelligence

Věra Ferdiánová and Petra Konečná
Department of Mathematics, Faculty of Science, University of Ostrava, Ostrava, Czech Republic

Vera.Ferdianova@osu.cz
Petra.Konecna@osu.cz

Abstract: This paper presents the findings of a pilot study aimed at gaining deeper insights into student errors in solving mathematics tasks from the Czech national school-leaving examination (*maturita*), while also exploring the potential of artificial intelligence (AI) to support error analysis and provide targeted feedback. The study began with an analysis of publicly available CERMAT data, focusing on tasks that have consistently shown low success rates over the years. Based on this analysis, a subset of tasks was selected and further tested on students preparing for the exam. The results were compared with national statistics to validate the relevance of the identified difficulties. A revised version of the test was then developed and administered to a new cohort of students, enabling the collection of a dataset of real student solutions for qualitative error analysis. The study adopted a nuanced framework for error classification, distinguishing between "slips" (minor, often procedural errors) and "true errors" stemming from a lack of conceptual understanding. Emphasis was placed on understanding the nature and origin of these errors, their recurrence, and implications for learning. Student work was analysed in all phases of the error-handling process, including detection, diagnosis, explanation, and correction. At the same time, the study evaluated selected AI tools, primarily ChatGPT 4.0—for their potential to solve exam-level mathematics tasks at the university level and identify errors in student solutions. Multiple test items were processed through the AI system, and its responses were compared with those of students. Particular attention was given to the AI's behaviour when confronted with incorrect or incomplete answers. The results revealed both the promise and limitations of current AI models in supporting formative assessment, particularly with respect to misinterpretation of task wording, difficulty in recognising alternative valid strategies, and occasional inconsistency in the quality of feedback. The findings contribute to the broader discussion on how AI can be effectively integrated into educational practice—not as a replacement for teacher judgment, but as a supplementary tool to enhance student understanding, develop metacognitive skills, and improve preparation for high-stakes assessments such as the *maturita* exam.

Keywords: Mathematics Education, National high-stakes assessment, Error Analysis, Student Misconceptions, Artificial Intelligence in Education, Diagnostic Feedback, Conceptual Understanding, Secondary Education, Test Validation, Educational Technology

1. Background

1.1 Errors in Mathematics and Their Classification

An error is generally understood, both in pedagogical and psychological contexts, as a deviation from the desired or expected performance (Průcha et al. 2009). In mathematics education, an error can be perceived not only as a failure but also as an opportunity for learning—provided it is approached with appropriate didactic intent (Borasi 1994, Brodie 2014).

The academic literature offers various approaches to classifying errors. They can be categorized according to the following:

- Cognitive value: meaningful vs. meaningless (i.e., based on some form of reasoning vs. random)
- Organisation: regular (systematically recurring) vs. sporadic
- Bearer: individual vs. collective (e.g., an entire class repeating the same mistake)
- Typicality: common vs. unusual
- Severity: significant (indicating serious misunderstanding) vs. insignificant
- Nature of deficiency: declarative (knowledge-related) vs. procedural (process-related)
- Cause: e.g., inattention, poor strategy, lack of understanding, etc. (see, e.g., Kulič 1971, Průcha et al. 2009, Heinze 2005)

From a didactic perspective, we are particularly interested in classifications that relate errors to mathematical thinking and the level of understanding demonstrated by the student in a given situation. Here, we draw inspiration from Reason's distinction between slips and mistakes, or more broadly, errors (see Pilous 2014).

Slips are errors that result from inattention, computational mistakes, or carelessness. They are usually insignificant in terms of mathematical thinking and understanding.

Mistakes, by contrast, indicate deeper issues—such as a lack of understanding of concepts, procedures, or problem-solving strategies. These errors tend to recur in similar situations and require a targeted intervention. They may be viewed as obstacles, in the sense described by the theory of didactic situations (Vondrová 2019).

This basic two-level classification allows us to analyse student errors not only in terms of their occurrence, but more importantly, in terms of their significance for the learning process.

1.2 Constructive Work with Errors

In mathematics education, errors have traditionally been perceived as failures, deviations from the norm, or signs of ignorance or misunderstanding. However, this view has been superseded by a modern constructivist approach, which regards errors as a natural part of the learning process and an opportunity for deeper understanding (Průcha et al. 2009, Brodie 2014). When approached with appropriate pedagogical guidance, working with errors can lead to genuine insight, encourage student self-reflection, and support the development of mathematical thinking. The key is to create a safe learning environment—one that does not penalise errors but instead deliberately uses them as a didactic tool.

Working with errors involves four fundamental steps: detection, identification, interpretation, and correction (Průcha et al. 2009). These phases can become part of everyday teaching if the teacher actively integrates them into instructional design. Importantly, the teacher does not need to be the sole actor in this process. Students should be guided to recognise, analyse, and correct their own errors. In doing so, they learn to take responsibility for their learning and develop metacognitive skills.

Research in cognitive science confirms that working with errors is one of the most effective strategies for fostering deep learning. Booth et al. (2017) summarise findings from cognitive psychology demonstrating that analysing incorrect examples can be more effective in acquiring mathematical concepts than working solely with correct ones. When students identify the nature of the error, its cause, and how it can be corrected, they engage more actively in the learning process and develop stronger mental models.

Similar conclusions are drawn in Rushton’s (2018) study, which found that working with mathematical errors improves long-term knowledge retention and supports metacognitive development. The findings show that students who learn through error analysis not only understand the content better but also feel more involved in the learning process and become more aware of their own thinking. According to the author, this experience is more valuable than simply acquiring the correct procedure.

Jo Boaler (2016) has also made a significant contribution to this topic with her concept of mathematical mindsets, emphasising that an error is not a threat, but a signal that the brain is learning. Boaler stresses the importance of fostering a growth mindset in students—that is, the belief that mathematical ability is not fixed or innate, but can be developed through effort, feedback, and reflection. A classroom environment that allows for experimentation, tolerates errors, and provides constructive feedback supports not only improved performance but also a more positive attitude toward mathematics. Boaler warns that a traditional emphasis on speed and accuracy can cause anxiety and a loss of confidence, whereas a learning environment that embraces errors as part of learning promotes creativity and perseverance.

A recent Czech study (Janíková, Janík & Pavelková 2024) confirms that most teachers still perceive errors primarily as indicators of insufficient knowledge rather than as learning opportunities. However, the study also shows that teachers who view errors as part of the learning process are more likely to create supportive classroom environments and contribute to deeper understanding. Their instruction is characterised by openness to errors and encouragement of student participation in analysing them.

Constructive work with errors is thus not merely a teaching strategy, but part of a broader learning culture. It requires an approach that respects individual learning pace, fosters self-reflection, creates opportunities to discuss thought processes, and provides meaningful feedback. The studies cited above make it clear that such a culture has the potential not only to improve students’ academic performance, but also to strengthen their relationship with mathematics as a subject.

1.3 Artificial Intelligence in Education: Adaptive Systems, Personalisation, and the Role of Language Models

The use of artificial intelligence (AI) in education has evolved rapidly in recent years, transforming how students engage with learning content and how educators approach instruction and assessment. Particularly important in this context are AI-powered adaptive learning systems, generative language models, and interactive feedback tools, which together form the foundation for personalised and self-regulated learning.

In a meta-analysis of 45 independent studies published between 2010 and 2022, Wang et al. (2024) demonstrated that AI-enabled adaptive learning systems have a medium to large positive effect on students' cognitive learning outcomes ($g = 0.70$). The effectiveness of these systems was significantly moderated by factors such as the duration of the intervention, student level, subject area, and study design. Interestingly, the type of AI used in the adaptive engine did not significantly influence learning outcomes. Instead, the most impactful elements were the extent to which the systems adapted to students' cognitive, affective, and behavioural characteristics, and the targeting of adaptive features, such as navigation and assessment.

Building on this line of inquiry, Yaseen et al. (2025) examined the impact of adaptive learning technologies and AI tools that provide personalised feedback on student engagement. Their findings suggest that such technologies improve student motivation and involvement, particularly among students with higher levels of digital literacy. The study emphasises the importance not only of the technological tools themselves, but also of students' ability to use them effectively.

A more specific example of AI integration in education is the use of generative language models, such as ChatGPT, which enable interactive dialogue, task creation, lesson planning, and formative assessment. In a scoping review, Pepin, Buchholtz, and Salinas-Hernández (2025) concluded that ChatGPT holds significant potential to support personalisation, collaboration, and the development of instructional design skills among teachers. However, the review also highlights concerns regarding the accuracy of responses, the possible erosion of critical thinking, and the need for pedagogical guidance when implementing such tools in practice.

This ambivalence is echoed in experimental research by Kumar et al. (2023), who investigated the effect of LLM-generated explanations on learning in high school mathematics. Their findings suggest that LLM explanations can facilitate learning—particularly when students attempt to solve problems independently before consulting the AI. Even when the explanations contained errors, they still yielded learning gains, indicating that interaction with incorrect outputs can be cognitively stimulating, especially when paired with appropriate instructional support.

1.4 The School-Leaving Examination in Mathematics in the Czech Republic

The school-leaving examination (*maturita*) in the Czech Republic represents a standardised method for completing upper secondary education. It consists of two parts: a common (state-administered) component and a profile (school-based) component. The common part is organised by the Ministry of Education, Youth and Sports through the Centre for the Evaluation of Educational Achievement (CERMAT), while the profile part is prepared and assessed by the individual secondary schools.

The examination in Czech language and literature is mandatory for all students. In addition, students choose one elective subject—either mathematics or a foreign language. Thus, in the current model, mathematics is an optional subject. However, for certain groups of students (e.g., those in technically oriented study programmes), mathematics may be made compulsory within the profile component, based on the decision of the school.

The didactic test in mathematics, which constitutes the state component of the examination, follows a standardised structure and assessment criteria. The test includes a combination of closed-ended and open-ended tasks. In closed-ended tasks, students select the correct answer from multiple choices, whereas in open-ended tasks, they are typically required to provide not only the final answer but also the full solution process. This format enables verification not only of students' ability to identify correct answers but also of their conceptual understanding, reasoning skills, and the ability to apply acquired knowledge in unfamiliar contexts. The test duration is 135 minutes, and the content is based on the publicly available Catalogue of Requirements for the Mathematics Examination (CERMAT, 2023a).

The mathematics exam is taken by final-year secondary school students who have chosen this subject. In addition to the standard spring term, former students who did not pass in previous years—or who wish to improve their scores—may also take the test. The test is centrally evaluated through a system managed by CERMAT. Each item has a predetermined point value. Students receive their results in the form of both raw scores and percentage evaluations. To pass the exam, students must reach a minimum threshold set by the Ministry of Education.

To support exam preparation, the Catalogue of Requirements for the Mathematics Examination outlines the scope of knowledge and skills to be assessed. It also includes sample tasks with solutions, helping both teachers and students to understand the test structure and expected performance. Furthermore, full versions of previous years' tests—including answer keys and scoring schemes—are published annually on the official website: https://maturita.cermat.cz.

For analytical and research purposes, summary reports of school-leaving examination results are also available. These documents, published by CERMAT, present statistics on student success rates, failure rates, average scores, and other indicators. Results are provided at both national and regional levels and further broken down by school and examination type. As such, they provide valuable data for monitoring educational trends, identifying problematic areas, and improving exam preparation at individual schools (CERMAT, 2023b).

2. Design of Our Study

When designing the test and selecting the tasks, we relied on publicly available data published by CERMAT. Specifically, we selected tasks from the spring 2024[1] school-leaving mathematics examination. The tasks were ranked in ascending order based on their success rates—that is, the average score achieved per task as a percentage of the maximum possible score. We then selected the 14 tasks with the lowest success rates at the national level.

Performance (or lack thereof) on these tasks corresponded to the results of the autumn tests conducted at School 1, the institution where the study was carried out. To this initial set, we added three additional tasks. Although tasks in the areas of statistics and probability showed high success rates in the national data (90.2% and 75.3%, respectively), students at School 1 performed significantly below average on them in the initial autumn test (27.3% and 0.0%, respectively). We also included a task on trigonometric functions. The test used in the autumn session was based on the 2022 version, which included a similar task with an extremely low success rate (0.0%). Since no comparable item appeared in the spring 2024 test, we included an analogous one from the spring 2022 test.

2.1 Analysis of Results

For each task, student responses were categorised according to the following status codes:

U = Task solved correctly, full marks awarded.

N = Task not attempted.

R0 = Task attempted or started, but the solution was incorrect or incomplete, no points were awarded.

R1 = Task attempted; a mistake was made, but partial credit was earned.

Figure 1 presents an overview of the task-solving outcomes in the observed schools, including achieved scores and comparisons with national performance data.

		Task 1	Task2	Task3	Task4	Task5	Task6	Task7.1	Task7.2	Task8	Task9.1	Task9.2	Task13	Task14	Task18	Task20	Task22	Task24	Task25	Task22*
School 1	Score (%)	83,3	79,2	70,8	50,0	41,7	27,1	54,2	41,7	29,2	58,3	66,7	83,3	66,7	87,5	70,8	62,5	66,7	62,5	29,2
	U	20	19	17	11	10	5	13	9	8	14	15	20	15	21	17	15	16	15	7
	N	1	0	2	1	0	6	3	5	7	4	5	0	0	0	0	1	0	0	1
	R0	3	5	5	12	11	11	8	10	9	6	4	4	7	3	7	8	8	9	16
	R1	0	0	0	0	3	2	0	0	0	0	0	0	2	0	0	0	0	0	0
School 2	Score (%)	25,0	54,2	41,7	20,8	33,3	22,9	20,8	20,8	33,3	20,8	45,8	45,8	29,2	45,8	25,0	25,0	37,5	27,1	16,7
	U	6	10	9	4	6	4	5	4	8	5	10	12	5	11	6	6	8	6	4
	N	2	0	5	1	0	4	4	9	7	5	5	2	4	2	5	4	3	4	5
	R0	9	1	3	9	7	6	8	4	2	7	2	3	4	4	6	7	6	6	8
	R1	0	6	0	3	4	3	0	0	0	0	0	0	4	0	0	0	0	1	0
National score		59,0	68,9	60,9	47,0	58,8	62,8	44,9		34,8	59,3		90,2	64,7	75,3	70,4	59,8	64,8	64,0	not available

Figure 1: Summary of the task response status in the observed schools

[1] Due to a reduced time allocation for the autumn testing (90 minutes, i.e., approximately 66.6% of the standard time limit), we decided to include 17 out of the 25 total tasks, with a combined maximum score of 32 points out of a possible 50.

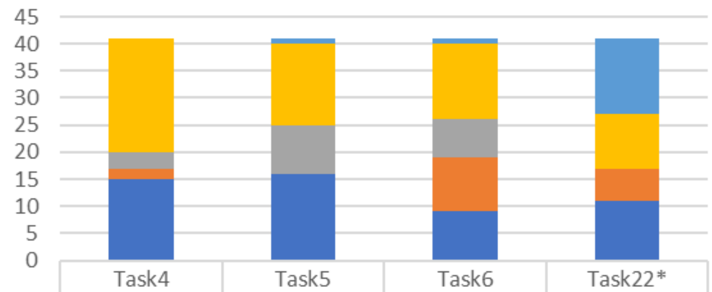
Based on the data analysis, four tasks were selected using the following two criteria:

1. Tasks with low average scores in both schools.
2. Tasks with a high proportion of attempted but incorrect solutions (i.e., categories R0 and R1).

Tasks 4, 5, 6, and Task 22* (on trigonometric functions) were selected for an in-depth analysis of student problem-solving strategies. The errors were classified into two categories:

Category 1: Slips: numerical mistakes, large leaps in reasoning, inattentiveness, or mistakes due to incorrect transcription.

Category 2: True errors, stemming from conceptual misunderstanding, a lack of grasp of fundamental ideas, processes, or an inability to apply or integrate knowledge into a coherent solution.



	Task4	Task5	Task6	Task22*
Unable to determine	0	1	1	14
Category 2	21	15	14	10
Category 1	3	9	7	0
Unsolved	2	0	10	6
Right Solution	15	16	9	11

Right Solution Unsolved Category 1 Category 2 Unable to determine

Figure 2: Distribution of Student Solutions by Error Categories

Figure 2 illustrates the distribution of student errors across the defined categories. Overall, a greater number of true errors were observed. These most often involved incorrect procedures resulting from a lack of understanding of fundamental mathematical principles (e.g., solving an equation instead of an inequality, or incorrectly determining a common denominator for rational expressions). In many cases, students were unable to complete the task due to such conceptual gaps.

In some cases, such as Task 6, students failed to solve the problem because of an inappropriate strategy. For example, some did not identify the least common denominator of rational expressions. Even when the remainder of the procedure was correct, this led them to work with unnecessarily complex expressions involving higher powers and large numbers they could not manage.

In some cases, it was not possible to clearly interpret the nature of the error. In one response to Task 5, it was unclear whether the errors resulted from a series of numerical errors and carelessness, or from misunderstanding of basic arithmetic rules. In Task 6, a student correctly followed each procedural step but failed to complete the final simplification. Because the overall approach was sound, it was uncertain whether this omission was due to oversight or a lack of knowledge.

The final task analysed (Task 22*) contained a higher number of similarly ambiguous cases. In this task, students were asked to determine the value of a function, with the condition that the solution must be within a specific interval. The task was a multiple-choice question, with the answer options including tangent values from both intervals. Based on students' responses alone, it was not clear whether the error stemmed from inattentive reading of the question—corresponding to category 1—or from a failure to understand how to evaluate the function within a specified interval, which would indicate a lack of understanding of the basic principles and rules related to periodic functions—i.e., category 2.

3. Working with AI

This section explores the potential of artificial intelligence (AI) tools in helping students solve test problems, critically assessing both the benefits and limitations of such use. The investigation was divided into two parts, reflecting two ways in which students might engage with AI:

1. To solve an entire test, including the presentation of solution procedures.
2. To identify errors in their own solutions to test problems.

3.1 Solving the Test Using AI

For this analysis, the selected test was uploaded to the ChatGPT environment (version 4.0), and the model was asked to solve the individual tasks. Because the tasks were presented in an image-based format, the user was informed at the outset that optical character recognition (OCR) would be required due to the model's reliance on converting visual inputs into readable text before processing them.

The first three tasks were correctly recognised and solved without error. However, with the fourth task—focused on inequalities—the model misinterpreted the task prompt, resulting in an incorrect solution method. In five subsequent tasks, the model generated multiple solution approaches and invited the user to choose their preferred method. This reflects the way AI learns from patterns in human understanding and provides various interpretative frameworks.

During the interaction, the model twice prompted the user to confirm whether it should continue solving the test—likely an optimisation measure related to token usage, which regulates the computational resources allocated per response.

Out of the seventeen tasks, eight were incorrectly recognised. In two additional cases, the model slightly simplified the task wording, although the overall meaning was preserved. Furthermore, the model generated two tasks that were not part of the original test. When asked about their origin, the AI claimed they were drawn from official CERMAT sources. After being informed that these tasks did not appear in the actual test, the model acknowledged that they had been "randomly included" based on similar test formats.

In conclusion, while AI demonstrated a capacity to accurately recognise and solve certain types of tasks, the overall results were inconsistent and cannot be considered a reliable solution model for students without further oversight.

3.2 Detecting Errors in Student Solutions

The second part focused on AI's potential to identify errors in student solutions. Again, this analysis concentrated on four selected tasks that had proven most challenging, providing a sufficiently rich sample of incorrect student responses for review.

3.2.1 Case 1 – Task 4

A handwritten, incorrect solution to Task 4 (including the task prompt) was photographed using a mobile phone. The image was then uploaded to the AI system, and the model was asked to evaluate the solution.

The AI correctly identified the error: the student had treated an inequality as if it were an equation. The model highlighted the mistake, provided commentary, and presented a revised solution. However, its final result was incorrect. We then uploaded a second image showing a correct student solution and asked the model to compare it with its previous output. Upon detailed review, the AI recognised its own mistake, acknowledged it, and then provided a corrected final result.

3.2.2 Case 2 – Task 5

Task 5 was also submitted as a photo, showing a student's incorrect attempt at solving an equation. This time, the AI failed to accurately recognise the problem statement. We then reformulated the task in a linear textual format—similar to how a student might enter the problem into an advanced calculator.

This enabled the model to correctly identify and solve the task. However, it did not relate its solution to the student's original work. When prompted with a follow-up question whether its own calculation was correct, the model recalculated the problem and confirmed its accuracy. Only after explicitly pointing out that the original solution appeared in the first image did the AI respond appropriately. It identified the error, and—despite noting the low legibility of the work (see Figure 3)—was able to extract the equation used by the student. However, this equation differed from the one the model considered correct.

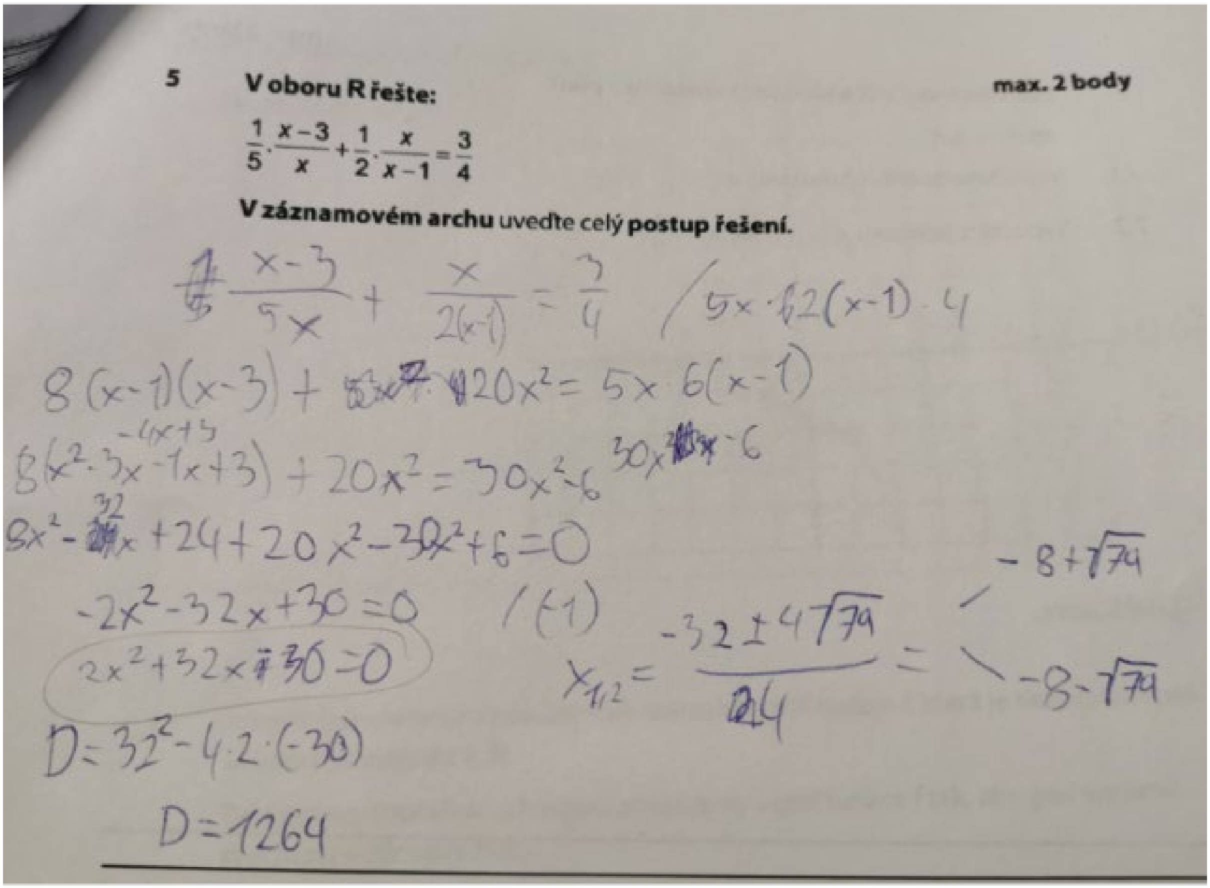


Figure 3: Student's solutions with error

This case highlights two key issues: the extended interaction required to evaluate student errors, and the potential for the AI to identify a different—though mathematically valid—intermediate equation, which it may incorrectly flag as erroneous. This shows the need for thorough verification of AI-generated outputs and careful pedagogical oversight, particularly when evaluating non-standard solution strategies.

3.2.3 Case 3 – Task 6

Task 6 involved simplification of an algebraic expression. Both the task prompt and the student's solution were uploaded as a photo taken with a mobile phone. Initially, the AI failed to correctly interpret the task from the image. To resolve this, we only uploaded the problem prompt, without the solution. At this point, the model accurately recognised the task and proceeded to solve it.

In this instance, the model remained focused on the original instruction to find an error. While simplifying the expression, it discovered a mistake in one particular step, based on a mismatch with its own intermediate result. Ironically, the student's solution was entirely correct—it was simply incomplete, with the final step (factoring out an expression that would lead to cancellation) omitted.

This case also indirectly supports the hypothesis outlined in Case 2: While AI tools are effective at identifying computational errors, their evaluation is guided by their own solution paths. As a result, they may fail to recognise alternative (yet valid) approaches or incomplete solutions. Once again, this highlights the importance of expert interpretation and consistent pedagogical oversight when integrating AI tools into education.

*3.2.4 Case 4 – Task 22**

In the final case, we analysed Task 22*, in which students were required to solve a trigonometric equation and determine the correct value of the tangent function for a given angle, constrained to a specific interval. The task prompt and the student's incorrect solution were uploaded via a mobile phone photo. The AI correctly solved the problem, identified the student's mistake, and provided a clear explanation of the error.

In this case, the AI fully met expectations—offering precisely the kind of support that would be highly beneficial to both students and educators.

4. Discussion

The analysis of student strategies in the selected tasks clearly indicates a high prevalence of true errors when working with equations and algebraic expressions. These errors stem from a lack of knowledge or insufficient mastery of fundamental principles that students are expected to know by the time they take the school-leaving (*maturita*) exam. Therefore, it is essential that the preparation for the exam includes thorough assessment of students' knowledge and skills in this area. If significant gaps are identified, emphasis should be placed on re-teaching and reinforcing foundational skills—e.g., by including simpler tasks that guide students through each step required for solving complex exam problems.

When working with trigonometric functions, we recommend including more exercises that require students to determine values across various intervals. This approach serves a dual purpose: it helps reinforce the importance of paying close attention to interval constraints specified in the task and provides students with valuable practice in interpreting and applying periodic properties of trigonometric functions. Such targeted practice supports effective remediation whether student errors stem from carelessness or a lack of understanding.

The case analyses described in Chapter 3.2 reveal several key findings that should be considered when integrating artificial intelligence (AI) tools into classroom instruction and assessment.

One of the most significant limitations observed is the inconsistent accuracy with which AI models recognise problem prompts—particularly when tasks are uploaded as images or when mathematical notation is generated using an equation editor in DOCX format. This may stem from limited exposure to such formats during training as the equation editor language is less familiar and less legible to the model. Consequently, misinterpretation or omission of task components can occur, leading to incorrect solutions or a failure to understand the task. In contrast, when the same files were converted to PDF format, the AI's recognition accuracy improved notably. This is likely because PDF formats render equations as fixed graphical elements, which the AI is better equipped to process using optical character recognition (OCR).

A second limitation lies in the AI's method of evaluating student responses based on internal solution algorithms. The model typically validates solutions by checking whether the intermediate steps and final answer align with its own procedural pathway. As a result, student responses that use alternative—but mathematically valid—methods may be incorrectly flagged as incorrect. A similar issue arises with incomplete solutions that are headed in the right direction but, for instance, omit a final step such as factoring or simplification.

Finally, a recurring usability issue was observed: interruptions in the flow of problem-solving, as the model frequently requested confirmation to proceed or clarification of instructions. While this behaviour is related to computational efficiency (i.e., token management), it can be distracting for users and impede the continuity of the task-solving process.

5. Conclusion

Based on these observations, several practical recommendations can be made. When submitting digital tasks, it is advisable to use PDF rather than DOCX format—particularly for mathematical content. Task prompts and student solutions should be uploaded separately, presented clearly, and provided in machine-readable form whenever possible.

AI-generated outputs should be considered a starting point for further discussion, rather than definitive evaluations. All assessments must remain under the supervision of a teacher, who can determine whether deviations from standard approaches represent alternative, but still valid, solution strategies.

AI can serve as a valuable assistant, especially for routine problem types with well-defined structures, such as algebraic manipulations, equations, or inequalities.

Ethical Statement

The submitted manuscript is original, has not been published elsewhere, and is not under review by another journal or conference. All authors contributed significantly and all sources are properly cited. The research is based solely on anonymous or publicly available data. In accordance with Czech legislation and institutional guidelines, no ethical approval was required.

AI Statement

This paper examines and analyses the role of artificial intelligence in education. Accordingly, AI tools (specifically ChatGPT 4.0 by OpenAI) were used during the research and preparation of this manuscript. These tools supported selected analytical tasks, data interpretation, and language refinement. All AI use was conducted under the supervision of the authors, and the final content reflects their independent critical judgment. The authors take full responsibility for the accuracy, originality, and integrity of the manuscript.

References

Boaler, J. (2016) Mathematical Mindsets: Unleashing Students' Potential through Creative Math, Inspiring Messages, and Innovative Teaching, Jossey-Bass, San Francisco, CA.

Booth, J.L., Lange, K.E., Koedinger, K.R., & Newton, K.J. (2017) Using Cognitive Science Principles to Enhance Educational Practice: Lessons from Mathematics. ERIC ED586306

Borasi, R. (1994) “Capitalizing on errors as “Springboards for Inquiry”: A teaching experiment”, Journal for Research in Mathematics Education, Vol 25, No. 2, pp 166–208.

Brodie, K. (2014) “Learning about learner errors in professional learning communities”, Educational Studies in Mathematics, Vol 85, No. 2, pp 221–239.

CERMAT. (2023a) “Katalog požadavků zkoušky společné části maturitní zkoušky z matematiky”, [online], CERMAT, https://maturita.cermat.cz/menu/maturitni-zkouska/katalogy-pozadavku.

CERMAT. (2023b). “Výsledky maturitní zkoušky – souhrnné statistiky”, [online], CERMAT, https://maturita.cermat.cz/menu/maturitni-zkouska/vysledky-maturitni-zkousky

Grassinger, R. et al. (2018) “The relevance of adaptive reactions to errors and a positive error climate for academic achievement”, High Ability Studies, Vol 29, No. 1, pp 37–49.

Heinze, A. (2005) “Mistake-handling activities in the mathematics classroom”. Paper read at PME 29, Melbourne, Australia, January.

Kulič, V. (1971) Chyba a učení: funkce chybného výkonu v učení a v jeho řízení, SPN, Praha.

Kumar, H., Rothschild, D., Goldstein, D., & Hofman, J. (2023) “Math education with large language models: Peril or promise?”, SSRN Electronic Journal, [online], https://doi.org/10.2139/ssrn.4641653.

Janíková, M., Janík, T., & Pavelková, M. (2024) “Dealing with student mistakes in mathematics at primary school”, Studia Paedagogica, Vol 29, No. 2, pp 66–89.

MŠMT. (2024) “Ukončování vzdělávání maturitní zkouškou”, [online], MŠMT, https://msmt.gov.cz/vzdelavani/stredni-vzdelavani/ukoncovani-vzdelavani-maturitni-zkouskou.

Pepin, B., Buchholtz, N., & Salinas-Hernández, U. (2025) “A scoping survey of ChatGPT in mathematics education”, Digital Experiences in Mathematics Education, Vol 11, pp 9–41.

Pilous, D. (2014) Vybrané pohledy na žákovskou chybu ve výuce matematiky, Disertační práce, PedF UK, Praha.

Průcha, J., Walterová, E. & Mareš, J. (2009) Pedagogický slovník, Portál, Praha.

Rushton, S.J. (2018) “Teaching and Learning Mathematics through Error Analysis”, Fields Mathematics Education Journal, Vol 3, No. 1.

Vondrová, N. (2019) Didaktika matematiky jako nástroj zvládání kritických míst v matematice, PedF UK, Praha.

Wang, X., Huang, R. T., Sommer, M., Pei, B., Shidfar, P., Rehman, M. S., Ritzhaupt, A. D., & Martin, F. (2024) “The efficacy of artificial intelligence-enabled adaptive learning systems from 2010 to 2022 on learner outcomes: A meta-analysis”, Journal of Educational Computing Research, Vol 62, No. 6, pp 1348–1383.

Wuttke, E., & Seifried, J. (2017) Learning from errors: A key competence for vocational education and training. In: Bauer, J. & Prenzel, M. (Eds.), Handbook of competence development.

Yaseen, H., Mohammad, A. S., Ashal, N., Abusaimeh, H., Ali, A., & Sharabati, A.-A. A. (2025) “The impact of adaptive learning technologies, personalized feedback, and interactive AI tools on student engagement: The moderating role of digital literacy”, Sustainability, Vol 17, No. 3, pp. 1133.