

Obsah

Přehledová studie

- Marina Babayeva
Enhancing physics hands-on lab through online educational tools..... 2

Výzkumné studie

- Irena Budínová
Úspěšnost nadaných žáků při řešení matematické slovní úlohy 15
- Alexandr Nikitin, Marie Snětinová
Do various groups involved in physics education appreciate the same aspects of physics demonstrations? – A methodological approach 34
- Irena Smetáčková, Anna Páchová
The development of self-assessment accuracy in mathematical word problems: A study of primary school learners 48

Enhancing physics hands-on lab through online educational tools

 Marina Babayeva¹

¹ Faculty of Mathematics and Physics, Charles University, V Holešovičkách 747/2, 180 00 Praha 8, Czech Republic; marina.babayeva@matfyz.cuni.cz

The integration of digital tools into physics education offers new opportunities to enhance student engagement, support differentiated instruction, and facilitate hands-on learning. Despite the growing availability of such tools, educators often face challenges in selecting and implementing platforms that align with pedagogical goals and classroom realities. This study aims to provide a structured overview of online educational tools relevant to secondary-level physics instruction, focusing on their categorization by functionality and potential classroom use. A preliminary review was conducted to identify and classify tools based on core features such as content delivery, assessment, collaboration, and simulation. A pilot study and two educator workshops were used to illustrate practical integration strategies and gather initial feedback from teachers. The findings highlight the value of platform consolidation, the importance of usability and accessibility, and the need for inclusive materials. The categorization framework and illustrative cases offer practical guidance for teachers designing their own lessons and selecting tools purposefully. Future research will explore AI-based solutions tailored to hands-on physics laboratories.

Key words:
physics education,
educational technology,
digital tools in
education, online
learning platforms,
instructional software.

Received 5/2025

Revised 12/2025

Accepted 12/2025

1 Introduction

The rapid technological development necessitated an abrupt and large-scale transition from traditional face-to-face instruction to fully or partially online education across schools and universities worldwide. This shift significantly impacted instructional strategies, especially in disciplines requiring hands-on engagement, such as physics. A bibliometric study by Jatmiko et al. (2021) identified “online learning” as the second most prominent research trend in physics education, especially during the pandemic period, following by “experiments”, highlighting the central role that digital platforms assumed in response to these unprecedented challenges.

While the immediate need for purely online instruction has diminished since COVID-19, educational practices continue to evolve toward hybrid or blended learning models. These approaches combine the flexibility and accessibility of online education with the interactive and practical benefits of in-person instruction. Studies by Guo et al. (2023) and Xu et al. (2023) have demonstrated that blended learning can enhance student outcomes, especially when effectively integrated with curriculum goals and supported by appropriate technologies. Blended models enable continued access to recorded lectures, diverse digital materials, and asynchronous collaboration while preserving opportunities for hands-on experimentation, face-to-face discussion, and real-time feedback.

2 Theoretical background

2.1 Technologies in physics education

However, the integration of such technologies into physics education presents both pedagogical opportunities and implementation challenges. Instructors must adapt to new instructional roles, navigate steep learning curves associated with unfamiliar tools, and ensure inclusive and equitable access for all learners. Consequently, research interest in educational technology within physics education has expanded significantly in recent years (Kurnia et al., 2022). A bibliometric analysis by Prahani et al. (2022) confirmed a marked increase in the number of publications in this area during 2020–2021, reflecting the field’s growing relevance and dynamism.

The educational technology market is highly diverse, with platforms tailored to different educational levels and instructional goals. Interfaces designed for primary education often prioritize simplicity and visual appeal, while platforms targeting secondary and post-secondary institutions tend to offer more complex functionality, such as advanced assessment tools, collaborative environments, and detailed performance analytics. Some tools are built for specific purposes—such as quiz creation, class management, or instructional content delivery, while others aim to integrate several functionalities within a single ecosystem. In this work, I examine several functional categories of educational software and illustrate their implementation with practical examples from classroom settings.

The increasing pervasiveness of technology in everyday life has encouraged educators to integrate digital tools into their instructional practices. Motivations for doing so include enhancing student engagement, supporting interactive and differentiated instruction, and equipping learners with digital competencies often referred to as 21st-century skills. Aprilo et al. (2023) conducted a comprehensive literature review on the integration of technology in 21st-century physical education, underlining the broader relevance of such technologies across various disciplines, including science education.

Educational technologies can also serve as scaffolds for the development of both cognitive and collaborative skills. For example, Schanze, Groß, and Hundertmark (2020) investigated the use of online tools to support collaborative writing in educational contexts. Their findings suggest that, while individual contributions remain important, technology-enhanced collaboration fosters a stronger sense of group identity and shared goals. Similarly, Salas-Rueda et al. (2022) examined the implementation of collaborative digital walls in physics education and found that such tools contributed positively to the teaching – learning process by encouraging student interaction and active knowledge construction. Collectively, these studies reflect a growing interest in the pedagogical potential of learning management software (LMS) and related technologies across educational contexts. As digital tools continue to evolve, their thoughtful integration into the science classroom offers promising opportunities to enrich both teaching practices and student learning outcomes.

The advantages and disadvantages of educational software in science education at schools and colleges have been extensively discussed in literature. The use of educational technology offers enhanced understanding of the material being studied, inclusive teaching, increased accessibility, and providing immediate feedback (Fonseca et al., 2013). Additionally, the integration of technology in education allows for the facilitation of the learning process, as it can be used both inside and outside the classroom, thereby promoting autoregulation of learning (Cacabelos et al., 2015). Simanullang et al. discuss the application of Moodle LMS in physics education. Moodle was chosen based on their previous research on available open-source LMS as one of the most popular systems. It offers video-based activities, forums, materials, and quizzes. According to Simanullang, students were able to successfully conduct all the activities without any obstructions (Simanullang & Rajagukguk, 2020). Setiawan et al. also talk about the LMS and use iSpring Free for creating engaging and responsive courses. They particularly underline that providing students with online materials increased accessibility of the materials anytime and anywhere given the existing internet connection (Setiawan et al., 2022). Rizal et al. discuss the development of LMS for pre-service teachers to increase their digital literacy. Overall, it is stated that the developed LMS was beneficial for students to stimulate them to develop new skills. One of the limitations was the difficulty in the case of limited/bad internet connection (Rizal et al., 2022).

Another noteworthy example of educational software in physics instruction is presented by Solvang and Haglund (2021), who investigate the use of GeoGebra, a dynamic mathematics software, in physics education. Their study offers a broad range of application scenarios across different physics domains such as mechanics, wave phenomena, and geometrical optics. By compiling and analyzing existing implementations, the authors demonstrate that the integration of GeoGebra can significantly enhance students' conceptual understanding, supporting more interactive and visual forms of learning.

2.2 Challenges and limitations of technology integration

Despite the promising benefits of educational technologies, their implementation is not without challenges. One such issue pertains to asynchronous learning environments, which may limit immediate feedback and student-instructor interaction. Levin (2023) notes that such scenarios necessitate advanced data analysis techniques, such as cluster analysis, to identify and address learning deficiencies effectively. Furthermore, the cognitive demands associated with digital learning environments are a matter of ongoing debate. Skulmowski and Xu (2021) examine factors contributing to cognitive load in digital education, identifying five critical elements: interactivity, immersion, disfluency, realism, and the presence of redundant information. These factors can variably influence learners' cognitive processing and educational outcomes.

The relationship between digital interactivity and learning outcomes remains inconclusive. For instance, Schubertová et al. (2023) report that although digital media offer enhanced interactivity, this does not consistently translate into improved academic performance. Additionally, concerns about students' attention spans have become increasingly prominent in discussions of blended and distance learning (Levitin, 2015; Suzuki, 2015). As students spend more time engaging with digital content, they are exposed to frequent distractions, particularly from social media platforms, which can disrupt sustained focus and diminish memory retention (Newport, 2019). Tripathi (2023) argues that the pervasive use of social media negatively impacts students' ability to concentrate, thereby undermining the learning process.

The integration of technology into education necessitates a thoughtful approach, particularly in addressing potential limitations such as disparities in student engagement and comprehension across syn-

chronous and asynchronous learning environments (Levin, 2023). While technology in science education offers considerable advantages, including improved conceptual understanding, personalized learning pathways, and greater flexibility in instructional delivery, they also introduce notable challenges. These include difficulties in maintaining effective communication in asynchronous contexts, complexities associated with managing distance learning, and the demand for robust systems to monitor and analyze student performance data. Consequently, successful implementation of educational technologies requires not only technical infrastructure, but also pedagogical strategies tailored to diverse learning needs and environments.

2.3 Research aim and questions

In this paper, I present an overview of online educational platforms relevant to secondary-level physics instruction, with a focus on their categorization by functionality and potential classroom use. The primary audience for this work is physics teachers, who often design their own lessons based on national or regional curricular frameworks, without the support of dedicated course designers. While the paper includes examples from a pilot classroom study and two educator workshops, these are intended to serve as illustrative cases that demonstrate how selected tools can be integrated into hands-on physics lessons. The main goal is to provide educators with a structured map of available digital resources and practical insights to support informed decision-making when selecting and implementing tools in their own teaching contexts.

To guide this exploration, the study is framed around the following research questions:

RQ1: What functional categories can be used to organize online educational tools for physics instruction, and what are the characteristics of tools within each category?

This question explores how digital tools can be systematically grouped based on their core functionalities, such as content delivery, assessment, collaboration, and simulation. It aims to provide educators with a structured overview of available platforms, highlighting their intended use, strengths, and limitations in the context of physics education.

RQ2: How can selected online educational tools be integrated into secondary-level physics lessons to support hands-on learning activities?

This question focuses on practical strategies for incorporating digital tools into classroom instruction. It draws on examples from pilot study and educator workshops to illustrate how tools like PhET, and Formative were used to enhance student engagement, facilitate collaboration, and support conceptual understanding in physics labs.

3 Method

3.1 Initial tools search and categorization

This study constitutes an initial phase of a broader research project investigating the role of digital tools in supporting hands-on physics laboratories. To inform the development of instructional materials, I conducted a preliminary review of freely available educational technologies that support three core functions: access to learning content, opportunities for problem-solving practice, and mechanisms for formative and summative assessment.

The search strategy was designed to reflect the typical behavior of educators seeking digital tools. It included:

- Keyword-based online searches using general-purpose search engines (e.g., Google)
- Review of widely used educational video tutorials
- Analysis of popular thematic blog posts focused on technology in science education

As part of the review, I identified commonly used digital assessment formats relevant to physics instruction, including:

- Multiple-choice questions
- Open-ended questions
- Matching tasks (e.g., text-to-text, text-to-image, image-to-image)
- Embedded questions within videos (e.g., multiple choice, open-ended)
- Drawing or diagram creation with teacher feedback

- Interactive tasks integrated with simulations
- Fill-in-the-blank items (text or numerical)
- Table completion tasks (text or numerical)
- Concept mapping tasks
- Drag-and-drop activities
- Interactive timelines and process flows
- Gamified challenges
- Peer review, collaborative assessment

To be considered suitable for use in the classroom, a tool ideally needed to support several of these assessment types. Following the initial search, the identified tools were categorized according to the following dimensions:

- **Functional Dimension:** Categorization was based on the tool's primary pedagogical function (e.g., learning management, assessment, collaboration, simulation).
- **Overlap and Multifunctionality:** Some tools span multiple categories. In such cases, the dominant use case was used for classification, with overlaps noted.

3.2 Pilot study

Based on the initial review and categorization, to test some of the tools and form an opinion of their usage, a pilot study was designed. The pilot took place at Liberty High School in Hillsboro, Oregon (USA) in March 2023 in collaboration with a local physics teacher with more than 20 years of experience teaching physics and math, master's degree in teaching and science education, and advanced math and physics certification. A unit from the existing curriculum was adapted using these tools to create a digital learning based module. The pilot involved two ninth-grade classes, with a total of 50 students aged 14–15. The topic covered was *Waves, Sound, and Sound Propagation*.

For the pilot implementation, five digital tools were selected:

- **Formative.com** and **Miro.com** for the delivery of self-paced learning materials
- **PhET Interactive Simulations** for visualizing physics concepts through interactive simulations
- **Wizer.me** for administering formative assessments
- **Google Classroom** for organizing materials, communicating with students, and managing deadlines.

The selection process was guided by a combination of practical availability, pedagogical diversity, and exploratory intent:

- **Formative.com**, **PhET Interactive Simulations**, and **Google Classroom** were chosen based on their **existing use** by the partnering teacher. This ensured a smoother integration into the classroom and allowed for authentic feedback from a practitioner already familiar with these platforms.
- **Miro.com** and **Wizer.me** were added to introduce **functional variety** and test the feasibility of integrating less commonly used tools. These platforms were selected to explore collaborative whiteboarding and interactive worksheet creation, respectively.

While the selection was not based on a formal evaluation framework, it was informed by the following practical criteria:

- **Usability:** Tools needed to be intuitive enough for both students and teachers to use with minimal training.
- **Cost-effectiveness:** All selected tools offered free versions or educational licenses suitable for classroom use.
- **Functionality:** Each tool addressed a distinct instructional need, content delivery, simulation, assessment, or classroom management.

This mixed approach allowed the pilot to reflect real-world conditions, where tool adoption is often shaped by teacher familiarity, institutional constraints, and the need to balance innovation with feasibility.

During the first session, students were introduced to the topic using Miro boards to explore fundamental concepts. This was followed by two demonstrations: a hands-on activity with a slinky to illustrate mechanical wave propagation, and an interactive simulation using the “Wave on a String” tool from PhET. Students then engaged in individual and group activities using laptops. Formative.com hosted a series of digital tasks incorporating videos, simulations, and short exercises. Group work included

hands-on investigations, such as examining how the thickness of a rubber band or guitar string influences vibration frequency. At the end of each session, students completed an assessment on Wizer.me. All resources, activities, and communications were managed via Google Classroom to ensure centralized access. The focus of the study was on teacher's and students' experience rather than on quantitative learning outcomes. To evaluate the usability of digital tools, feedback was gathered through a teacher interview consisting of two parts: an unstructured discussion and a written survey. The survey included questions about participants' background (e.g., education level, learners' age group, prior experience with classroom technologies) and tool usage, such as "Which of the following tools have you used in your teaching?" with a list of software options. It also asked about classroom dynamics after the pilot, for example, "Have you noticed any changes in student engagement, such as increased interest, frustration, or activity?" This feedback provided initial insights into the effectiveness and accessibility of the selected platforms in a real classroom environment.

3.3 Workshops

Following the pilot study, it was decided to introduce the research to physics teachers in the form of a workshop. Based on the classroom experience, the implementation strategy was changed and streamlined by consolidating multiple tools into a single platform. Nearpod.com was selected as the primary environment for content delivery, interaction, and assessment due to its versatility and ease of integration.

To disseminate the updated curriculum and gather broader feedback, the materials were presented in teacher workshops at two international physics education conferences: GIREP 2023 (Groupe International de Recherche sur l'Enseignement de la Physique) and MPTL 2023 (Multimedia in Physics Teaching and Learning). These sessions aimed to showcase the instructional design and technological tools while exploring their applicability in diverse educational settings.

The first workshop was attended by approximately 20 participants, and the second by 6 participants. Attendees primarily included physics educators and education researchers from different European countries. The participants were diverse in their experience level spanning from 3 to 22 years of teaching as well as the learners' age group: from middle school to bachelor level. Most of the participants were primarily physics teachers with some also teaching math and science.

During the workshop, participants experienced a simulated physics lesson, the same as in earlier classroom implementation, using digital tools from both student and teacher perspectives. Activities included real-time interaction with embedded simulations and questions, as well as an overview of the teacher interface, which allows instructors to monitor student progress, respond to queries, and provide immediate feedback.

At the end of each session, participants shared their impressions and completed an optional online survey. The survey included background questions and asked about their overall experience with the tools, for example: "What did you find particularly beneficial in using the offered tools?" Four completed surveys were submitted, offering valuable insights into the perceived usability, pedagogical potential, and possible limitations of the tools demonstrated.

4 Results

4.1 Tool categories and examples

Based on the analysis of core functionality and features, the identified digital tools were grouped into the following categories:

1. **Learning Management Systems (LMS)**

Platforms in this category support the delivery and organization of instructional content. They allow educators to create lessons, distribute files or presentations, assign tasks, and share materials with students. Importantly, these resources are accessible outside scheduled class time, enabling flexible, asynchronous learning.

2. **Video Conferencing Tools**

This category includes platforms designed for real-time communication, supporting scheduled virtual meetings and classes with multiple participants.

3. **General Assessment Tools**

Tools in this category support the creation of diverse assessment formats, such as multiple-choice, open-ended questions, and fill-in-the-blank tasks. They typically allow for the integration of multimedia elements (images, video, audio) and the combination of various task types within a single worksheet.

4. **Content Creation Tools**

Platforms that specialize in the development of instructional materials in specific formats (e.g.,

images, videos, digital books, presentations). These tools often offer templates and editing features tailored to particular media types.

5. AI-based Tools

These are platforms or applications that leverage artificial intelligence to assist teaching and learning. They may provide automated grading, personalized learning pathways, content generation, adaptive quizzes, or even AI-driven tutoring systems.

6. Collaboration Tools

Designed to facilitate group work and joint content development, these tools may include features such as shared whiteboards, discussion forums, and real-time commenting. While not always education-specific, they are widely used in instructional contexts to support peer interaction and project-based learning.

7. Media-Specific Assessment Tools

These platforms are assessment-focused but rely predominantly on one or two media types as the primary basis for evaluation, for example, tools designed specifically for video-based quizzes or reading comprehension tasks.

8. Quiz Platforms

This category includes software dedicated to creating and administering online quizzes. Unlike general assessment tools, these platforms are typically used synchronously during class sessions and often support group discussion and immediate feedback.

9. Specialized Simulations & Virtual Labs

These tools focus on the creation or deployment of domain-specific content, such as scientific simulations or interactive models. They often require subject-specific knowledge and technical expertise to use effectively and are particularly valuable in STEM education.

10. Data Collection & Analysis Tools

Platforms designed to support the gathering, processing, and visualization of experimental or observational data. These may include digital lab notebooks, statistical software, or mobile apps for sensor-based measurements.

11. Virtual & Remote Labs

Systems that allow learners to perform experiments in a simulated or remote-controlled environment. These tools replicate the conditions of a physical laboratory, enabling experimentation without requiring on-site lab access.

12. Visualization & Astronomy Tools

Specialized software that helps represent abstract or large-scale scientific concepts, particularly in physics and astronomy. These tools often include star maps, particle simulators, or graphing utilities to make complex data and phenomena more comprehensible.

13. Gamification Platforms

Platforms that integrate game-like elements, such as points, leaderboards, badges, or quests, into the learning process. These tools aim to boost motivation, engagement, and competition among students.

14. Educational Resource Hubs

Aggregators or repositories that provide structured access to instructional content such as lesson plans, open educational resources, simulations, videos, and datasets. They support teachers in discovering and reusing quality materials.

15. Coding & Computational Tools

Platforms that enable students to learn and apply programming, numerical modeling, or algorithmic thinking in scientific contexts. These tools often integrate coding with visualization and data analysis.

16. 3D Modeling & AR/VR Tools

Software that allows the creation or exploration of three-dimensional representations, augmented reality (AR) environments, or fully immersive virtual reality (VR) scenarios. These tools are especially useful for visualizing complex structures and abstract physics concepts.

17. Note-taking & Study Tools

Applications designed to support personal learning management, including digital notetaking, flashcards, annotation, and concept-mapping. They facilitate information organization, retrieval, and review.

Examples of representative tools for each category are summarized in Table 1 including their primary use cases.

Table 1: Online resources examples based on outlined categories

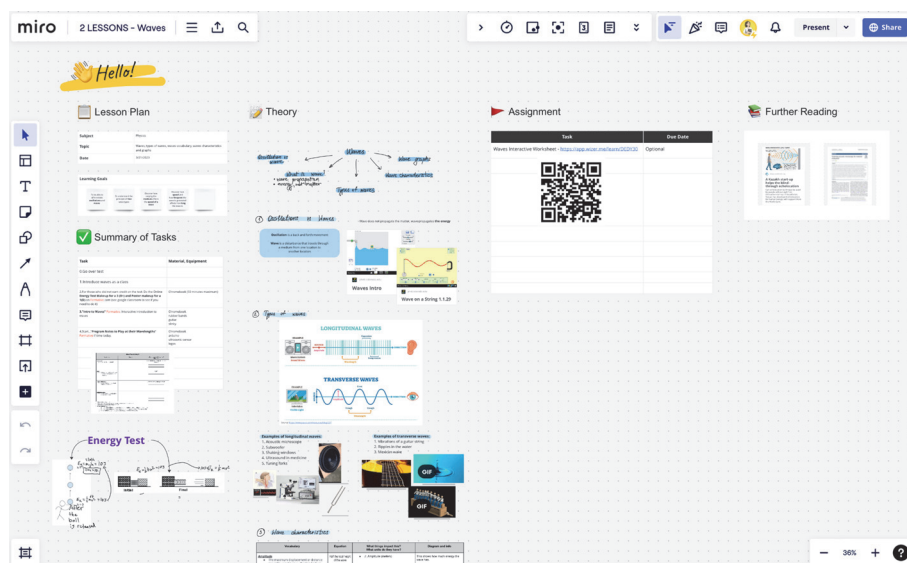
Category	Examples	Use Cases
LMS	Nearpod, Classkick, Kami, Pear Deck, Moodle, Google Classroom, Schoology, Microsoft Teams	Deliver structured materials, lessons, assignments, quizzes, integrate media & track progress.
Video Conferencing Tools	Google Meet, Zoom, BigBlueButton, Webex	Remote/hybrid teaching, breakout rooms, polls, recording lectures, integration with LMS.
General Assessment Tools	Formative, Wizer.me, Socrative, Wayground	Real-time assessments, surveys, performance tracking, feedback dashboards.
Content Creation Tools	Canva Education, Genially, Flip (Flipgrid), Iorad, Prezi, ThingLink, Powtoon, Piktograph	Create interactive presentations, videos, graphics, tutorials, e-books.
AI-based Tools	ChatGPT, Perplexity, Eduaide.AI, Diffit, MagicSchool AI, Quizgecko, Quillionz	Generate lesson plans, adapt materials, create practice problems, automate feedback.
Collaboration Tools	Wakelet, Miro, Padlet, Trello	Group brainstorming, project boards, real-time collaboration, resource sharing.
Media-Specific Assessment Tools	ActivelyLearn, Edpuzzle, PlayPosit, Kami Assignments, InsertLearning, GoReact	Annotate texts, embed quizzes in videos, gamify content, interactive assignments.
Quiz Platforms	LearningApps, Quizlet, Quizalize, Mentimeter, Wordwall, Kahoot, Gimkit, Blookey, Slido, AhaSlides, Quizalize	Gamified learning, live quizzes, self-paced practice, interactive polls.
Specialized Simulations & Virtual Labs	PhET Simulations, GeoGebra, Wolfram Alpha, Algodoo, Phision, myPhysicsLab, OPhysics, ROQED Virtual Lab, Yenka, ExploreLearning Gizmos	Physics visualization, modeling, and interactive experiments for conceptual understanding.
Video & Experiment Tools	Tracker, Pivot Interactives, FizziQ, Vernier Video Physics	Connect real-world motion and experiments to models with analysis tools.
Data Collection & Analysis Tools	Vernier Graphical Analysis, Logger Pro, PASCO Capstone, Phyphox, LabQuest, Physics Toolbox Sensor Suite, SageMath, SparkVue, LabArchives	Collect sensor data, analyze graphs, fit models, support experimental research.
Virtual & Remote Labs	Labster, PraxiLabs, Go-Lab, WhimsyLabs, PNX Physics Virtual Labs, OLabs, Praxilabs XR, ChemCollective	Run realistic experiments online; VR/AR options available.
Visualization & Astronomy Tools	Stellarium, Universe Sandbox, Celestia, NASA's Eyes, Cosmosium	Visualize large-scale or abstract concepts (astronomy, cosmology, circuits).
Gamification Platforms	Classcraft, Minecraft Education, Quizalize, PlayBrighter	Increase motivation via role-play, building, and game-based physics learning.
Educational Resource Hubs	The Physics Classroom, HyperPhysics, Spongelab Interactive, MERLOT Physics, OER Commons, Khan Academy, OpenStax Physics	Tutorials, reference maps, structured lessons, free textbooks and open resources.
Coding & Computational Tools	GlowScript, Trinket, Jupyter Notebooks	Simulation, modeling, computational problem-solving, AR/VR development.
3D Modeling & AR/VR Tools	CoSpaces Edu, Merge Cube, Unity, Blender, OpenSpace3D	Immersive 3D/VR content for mechanics, astronomy, and circuits.
Note-taking & Study Tools	Notion, OneNote, Quizlet (spaced repetition), Evernote, Obsidian, Google Keep, Coggle, MindMeister	Knowledge organization, mind mapping, spaced repetition, collaborative notes.

The list of tools presented in this study is not intended to be exhaustive. Given the continuous emergence of new educational technologies, the selection reflects those platforms considered most relevant at the time of the research.

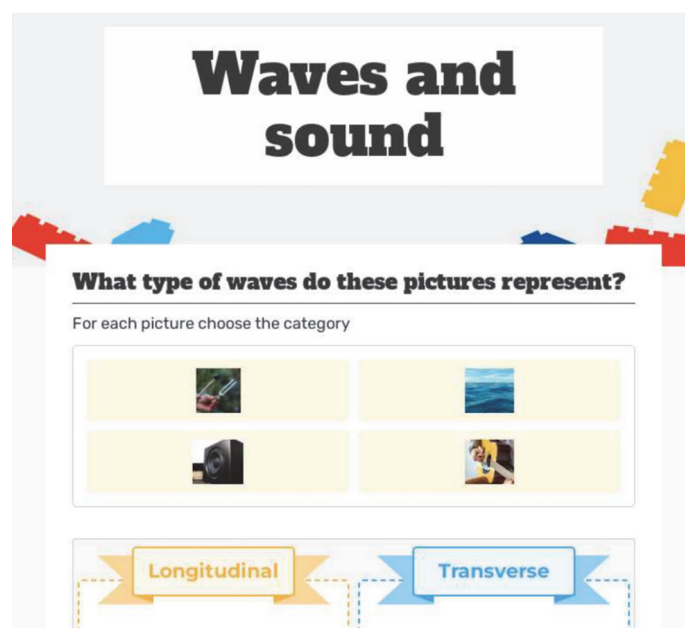
4.2 Tools application

Following the initial research, pilot implementation of five educational tools, **Miro.com**, **Formative.com**, **Wizer.me**, **PhET simulations**, and **Google Classroom**, was conducted in classroom settings. Each tool offered distinct advantages but also presented limitations that informed subsequent adjustments in tool selection.

Miro.com offered a highly collaborative environment with an infinite whiteboard interface suitable for teamwork and brainstorming. Figure 1a shows an illustrative only overview of how the whiteboard might look like. However, teachers reported significant technical difficulties including long loading times, connection instability, and a non-intuitive interface, especially for first-time users without touchscreen devices. For example, after the pilot project, the teacher noted: “I struggled making my own Miro board, but it reinvigorated me to use OneNote. It took too long to load, so I assumed it would do the same for my students.” These usability issues outweighed the collaborative potential in practice.



a)



b)

Fig. 1: a) Miro board created for waves topic, b) preview of Wizer worksheet

Wizer.me enabled the creation of interactive worksheets using visually appealing templates and access to a large library of shared resources. An example of an assessment question is shown in Figure 1b. Yet, the necessity to publish resources publicly in the free version and the limited types of questions and analytics available posed challenges in classroom implementation.

Formative.com stood out for its real-time analytics, diverse question types, and seamless integration with Google Classroom. However, the limited visual customization and styling of worksheets often made it difficult to emphasize key information, resulting in students skipping important content unintentionally. Figure 2 shows an illustrative example question of student's worksheet.

LEGO setup process

Assigned to Guests | Student Paced | 3 Questions

1 2 3 <>

INSTRUCTIONS

1. Connecting LEGO block to the device

1. Follow the steps below to connect your LEGO block to your computer/tablet. Tick the actions you have done to keep track of your actions

☐ Now grab your LEGO block and hit the big button to turn on your block.

☐ Click on the small bluetooth button, it will start blinking blue. Your block should look like this:

☐ Go back to the Spike App on your computer and click Connect button in Spike

2. Make sure everything is functioning

☐ Connection menu will open, choose the block and click Pair

Fig. 2: Formative worksheet overview

PhET simulations were primarily used to demonstrate abstract physics concepts through interactive visualizations. These simulations complemented hands-on activities by providing dynamic models of wave behavior. No major usability issues were reported, as the tool is intuitive and widely recognized in physics education.

Google Classroom, already integrated into the school's ecosystem, served as the central hub for classroom management. It streamlined the distribution of materials, communication, and assignment tracking, ensuring that all resources were accessible in one place. Teachers reported smooth functionality and minimal technical challenges during implementation.

While each of these platforms addressed specific classroom needs, simultaneous use of multiple platforms resulted in a fragmented experience. The switching between tools was found to be confusing for both students and teachers and reduced the overall effectiveness of the digital lesson flow.

In response, subsequent workshops adopted **Nearpod.com** as a unified solution. The illustrative overview of the Nearpod's interface is shown in Figure 3. Nearpod combines presentation features with built-in assessments, real-time feedback, and student-paced modes. Teachers highlighted its accessibility without user registration, its visual presentation style that reduced skipped content, and its capacity to integrate multiple learning activities into one coherent experience. Drawbacks included limited media integration (some content opened in separate windows) and the difficulty of processing post-lesson data in numerical form.

Workshop feedback and teacher interviews supported these findings. Teachers across all levels found collaboration, content sharing, and tool accessibility to be the most valuable aspects. Reported challenges

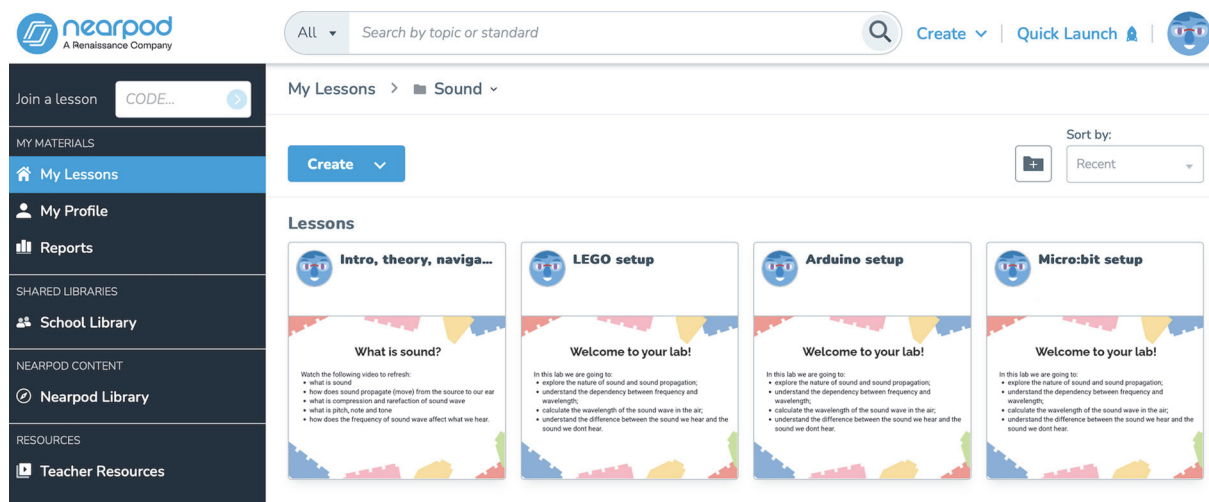


Fig. 3: Nearpod lessons library

included internet connectivity issues, steep learning curves for some tools, and the need for clearer student instructions. Several teachers also highlighted the lack of inclusive materials for learners with special needs and requested more targeted resources and support.

Despite these challenges, nearly all respondents reported a high likelihood of incorporating the tested tools into future teaching. Additionally, several educators expressed interest in receiving further training, both self-paced and instructor-led, on the use of collaborative and digital tools.

5 Discussion

This study proposed a categorization framework based on pedagogical functionality, distinguishing between learning management systems, assessment platforms, collaboration tools, simulation environments, content creation tools, etc., to address the first research question. The analysis revealed that many tools exhibit **multifunctionality**, with overlapping features across categories. For example, Classkick supports both assessment and collaboration, while Nearpod integrates content delivery, assessment, and feedback.

Teacher and workshop participant feedback emphasized the importance of **collaboration**, **access**, and **adaptability**. Tools that supported easy content sharing and peer-to-peer interaction were rated highest, especially in environments where infrastructure (such as stable internet connections or available hardware) could not be guaranteed. The importance of **accessibility** and **inclusivity** also emerged as a recurring theme in teacher feedback. One of the educators highlighted the need for materials that could accommodate learners with physical or sensory impairments as “Using these tools for an inclusive activity”.

The findings from the pilot study and subsequent educator workshops demonstrate the significant potential of digital tools to enhance physics instruction, particularly when thoughtfully integrated into lesson design. However, the implementation also revealed practical challenges that must be addressed for these tools to be effective in real-world classrooms.

A key takeaway from this study is the value of platform consolidation. While tools such as Miro, Wizer, and Formative each supported specific instructional functions, collaboration, formative assessment, and real-time feedback, their simultaneous use often resulted in fragmented lesson structures. This fragmentation led to cognitive overload for students and disrupted instructional flow, particularly in younger learners. These findings echo the concerns raised by Skulmowski and Xu (2021).

In contrast, Nearpod, which was adopted for the workshops, provided a more streamlined and coherent digital experience. Its integrated features, such as interactive slides, embedded assessments, and real-time monitoring, enabled teachers to deliver lessons in a unified format, reducing the need for transitions between platforms. Participants appreciated its visual clarity, intuitive interface, and accessibility, especially the ability to join sessions without student accounts. These characteristics align with recommendations in the literature advocating for tools that support blended and self-paced learning environments (Guo et al., 2023; Xu et al., 2023). Nevertheless, even Nearpod presented limitations. For instance, it has a restricted media integration (e.g., certain content opening in separate windows) and its limited capacity for exporting and analyzing detailed assessment data. These issues underscore a broader challenge in educational technology: balancing usability with feature richness. Tools that are simple and easy to adopt may

lack the flexibility needed for differentiated instruction, while feature-rich tools may require substantial training and support to use effectively.

It is also noteworthy that educators reported changes in classroom dynamics when new tools were introduced. Some observed increased engagement from students who were typically less active, suggesting that digital tools can reach learners who may not respond to traditional teaching methods: *“I had some kids who never do the honors work, but jumped in to do this one”* (honors – optional students’ tasks for additional points). Others highlighted the novelty of hands-on materials as a motivating factor, especially when paired with digital instructions or collaborative tasks.

These insights directly address the second research question: the pilot and workshops illustrated that digital tools could complement physical experiments by scaffolding conceptual understanding, facilitating collaboration, and enabling formative assessment. However, successful integration depends on thoughtful instructional design, teacher familiarity, and infrastructure readiness.

Looking ahead, the evolution of educational technology introduces new opportunities and challenges. One emerging trend is the integration of artificial intelligence (AI) into physics education. While AI tools hold promises for personalized learning, adaptive feedback, and intelligent content delivery, they also raise new concerns around transparency, teacher readiness, and ethical use. Effective AI deployment will depend on robust data analytics and careful interpretation of learner behavior, capacities that most current platforms do not yet support at scale. Building on the insights from this study, the author’s ongoing research focuses on developing tailored AI-based solutions specifically designed to support hands-on physics laboratories, bridging the gap between digital innovation and experimental learning (Kregear et al., 2025).

Ultimately, the findings highlight a need for teacher-centered implementation strategies, where the selection and use of tools are guided by classroom context, infrastructure limitations, and learner diversity. The study reinforces the idea that technology alone does not immediately improve education; it must be aligned with pedagogical goals, adapted to the teaching context, and supported by adequate training and infrastructure. Even well-designed tools require support structures such as clear instructions, technical troubleshooting, and targeted training to be fully effective. Future research should explore long-term implementation across diverse learning environments and evaluate the impact of newer technologies, including AI-driven systems, on student learning outcomes and teacher practices.

6 Limitations

This study was exploratory in nature and subject to several limitations. First, its primary aim was to provide an overview of the current landscape of online educational tools relevant to physics instruction, with a focus on categorization and practical classroom integration. The pilot implementation and educator workshops were conducted not as formal evaluations, but rather as illustrative case studies to verify initial hypotheses and gather preliminary feedback from teachers. The sample of educators was small, which may not fully represent the diversity of teaching contexts, school infrastructures, and student populations. These examples were intended to demonstrate how selected tools might be used in real classroom settings, rather than to produce generalizable findings or comparative effectiveness data.

Second, the duration of the tool implementation was relatively short. Teachers and students may require more extended use to fully adapt to the tools and provide deeper insights into their long-term impact on learning and engagement.

Third, the effectiveness of the tools was influenced by external factors such as internet connectivity, hardware availability, and individual teachers’ prior experience with educational technology. These factors varied across participants and could have skewed perceptions of tool usability and usefulness.

Finally, while feedback was collected from both interviews and written reflections, it remained largely qualitative and self-reported. A more systematic approach, including quantitative performance measures and classroom observations, would provide a more robust understanding of the tools’ impact.

7 Conclusion

Successful implementation of online learning tools and management software can support students’ learning processes, increase accessibility, and promote inclusiveness in education. This paper presented an overview and categorization of available digital tools that can assist educators in creating scaffolding, organizing content, conducting assessments, and fostering collaborative learning environments. The aim was to identify practical and accessible tools that could enrich lesson delivery, enhance student engagement, and provide teachers with flexible instructional strategies. The categorization process presented in this study offers a practical tool for physics teachers navigating the increasingly complex landscape of

educational technology. By organizing digital platforms according to pedagogical function and classroom applicability, the framework supports informed decision-making and helps educators select tools that align with their instructional goals, available resources, and student needs. Importantly, the study emphasizes how these tools can be meaningfully integrated into hands-on physics laboratories, not as replacements but as complementary supports that scaffold conceptual understanding, facilitate collaboration, and enhance assessment. This alignment between digital tools and experimental learning environments is essential for maintaining the integrity of physics education in blended and technology-rich classrooms.

Looking ahead, the integration of AI into physics education represents a rapidly evolving frontier. AI offers promising applications, from intelligent tutoring systems and adaptive assessments to automated feedback and data-driven personalization. However, it also presents new challenges related to pedagogical design, equity, transparency, and teacher readiness. As educational technology continues to advance, supporting educators in understanding and thoughtfully implementing AI will be a crucial next step in the journey toward more effective and future-ready physics instruction.

Acknowledgment

The work was supported by the grant SVV no. 260828.

References

- Aprilo, I., Arfanda, P. E., Mappaompo, M. A., Saleh, M. S., & Hasbunallah. (2023). Technology adaptation in 21st century physical education learning: Literature review. *Jurnal Dunia Pendidikan*, 3(2), 102–107. <https://doi.org/10.55081/jurdip.v3i2.858>
- Cacabelos, A., Arce, M. E., Míguez, J. L., & Míguez, C. (2015). Integration of the free software GenOpt for a thermal engineering course. *Computer Applications in Engineering Education*, 24(3), 356–364. <https://doi.org/10.1002/cae.21713>
- Fonseca, L. M. M., Aredes, N. D., Leite, A. M., Santos, C. B. dos, Lima, R. A. G. de, & Scochi, C. G. S. (2013). Evaluation of an educational technology regarding clinical evaluation of preterm newborns. *Revista Latino-Americana de Enfermagem*, 21(1), 363–370. <https://doi.org/10.1590/s0104-11692013000100011>
- Guo, Y., Wang, Y., & Ortega, L. (2023). The impact of blended learning-based scaffolding techniques on learners' self-efficacy and willingness to communicate. *Porta Linguarum Revista Interuniversitaria de Didáctica de Las Lenguas Extranjeras*, 40, 253–273. <https://doi.org/10.30827/portalin.vi40.27061>
- Jatmiko, B., Prahani, B. K., Suprpto, N., Admoko, S., Deta, U. A., Lestari, N. A., Jauhariyah, M. N. R., Yantidewi, M., & Mulyati, D. (2021). Bibliometric analysis on online physics learning during COVID-19 pandemic: Contribution to physics education undergraduate program. *Journal of Physics: Conference Series*, 2110(1), 12–18. <https://doi.org/10.1088/1742-6596/2110/1/012018>
- Kregear, T., Babayeva, M., & Widenhorn, R. (2025). Analysis of student interactions with a large language model in an introductory physics lab setting. *International Journal of Artificial Intelligence in Education*, advance online placement. <https://doi.org/10.1007/s40593-025-00489-3>
- Kurnia, B., Zaidi, M., Suprpto, N., Alan, U., & Cheng, T.-H. (2022). The trend of physics education research during COVID-19 pandemic. *International Journal of Educational Methodology*, 8(3), 517–533. <https://doi.org/10.12973/ijem.8.3.517>
- Levin, S. (2023). LMS cluster data analysis for correcting students' learning. *2nd International Conference on Computer Applications for Management and Sustainable Development of Production and Industry (CMSD-II-2022)*, 1256416. <https://doi.org/10.1117/12.2669585>
- Levitin, D. J. (2015). *The organized mind: Thinking straight in the age of information overload*. Penguin.
- Newport, C. (2019). *Digital minimalism*. Penguin.
- Prahani, B., Alfin, J., Fuad, A., Saphira, H., Hariyono, E., & Suprpto, N. (2022). Learning Management System (LMS) research during 1991–2021: How technology affects education. *International Journal of Emerging Technologies in Learning (IJET)*, 17(17), 28–49. <https://doi.org/10.3991/ijet.v17i17.30763>
- Rizal, R., Rusdiana, D., Setiawan, W., & Siahaan, P. (2022). Learning Management System supported smartphone (LMS3): Online learning application in Physics for school course to enhance digital literacy of preservice Physics teacher. *Journal of Technology and Science Education*, 12(1), 191–203. <https://doi.org/10.3926/jotse.1049>
- Salas-Rueda, R.-A., De-La-Cruz-Martínez, G., Alvarado-Zamorano, C., & Prieto-Larios, E. (2022). The collaborative wall: A technological means to improving the teaching-learning process about physics. *Center for Educational Policy Studies Journal*, 12(4), 205–231. <https://doi.org/10.26529/cepsj.1167>

- Schanze, S., Groß, J., & Hundertmark, S. (2020). Computerbasiertes und kollaboratives Lernen in den Naturwissenschaften anhand von Aufgaben entwickeln und fördern. In S. Becker, J. Meißner-Koppelt, & C. Thyssen (Hrsg.), *Digitale Basiskompetenzen: Orientierungshilfe Und Praxisbeispiele Für Die Universitäre Lehramtsausbildung in Den Naturwissenschaften* (pp. 62–65). Joachim Herz Stiftung Verlag.
- Schubertová, K., Lukavský, J., Drobná, A., Volná, K., & Brom, C. (2023). Contextual animation in multimedia learning materials for pre-adolescents: The saga of null results continues. *Learning and Instruction, 87*, 101803. <https://doi.org/10.1016/j.learninstruc.2023.101803>
- Setiawan, A., Zuhri, M., & Zhang, Y. (2022). Learning Management System using Scorm and iSpring Free on physics learning. *Journal of Science Education and Practice, 6*(2), 95–106. <https://doi.org/10.33751/jsep.v6i2.5941>
- Simanullang, N. H. S., & Rajagukguk, J. (2020). Learning Management System (LMS) based on Moodle to improve students learning activity. *Journal of Physics: Conference Series, 1462*(1), 012067. <https://doi.org/10.1088/1742-6596/1462/1/012067>
- Skulmowski, A., & Xu, K. M. (2021). Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. *Educational Psychology Review, 34*(1), 171–196. <https://doi.org/10.1007/s10648-021-09624-7>
- Solvang, L., & Haglund, J. (2021). How can GeoGebra support physics education in upper-secondary school—a review. *Physics Education, 56*(5), 055011. <https://doi.org/10.1088/1361-6552/ac03fb>
- Suzuki, W., & Fitzpatrick, B. (2015). *Healthy brain, happy life*. Harper Collins.
- Tripathi, N. (2023). A study on “attention harvesting” strategies used by mass media and its implications. *Journal of Communication and Management, 2*(02), 153–157. <https://doi.org/10.58966/JCM20232215>
- Xu, Z., Zhao, Y., Liew, J., Zhou, X., & Kogut, A. (2023). Synthesizing research evidence on self-regulated learning and academic achievement in online and blended learning environments: A scoping review. *Educational Research Review, 39*, 100510. <https://doi.org/10.1016/j.edurev.2023.100510>

Úspěšnost nadaných žáků při řešení matematické slovní úlohy

Success in solving mathematical word problems in gifted pupils

 Irena Budínová¹

¹ Pedagogická fakulta, Masarykova univerzita, Poříčí 31, 602 00 Brno; budinova@ped.muni.cz

Jedním z přístupů k odhalování matematického nadání žáků je práce s komplexními slovními úlohami, jejichž řešení může ovlivňovat charakter úloh i rozdílné profily nadání. Studie se zaměřuje na to, jak tyto faktory ovlivňují identifikaci matematicky nadaných žáků 5. ročníku základní školy. Cílem bylo (1) zjistit míru korelace mezi třemi metodami identifikace nadání – hodnocením učitele, výsledky standardizovaného testu a výsledky výzkumného didaktického testu – a (2) identifikovat úlohy, které nejlépe odlišují nadané žáky od ostatních. Výzkumný test obsahoval pět subtestů, čtyři algebraické a jeden kombinatorický, zaměřené na projevy typické pro nadané žáky, jako je flexibilita myšlení, schopnost zobecňování a tvořivost při řešení problémů. Studie klade důraz na analýzu subtestu s úlohami typu „myslím si číslo“, které umožňují odhalit skryté projevy nadání a přispět k jeho přesnější diagnostice. Výzkumu se zúčastnilo 45 žáků školy zaměřené na podporu nadaných. Pilotní charakter studie má posloužit jako základ pro tvorbu standardizovaného testu pro žáky 5.–7. ročníku. Výsledky ukazují střední až silnou korelaci mezi použitými metodami, ale i existenci jevů, které mohou vést k přehlédnutí matematicky nadaných žáků.

Klíčová slova:

nadaní žáci, identifikace nadání, matematická problémová úloha.

Zasláno 10/2024

Revidováno 8/2025

Přijato 10/2025

One approach to identifying mathematical talent in students is to work with complex word problems, the solution of which can be influenced by the nature of the problems and different talent profiles. The study focuses on how these factors influence the identification of mathematically gifted students in the 5th grade of primary school. The aim was (1) to determine the degree of correlation between three methods of identifying talent – teacher assessment, standardized test results and research didactic test results – and (2) to identify tasks that best distinguish gifted students from others. The research test included five subtests, four algebraic and one combinatorial, focused on manifestations typical of gifted students, such as flexibility of thinking, ability to generalize and creativity in problem solving. The study emphasizes the analysis of the subtest with tasks of the “I think of a number” type, which allow to reveal hidden manifestations of talent and contribute to its more accurate diagnosis. The research involved 45 students from a school focused on supporting the gifted. The pilot nature of the study is intended to serve as a basis for the creation of a standardized test for students in grades 5–7. The results show a moderate to strong correlation between the methods used, but also the existence of phenomena that may lead to the overlooking of mathematically gifted students.

Key words:

gifted students, identification of giftedness, mathematical problems.

Received 10/2024

Revised 8/2025

Accepted 10/2025

1 Úvod

Snaha o přesné určení matematických schopností nadaných žáků nás provází již desítky let. Už na konci 19. století se objevily první pokusy o stanovení individuálních matematických schopností a již v té době byly pozorovány v kontextu řešení matematických problémů (Krutetskii, 1976; Szabo et al., 2024). Od 60. let 20. století byly do výzkumu aplikovány psychometrické metody, které měly určovat matematické schopnosti jednotlivců (Szabo et al., 2024). Tyto metody však nevedly k vymezení charakteristik matematicky nadaných žáků a od posledních desetiletí 20. století byly podrobeny kritice jakožto nespolehlivá metoda pro indikování matematických schopností (Szabo et al., 2024). Již v průběhu minulého století si totiž výzkumníci začali všimnout toho, že lidé, kteří se projevují jako nadaní v matematice, se raději zhloubají do problému, než aby předváděli rychlé řešení podle známého algoritmu. Například Hadamard (1945) upozornil, že špičkoví vědci, jako byl Albert Einstein, při řešení obtížného matematického problému zažili poměrně dlouhou, nevědomou dobu inkubace, než byli schopni transformovat intuitivní myšlenky na vědomé, užitečné nápady. Sternberg a Williams (2002) v této souvislosti zase hovoří o tendenci nadaných žáků používat složitější způsoby myšlení, než je nutné, a nazývají ji *komplexním myšlením*.

Identifikace nadaných žáků představuje důležitý krok k jejich dalšímu rozvoji a vhodnému vedení v hodinách matematiky. V kontextu matematického vzdělávání si výzkumníci již řadu let kladou otázku, jak se projevuje matematicky nadaný žák a jaké jsou možnosti jeho identifikace. Identifikace některých typů nadaných žáků je však ve školním prostředí problematická, což může vést k absenci rozvoje nadání žáka a následně k jeho podvýkonnosti (Betts & Neihart, 1988). Jak dokládají data, v České republice není

řada nadaných žáků identifikována. Podle České školní inspekce (ČŠI, 2022, s. 12) polovina základních škol v České republice, které mají pouze první stupeň, nehlásí žádné nadané žáky a mezi plně organizovanými základními školami je to pětina takových škol. V základních školách bylo ve školním roce 2023/2024 pouhých 0,27 % žáků identifikováno jako nadaných (MŠMT, 2024).

Identifikace a rozvoj nadaných žáků ve škole jsou však klíčové pro naplnění jejich potenciálu, podporu motivace a seberealizace, prevenci školního neúspěchu a přínos celé společnosti. Prvotní identifikaci nadání může provést učitel přímo ve třídě na základě svého posouzení, které však může být subjektivní. Jak upozorňují Jabůrek et al. (2022), učitelovo posouzení vychází z řady kritérií, jako jsou kognitivní schopnosti žáka, akademické schopnosti žáka, ale také například socio-emocionální nebo motorické dovednosti žáka. Pokud učitel žáka na základě svého posouzení vytipuje, má možnost si svůj soud ověřit jednak pomocí standardizovaného testu, který mu nabídne repertoár úloh, které nejsou zaměřeny na běžně probírané učivo a obsahují úlohy citlivěji měřící nadání, a jednak pomocí běžných školních testů. Ty ale obvykle obsahují specifické a omezené spektrum matematických úloh, které nemusí být pro identifikaci nadání vhodné. Jak upozorňuje Warne (2012), testy pro ověřování běžných školních znalostí a dovedností nejsou vhodné pro identifikaci nadání, neboť měří střední úroveň dovedností a pro nadané žáky jsou příliš snadné. Matematická úloha v sobě může obsahovat i různá další rizika, kvůli kterým nemusí spolehlivě odlišovat nadané žáky od ostatních. V některých úlohách z dřívějšího výzkumu se projevilo to, že zatímco běžní žáci volili algoritmické postupy, které se naučili ve škole, nadaní žáci promýšleli náročnější postupy a tím se snížila jejich úspěšnost (Budínová, 2018).

Hlavním cílem výzkumu bylo zjistit, jaký je vztah mezi identifikací nadání pomocí standardizovaného testu a pomocí výzkumného didaktického testu sestaveného z algebraických slovních úloh, a dále jaký je vztah mezi identifikací nadání pomocí standardizovaného testu a na základě posouzení učitelem. Výzkumný didaktický test byl sestaven na základě poznatků z odborné literatury i vlastních zkušeností autorky. Nadaní žáci často prokazují specifické schopnosti v řešení matematických problémů, zejména těch, které vyžadují systematizaci dat, abstraktní uvažování a schopnost nacházet neobvyklá řešení (Machů et al., 2013). Jedním z prostředků, jak tyto schopnosti odhalit, je zaměřit se na specifické dovednosti nadaných žáků, jako je například schopnost identifikovat vzory a vztahy mezi různými prvky nebo zobecnovat matematické poznatky (Gutiérrez et al., 2018). Jako typ úloh, který dokáže odhalovat právě tyto dovednosti, jsem vybrala algebraické slovní úlohy. Druhým vhodným typem úloh, ve kterých řešitel hledá více řešení, což klade nároky na organizaci a systematizaci dat, jsou úlohy kombinatorické (Vondrová, 2020).

Druhým cílem bylo zjistit, které úlohy z jednoho subtestu didaktického testu obsahujícího algebraické slovní úlohy odlišují nadané žáky od ostatních. Konečně bylo mým záměrem zjistit, jaké fenomény se projeví v písemných řešeních algebraických slovních úloh u nadaných žáků a u žáků ostatních. Tato analýza byla součástí kvalitativního vyhodnocení dat a shledávám ji jako důležitou zejména z důvodu posouzení toho, zda nadaní žáci používali jiné postupy nebo prováděli jiné typy chyb než ostatní žáci.

Výzkum měl smíšený design a zúčastnili se ho žáci tří tříd 5. ročníku jedné základní školy.

2 Teoretický rámec

2.1 Matematické schopnosti nadaných žáků

Krutetskii a jeho tým dospěli k závěru, že matematicky nadaní žáci jsou schopni (Krutetskii, 1976)¹

- efektivně získávat a formalizovat matematické informace;
- efektivně zpracovávat matematické informace – logickým myšlením, rychlým a širokým zobecněním matematických pojmů a vztahů a snahou o jasná a jednoduchá řešení;
- efektivně uchovávat a vybavovat zobecněné matematické vztahy;
- představovat obecně propojené matematické komponenty.

Matematicky nadaní žáci jsou dále schopni kompenzovat své méně rozvinuté schopnosti těmi více rozvinutými (Krutetskii, 1976), což jim umožňuje snadněji hledat alternativní řešení úlohy, když neznají algoritmus, který je na řešení potřeba. Při řešení problémových úloh mají tendenci zvnitřňovat intuitivní myšlenky a přístupy a zkracovat své úvahy (Krutetskii, 1976; Szabo et al., 2024). V důsledku toho je pro učitele obtížné sledovat matematické uvažování nadaných žáků při řešení problémových úloh.

Když se v průběhu 20. století začalo ukazovat, že intelektově nadaní lidé nemusí v dospělém životě své schopnosti využívat podle toho, jak se od nich očekávalo na základní škole, začal se Renzulli (1978) zabývat

¹Přestože je tento popis už více než 40 let starý, v literatuře je stále citován (Szabo et al., 2024).

výzkumem tvořivě-produktivních lidí, kteří dosáhli v určité oblasti jedinečných výsledků. Zjistil, že se tyto lidé vyznačovali interakcí tří vlastností, a to nadprůměrnými schopnostmi, tvořivostí a angažovaností v úkolu. Odlišil školní nadání, které lze dobře využít ve škole, a tvořivě-produktivní nadání, které je potřeba mimo školu. I z tohoto pohledu je tedy obtížné říct, kdy již budeme žáka ve škole považovat za nadaného a dostaneme možnost (zejména časovou a finanční) se mu více věnovat.

Pro různé *stupně nadání* bývá využívána výše inteligenčního kvocientu, a to zpravidla tak, jak ukazuje tab. 1 (např. Fořtík & Fořtíková, 2007; Havigerová, 2011).

Tab. 1: Jeden ze způsobů rozdělení výkonu v inteligenčním testu do intervalů

IQ	Označení úrovně kognitivních schopností jedince	Výskyt v populaci
115–129	Bystrý jedinec	14,0 %
130–144	Nadprůměrně nadaný	2,0 %
145–159	Vysoce nadaný	0,1 %

Zcela jiné intervaly pro míru intelektového nadání však nacházíme u Gagného (2005, s. 109, tab. 2).

Tab. 2: Gagného metrický systém intervalů v populaci nadaných/talentovaných (Gagné, 2005)

IQ	Označení	Výskyt v populaci
120–134	Mírně nadaný	10 %
135–144	Středně nadaný	1 %
145–154	Vysoce nadaný	0,1 %
155–164	Mimořádně nadaný	0,001 %
165 a více ²	Extrémně nadaný	0,0001 %

Renzulli vytvořil identifikační systém, který nevychází pouze z výše inteligenčního kvocientu, ale zaměřuje se rovněž na tvořivost a angažovanost v úkolu. Navrhl, aby v tomto systému bylo 15 % studentů považováno za nadané (Renzulli, 2005).

Z těchto různých pojetí nadání vyplývá, že i ve školním prostředí bývá obtížné určit, koho označit jako nadaného a koho už ne. Hříbková (2009) uvádí, že za nadaného bývá ve škole považován ten žák, který v daném předmětu dosahuje mimořádných výkonů ve srovnání s výkony svých vrstevníků. To však může v běžné výuce být i žák, který se dobře učí, avšak nadaný není. Pokud se však chceme zabývat nadanými žáky, a to jak na úrovni teoretické, tak praktické, některé rozdělení přijmout musíme.

2.2 Možnosti identifikace nadání

Nadání tedy můžeme chápat různým způsobem a podle toho také provádět jeho identifikaci. Jednoduchý způsob identifikování nadání vychází z IQ definice nadání (Budínová, 2018). Míra nadání v tomto modelu je posuzována od IQ 120. Při hranici 120 se jedná přibližně o 9,1 % populace a při hranici 130 o 2,2 % populace (McBee & Makel, 2019). Pokud se na nadání budeme dívat v širším spektru, např. pomocí Renzulliho Tříkruhového modelu nadání (Renzulli, 1978), budou počty nadaných žáků značně jiné. Podle Renzulliho identifikačního systému, který se zaměřuje na schopnosti, tvořivost a angažovanost v úkolu, by mělo být za nadané považováno přibližně 15 % studentů (Renzulli, 2005), jak již bylo uvedeno výše.

Zásadní roli v tom, jak se nadání u jednotlivých žáků projevuje, hrají tzv. *profily nadání*. Nadání jedinci nejsou homogenní skupina, liší se schopnostmi, silnými i slabými stránkami, komunikací s okolím, způsoby učení se aj. Profily nadání určují, v jakých oblastech a jakým způsobem se nadání projevuje, a jsou proto klíčové pro správnou identifikaci i efektivní pedagogickou podporu. Betts a Neihart (1988) zavedli šest profilů nadaných žáků, které se zaměřují na různé projevy nadání. Pojmenovali je *úspěšní nadaní*, *autonomní žáci*, *skrývači nadání*, *defenzivní odpadlíci*, *provokatéři* a *žáci s dvojitou výjimečností*. Z těchto šesti profilů nadaných žáků jsou ve škole snadno identifikovatelné pouze první dvě skupiny žáků. Betts a Neihart (1988) tvrdí, že asi 90 % identifikovaných žáků ve školních programech má profil úspěšných nadaných. U žáků z ostatních profilů je identifikace nadání komplikovaná, a to z důvodu možného nesouladu mezi vnímáním nadání a jeho projevů a skutečnými projevy dítěte.

Příkladem jsou nadaní žáci s dyslexií, kteří jsou písemným testováním velmi obtížně identifikovatelní, neboť jsou schopni při běžném školním testování dosahovat zlomku výkonu, který odpovídá jejich potenciálu. Výkony těchto žáků navíc klesají při časově omezeném testování (Portešová, 2011; Portešová et al., 2014). Žáci s dyslexií tak mohou dosáhnout jen částečného úspěchu oproti tomu, kdyby měli možnost napsat test v nestresových podmínkách bez časového omezení. Dalšími faktory omezujícími výkony

²V České republice neexistuje test, který by měřil intelektové nadání v takto vysokých pásmech.

žáka s dyslexií jsou např. deficit ve schopnosti rychle si vybavit pojmy z dlouhodobé paměti, v pracovní rychlosti, v zapamatování či v nedostatečné grafomotorické rychlosti apod. (Portešová et al., 2014). Jak Portešová et al. (2014) upozorňují, u nadaných žáků s dyslexií vznikají různé paradoxy mezi jejich kognitivními schopnostmi a jejich handicapem. Proto jsou tito žáci často označováni jako *paradozní žáci a studenti* (Tannenbaum & Baldwin, 1983).

Identifikace nadaných žáků je komplexním procesem, který vyžaduje kombinaci různých metod a nástrojů. Podle Machů et al. (2013) stojí na počátku identifikace nominační metody (pozorování dítěte, nominace učitelem, spolužákem nebo rodičem), poté následují hromadné metody (didaktické testy, skupinově zadávané IQ testy) a vše je završeno individuálními metodami, vedenými odborníkem, např. z pedagogicko-psychologické poradny (rozhovor, individuálně zadávané testy). Diagnostika prostřednictvím didaktických testů tedy figuruje v prostřední fázi identifikace nadání a může proběhnout na půdě školy. Testová úloha má však také své různé limity a rizika, kvůli kterým nemusí k identifikaci nadání dojít. Dřívější studie poukázala na fakt, že některé úlohy zvládnou lépe žáci, kteří nejsou označeni jako nadaní (Budínová, 2018). Jsou různé faktory, které úspěšnost nadaných žáků ovlivňují – aritmetická snadnost či náročnost úlohy, možnost použití náročnějšího matematického aparátu, ke které inklinují právě nadaní žáci z důvodu komplexního myšlení, které popsali Sternberg a Williams (2002), nebo volba problémových úloh, u nichž žák nezná algoritmus řešení a musí postup řešení odhalit na základě svých dosavadních znalostí, dovedností a zkušeností a které jsou náročné pro všechny skupiny žáků.

Z uvedených důvodů může vzniknout nesoulad mezi identifikací nadání jedince různými metodami. Běžné školní testy prověřují zvládnutí základního učiva, zpravidla neobsahují problémové úlohy a z toho důvodu se pro identifikaci nadaných žáků nehodí. Existuje například nesoulad mezi aritmetickými schopnostmi a schopností logického uvažování. Straker (1980) uvádí, že mnozí úspěšní matematici o sobě přiznávají, že nejsou příliš dobří v aritmetice. Školní identifikace nadání může být dále poznamenána tím, jací žáci jsou považováni za nadané. Hříbková (2009) uvádí, že za nadaného je často považován ten žák, který svými výkony převyšuje své vrstevníky. Může se ale právě jednat o dobré počtáře, kteří nemají rozvinuté logické myšlení. Oproti tomu těžkopádně působící nadaný žák, který při řešení úlohy používá komplexní myšlení (Sternberg & Williams, 2002), za nadaného považován být nemusí.

2.3 Problémové matematické úlohy jako zdroje identifikace nadání

Matematické problémové úlohy a proces jejich řešení jsou často považovány za ústřední součásti matematického vzdělávání (Liljedahl et al., 2016). Nemusí být ovšem tak jednoznačné a jednoduché určit, která úloha je pro řešitele problémová. Obecně lze říci, že pokud se řešitel s daným typem úlohy již setkal a osvojil si algoritmus řešení, problémová pro něj není, ať by byla jakkoli náročná či nestandardní. Matematický problém by proto měl obsahovat určitou část řešení, která je pro řešitele zcela nová. Taková úloha by mohla lépe odlišovat nadané žáky od ostatních, protože nadaní žáci by měli mít větší schopnost se v nové situaci zorientovat a volit alternativní způsoby řešení (Krutetskii, 1976). Nadaní žáci mohou při řešení nerutinního úkolu projevit kreativní matematické myšlení (Lithner, 2013), zatímco ostatní žáci mohou inklinovat k tomu, že přebírají předkládané postupy, aniž by byli jejich tvůrci.

Matematická úloha, která by se dala využít v testu tak, aby odlišila nadaného žáka, by však měla splňovat určité nároky, které nyní zmíním.

Za prvé, některé úlohy mohou být příliš náročné i pro matematicky nadané žáky, pokud je mají zvládnout v časovém limitu. V některých úlohách dokonce mohou být průměrní žáci úspěšnějšími řešiteli než žáci nadaní (Budínová, 2018). Bývají to ty úlohy, u kterých žáci s průměrnými výsledky v matematice mohou použít jednoduché algoritmické řešení, zatímco nadaní žáci vymýšlejí nevšední postup.

Za druhé, u nadaných žáků by se podle Bloomovy taxonomie kognitivních cílů (Bloom, 1956; Krathwohl, 2002) měla pozornost přesunout z úrovně aplikace na úroveň analýzy. Na této úrovni je žák nucen úkolu hlouběji porozumět, protože zadání není koncipováno tak, aby bylo řešení zřejmé na první pohled. K vyřešení úlohy je třeba hlubšího vhledu a schopnosti rozpoznat vztahy mezi jednotlivými částmi problému. Tyto úlohy obvykle zahrnují více kroků a kladou vyšší nároky na logické uvažování a schopnost strukturovat řešení.

Za třetí, s nadáním úzce souvisí i úroveň metakognice žáka. Metakognice je vnímána jako důležitá součást žákova rozvoje a poznávacího procesu. Jedna z prvních definic metakognice ji vymezuje jako znalost a regulaci vlastního kognitivního systému (Brown, 1987). Můžeme říci, že metakognice je schopnost jedince uvědomovat si vlastní poznávací schopnosti – myšlení, vnímání, paměť apod. – a využívat výsledky tohoto uvědomění při učení, řešení problémů a při dalších aktivitách, do kterých se tyto poznávací schopnosti zapojují (Straka, 2021). Jiní autoři ji vymezují jako schopnost žáka monitorovat svoji aktuální úroveň porozumění danému tématu (Bransford et al., 2004). Žák se učí poznávat různé postupy, zvažuje jejich efektivitu, učí se vhodně volit řešitelské strategie, ale také se učí poznávat sám sebe, své silné a slabé stránky.

Nadaní žáci často vykazují oproti vrstevníkům zvýšené metakognitivní dovednosti, tedy schopnost plánovat, monitorovat a hodnotit své myšlení a učení (Alexander et al., 1995; Machů et al., 2013). Avšak i nadaní žáci potřebují svoji metakognici rozvíjet, což je možné zejména prostřednictvím problémových úloh. Pokud se žák s takovými úlohami nesetkává, metakognice se rozvíjet nemusí a nadaný žák nerozšiřuje svoji znalostní, dovednostní a metakognitivní základnu (Alexander et al., 1995). Problémové matematické úlohy umožňují nadaným žákům rozvíjet metakognici, pokud se nezaměřujeme pouze na správnost výsledku, ale rovněž na postup řešení – zda žák nepoužívá povrchové strategie řešení, zda si data vhodně systematizuje, aby byl schopen úspěšně najít řešení, zda kontroluje správnost svého postupu i výsledku aj.

Jako vhodný typ matematických úloh, které splňují výše uvedené nároky, se jeví slovní úlohy. Slovní úlohy jsou důležitou součástí výuky matematiky. Umožňují aplikovat osvojené matematické učivo, rozvíjí u žáků schopnost použít matematické poznatky v běžných situacích, rovněž rozvíjí matematické myšlení, neboť často neumožňují žákům použít mechanicky osvojené procedury (Vondrová et al., 2019). Pro řešení slovních úloh musí žák mimo jiné dobře porozumět textu. Např. Cummins et al. (1988) uvádějí, že stejná úloha zadaná jednou aritmeticky a podruhé jako slovní úloha má podstatně větší úspěšnost v prvním případě, a současné výzkumy potvrzují úzkou souvislost mezi řešením slovních úloh a čtenářskou gramotností (Vondrová et al., 2019). To naznačuje, že k úspěchu při řešení slovních úloh přispívají i jiné faktory než pouze matematické dovednosti.

Boonen et al. (2016) uvádějí, že pro řešení slovních úloh jsou zapotřebí dvě klíčové dovednosti. Za prvé, důležitým faktorem přispívajícím k hlubšímu porozumění textu slovní úlohy je schopnost sestavit smysluplnou a koherentní mentální reprezentaci obsahující všechny prvky relevantní pro řešení, které jsou odvozeny z textové báze slovní úlohy. To znamená, že řešitelé slovních úloh si musí vytvořit kvalitní situační model, v němž si uvědomují aktéry a vztahy mezi nimi skryté v textu (Boonen et al., 2016; Vondrová et al., 2019). To může pro žáka představovat obtíž v případě, kdy si takový situační model nevytvoří a použije *povrchové strategie řešení* (Vondrová et al., 2019), ve kterých se např. zaměřuje pouze na čísla v zadání bez toho, aby je analyzoval (Boonen et al., 2016). Druhou důležitou individuální dovedností, která rozhoduje o úspěšnosti řešení slovních úloh a která je podložena výzkumnými důkazy, je již zmíněná schopnost žáka číst s porozuměním (Boonen et al., 2016).

Z těchto závěrů vyplývá fakt, že při řešení slovních úloh mohou být úspěšnější ti řešitelé, kteří se více vyznaží v textu, mají širší slovní zásobu a umí si představit kontext úlohy. U matematicky nadaných žáků to může platit pro ty žáky, kteří mají své silné stránky také v jazykové oblasti. To ale jistě nejsou všichni matematicky nadaní žáci. Dle Gardnerovy teorie mnohačetných inteligencí (Gardner, 2006) jsou jazyková a logicko-matematická inteligence navzájem nezávislé. Na druhou stranu, jazykově i matematicky nadaní žáci vykazují obecné charakteristiky nadaných žáků, kterými jsou především bohatá slovní zásoba, schopnost abstrakce a generalizace, metakognitivní dovednosti, kritické myšlení, flexibilita a originalita řešení (Machů et al., 2013). Tyto charakteristiky tvoří určité překryvy mezi jazykovým a matematickým nadáním.

Kromě těchto dvou klíčových dovedností sestává komplexní teorie o řešení slovních úloh z dalších komponent, jako je například potřeba formálního výpočtu pro získání výsledku (Nathan et al., 1992). Všechny vyžadované dovednosti vedou k tomu, že jsou slovní úlohy pro mnoho žáků, nadaných i ostatních, výzvou a mohou v nich selhávat ve větší míře než v úloze vyžadující provedení konkrétní početní operace.

2.4 Výzkumné otázky

Výzkum vychází z popsaného teoretického rámce a má dva cíle. Tím prvním je zjistit vzájemný vztah mezi stanovením matematického výkonu žáků třemi metodami: standardizovaným testem, vlastním testem, který budu nazývat *výzkumným testem*, a posouzením učiteli. Výzkumný test měl za cíl zjistit, které úlohy jsou vhodné pro identifikaci nadání. Druhým cílem je prozkoumat schopnost konkrétní testové úlohy odlišovat žáky vzhledem k nadání.

Pro výzkum byly tedy stanoveny následující výzkumné otázky:

- O1: *Jak si odpovídají výsledky identifikace nadání ve standardizovaném testu a ve výzkumném testu?*
- O2: *Jak si odpovídá posouzení matematických schopností žáka učitelem a výsledek ve standardizovaném testu?*
- O3: *Jak odlišuje nadání konkrétní úloha jednoho ze subtestů výzkumného testu?*
- O4: *Jaké fenomény se vyskytují v řešeních žáků, nadaných i ostatních, u úloh v jednom ze subtestů výzkumného testu?*

3 Metodologie

Pro zodpovězení výzkumných otázek byla zvolena smíšená metodologie. Byly vyhodnoceny výsledky celého výzkumného testu a každého ze subtestů. Sledovala jsem korelace mezi třemi uvedenými metodami stanovení nadání žáka – mezi stanovením nadání učitelem, percentilem určeným standardizovaným testem a výsledkem získaným ve výzkumném testu. Na stanovení vztahu mezi výsledkem ve standardizovaném testu a stanovením nadání učitelem byla použita Spearmanova korelace. Rovněž na stanovení vztahu mezi výsledkem ve standardizovaném testu a výsledkem ve výzkumném testu byla využita Spearmanova korelace. Kromě deskriptivní statistiky byly stanoveny hypotézy popisující vztah mezi úspěšností v úloze zkoumaného subtestu a výsledkem standardizovaného testu. Hypotézy byly testovány pomocí Pearsonova chí-kvadrátu.

Kvalitativní šetření se zaměřovalo na fenomény vyskytující se v řešeních žáků.

3.1 Účastníci výzkumu

Výzkum probíhal v průběhu školního roku 2023/2024 v jedné základní škole Jihomoravského kraje, která má jako jednu ze svých priorit péči o nadané žáky. Nejedná se však o školu výkonově zaměřenou nebo vzdělávající pouze nadané žáky. To znamená, že zde najdeme všechny výkonnostní skupiny žáků. Škola se snaží o vnitřní diferenciaci a žáky separuje podle výkonu pouze jednou týdně, kdy se učitelé zaměřují na výuková specifika každé ze tří skupin (žáci s poruchami učení a žáci s podprůměrnými výsledky v matematice, žáci s průměrnými výsledky v matematice, nadaní žáci). Žáci nejsou do těchto skupin rozdělováni podle známek, ale podle jejich schopností a motivace k rozšiřujícímu studiu matematiky. Například skupina nadaných žáků, která je vedena v rámci klubu pro nadané, je vybírána na základě speciálního školního testování a také dlouhodobého pozorování žáků, aby se do skupiny dostali žáci motivovaní pro vzdělávání.

Spolupráce probíhala se třemi učiteli. Účast ve výzkumu byla pro žáky dobrovolná a z celkového počtu 83 žáků ze tří tříd 5. ročníku se do testování zapojilo 45 žáků. Pátý ročník byl vybrán ze dvou důvodů. Prvním důvodem bylo to, že standardizovaný test TIM³⁻⁵, který bude představen dále, je určen pro žáky 3. až 5. ročníku. Druhým důvodem byl záměr vytvořit na základě výsledků tohoto výzkumného šetření standardizovaný test pro žáky 5. až 7. ročníku.

Žáky jsem znala osobně. Se skupinou nadaných žáků jsem v daném školním roce pracovala v rámci jejich klubu pro nadané a všechny zapojené žáky jsem pozorovala při řešení úloh během jejich plnění výzkumného testu.

3.2 Popis výzkumného šetření

Na začátku výzkumu byli vyučující požádáni, aby uvedli, jak žáky posuzují vzhledem k jejich výkonu v matematice. Ve vzorku bylo podle nich 5 matematicky nadaných žáků, 8 žáků nadaných, 1 nadaný žák s dysortografií, 1 žák s nadprůměrnými výsledky, 28 žáků s průměrnými výsledky, 2 žáci s podprůměrnými výsledky.

Žáci byli nejdříve podrobeni testování standardizovaným *Testem pro identifikaci nadaných žáků v matematice TIM³⁻⁵* (Cígler et al., 2017). Dle Cíglera et al. (2017) test TIM³⁻⁵ vykazuje dobrou vnitřní konzistenci (Cronbachovo α v rozmezí 0,67–0,91 dle ročníku a formy), uspokojivou stabilitu ($r = 0,68$ – $0,78$) a velmi vysokou shodu mezi hodnotiteli (ICC = 0,99). Faktorová analýza potvrzuje jednodimenzionální strukturu testu (TLI = 0,97; RMSEA = 0,02). Souběžná validita s inteligenčními testy je středně silná ($r \approx 0,64$) a test zároveň dobře rozlišuje nadané žáky od běžné populace (Cohenovo $d \approx 1,2$ – $1,8$).

Test (Cígler et al., 2017) určil pro jednotlivé žáky jejich percentil, na kterém se pravděpodobně pohybují v matematických schopnostech. Na základě výsledků v testu byli žáci rozděleni do skupin nadaných a ostatních. Při stanovování hranice nadání jsem vycházela z teoreticky popsanych způsobů rozdělení nadání, v nichž jsem se přiklonila k Renzulliho pojetí, v kombinaci s dlouhodobým pozorováním žáků, ke kterému jsem dostala příležitost při sledování jejich práce. Mnou vybraná hranice nadání byl 85. percentil ve standardizovaném testu TIM³⁻⁵. Z dlouhodobého hlediska byli žáci z tohoto pásma schopni podávat stabilně nadprůměrné výkony, a to i v nestandardních úlohách.

Následně žáci řešili postupně v měsících prosinec roku 2023, leden, únor a březen roku 2024 pět subtestů. Na řešení měli tolik času, kolik potřebovali (žák test odevzdal, až byl hotov), aby se odstranily například hendikepy žáků s dyslexií. Byli vždy požádáni, ať uvádí také postup řešení úlohy.

3.3 Výzkumný test

Výzkumným nástrojem byl didaktický test, sestávající z pěti subtestů o šesti slovních úlohách (úlohy typu „myslím si číslo“, dělení celku na nestejně části, úlohy vedoucí na soustavy rovnic, diofantické

rovnice, kombinatorika, viz příloha). Úlohy v subtestech byly gradovány od těch, které by dle očekávání autorky, podloženého dlouholetým praktikováním diferencované výuky matematiky, měli zvládnout i žáci s průměrnými výkony v matematice, až po ty, které by měly být výzvou i pro nadané žáky. Vzhledem k tomu, že se jednalo o pilotní studii, nebyla gradace úloh předem ověřována.

Do testu byly vybrány slovně formulované úlohy s kontextem i bez něj. Slovní úlohy jsou, jak bylo uvedeno výše, pro žáky výzvou, protože je nutné zorientovat se v textu, najít podstatné informace a ty zapsat a zpracovat matematicky. Algebraické slovní úlohy jsou náročné na nalezení vztahů mezi veličinami. V nižších ročnících jsou řešeny experimentální či aritmetickou strategií, od 8. ročníku, kdy se obvykle probírají rovnice, mohou žáci použít také strategii algebraickou. I když v dřívější studii bylo zjištěno, že u nadaných žáků dochází někdy k osvojení algebraické strategie dříve, než je probírána ve škole (Budínová, 2018). Je pravděpodobné, že se u těchto úloh projeví aritmetická zručnost žáka, jeho aritmetická představivost i dovednost najít pro situaci ze zadání správný matematický model. Kombinatorické úlohy kladou důraz na porozumění struktuře problému a vytvoření systému při řešení, který umožní najít všechny možnosti (Vondrová, 2020). Nadaní žáci mohou mít v těchto úlohách výhodu oproti svým vrstevníkům díky kreativitě (Krutetskii, 1976; Renzulli, 1978), flexibilitě a originalitě myšlení (Machů et al., 2013) a schopnosti přecházet k alternativním postupům řešení (Krutetskii, 1976).

Úlohy do výzkumného testu byly voleny vzhledem k autorce známým parametrům, které mohou ovlivňovat obtížnost úlohy. Např. náročnost úlohy ovlivňuje volba čísel. Žáci nejsnadněji pracují s malými přirozenými čísly, velká a složitá čísla (například čísla s velkým počtem cifer, zlomky s velkým jmenovatelem aj.) mohou snížit úspěšnost (Vondrová et al., 2019). Například dělení $24 : 6$ zvládnou žáci i z paměti, oproti tomu $14\,652 : 37$ může některé žáky od řešení úlohy odradit. V úlohách jsem proto čísla volila tak, aby žáka nenaváděla na požadovanou operaci, ale zároveň aby nebyla zdrojem dalších chyb, které jsem nesledovala. Do slovních úloh můžeme dále přidávat *antisignály* (*distraktory*) nebo nadbytečné údaje, které úlohu dále ztěžují. Antisignál můžeme popsat jako slovo nebo slovní spojení, které navádí řešitele na nesprávnou operaci (např. Adetula, 1990). Různé studie (mezinárodní i tuzemské) ukázaly, že při použití antisignálu je slovní úloha pro žáky obtížnější (Vondrová et al., 2019), proto jsem ji zařadila do výzkumného testu.

3.4 Popis úloh subtestu Úlohy typu „Myslím si číslo“

V tomto článku se zaměřuji na první subtest výzkumného testu. Při vyhodnocování byly posuzovány:

- správnost výsledku;
- zapsaný postup (zda žák zapsal jen výsledek, zda rozepsal postup, jaký postup řešení byl použit);
- zapisování odpovědi;
- písemné provedení zkoušky správnosti;
- specifické fenomény (implikační zápis, volba chybné metody řešení).

V následujícím textu budou představeny jednotlivé úlohy subtestu a autorčino očekávání řešení žáků.

Úloha 1. *Myslím si číslo. Když k němu přičtu 8, dostanu 22. Které číslo si myslím?*

Úloha 1 je jednokroková, k jejímu řešení stačí jeden výpočet. Jednoduchost úlohy vede u řady žáků k tomu, že nemají potřebu zapisovat postup a píšou jen výsledek, který určili z paměti. Očekávala jsem, že žáci zapíšou pouze výsledek (14) nebo odpověď („myslíš si 14“), případně aritmetický výpočet $22 - 8 = 14$. Neočekávala jsem problémy s antisignálem (je použito slovo přičtu, ačkoli se má odčítat), neboť úloha je snadná na vytvoření představy. V testu úloha plní roli spíše motivační, aby žáci dosáhli úspěchu na začátku testu.

Úloha 2. *Myslím si číslo. Když ho vynásobím dvěma a od výsledku odečtu 3, dostanu 7. Které číslo si myslím?*

Úloha 2 je dvoukroková, mnoho žáků však zřejmě nebude mít potřebu postup zapisovat a výsledek určí z paměti. Očekávala jsem zapsání výsledku (5) nebo odpovědi („myslíš si 5“), případně aritmetického výpočtu $7 + 3 = 10$, $10 : 2 = 5$.

Úloha 3. *Myslím si číslo. Když od něj odečtu 7 a výsledek vydělím dvěma, dostanu 8. Které číslo si myslím?*

Úloha 3 je také dvoukroková a na první pohled se až na pořadí operací příliš neliší od druhé úlohy. Rozdíl oproti druhé úloze je však ve větších číslech, což by mohlo snížit úspěšnost. Očekávala jsem aritmetické řešení $8 \cdot 2 = 16$, $16 + 7 = 23$. Dále bylo sledováno, zda někteří žáci neprovedli operaci $8 : 2$ místo $8 \cdot 2$, což by mohli provést v případě, že použijí povrchovou strategii řešení.

S úlohami podobnými uvedeným třem se zkoumaní žáci občas setkali ve výuce a byli seznamováni s tzv. postupem řešení od konce. V posledních minimálně pěti týdnech před testováním však takové úlohy do výuky zařazeny nebyly.

Úloha 4. *Myslím si číslo. Když ho vynásobím samo sebou a k získanému součinu přičtu myšlené číslo, dostanu 90. Které číslo si myslím?*

Úloha 4 je nestandardní úloha a lze předpokládat, že se žáci s tímto typem setkali poprvé až v testu. Nelze na ni použít metodu od konce a je nutné vytvořit situační model, který umožní úlohu vyřešit. Čísla jsou úmyslně volena tak, aby při znalosti malé násobilky žák řešení rychle odhalil: $9 + 9 \cdot 9 = 9 + 81 = 90$. Pokud by žák řešil úlohu na základě této představy, očekávala jsem, že zapíše pouze výsledek. V opačném případě může postupovat experimentálně. Zvolí např. číslo 6 a provede s ním operace: $6 + 6 \cdot 6 = 6 + 36 = 42$, postupně zvětšuje volené číslo.

Úloha 4 má druhé řešení v oboru celých čísel: $-10 \cdot (-10) - 10 = 90$. Vzhledem k věku žáků jsem však neočekávala, že budou uvažovat tímto způsobem, a plný počet bodů obdrželi po nalezení řešení z oboru přirozených čísel.

Úloha 5. *Myslím si dvojciferné číslo. Když vytvořím jeho ciferný součet (například ciferný součet čísla 25 je $2 + 5 = 7$) a vynásobím ho dvěma, dostanu myšlené číslo. Které číslo si myslím?*

Rovněž úloha 5 je nestandardní a vyžaduje nejen vytvoření modelu situace, ale také porozumění všem termínům v zadání. Očekávala jsem experimentální řešení, buď metodou pokusu a omylu, nebo řízeným experimentem. Při řízeném experimentu může žák postupovat tak, že vybere dvojciferné číslo, určí jeho ciferný součet, vynásobí ho dvěma a z výsledku posoudí, jak vybrané číslo změnit: např. $36; 3 + 6 = 9; 9 \cdot 2 = 18$, číslo 18 je o 18 menší než 36, pokusí se zmenšit vybrané číslo, např. $27; 2 + 7 = 9; 9 \cdot 2 = 18$, číslo 18 je o 9 menší než 27. Postupně lze vypořádat, že ciferný součet musí být alespoň 5, jinak jeho vynásobením dvěma nezískáme dvojciferné číslo. Největší ciferný součet je 18, což je vynásobeno dvěma 36, a dané číslo musí být menší než 36. Výsledkem je 18; $1 + 8 = 9; 9 \cdot 2 = 18$.

Úloha 6. *Myslím si číslo. Když k němu přičtu trojnásobek tohoto čísla, dostanu 24. Které číslo si myslím?*

Úloha 6 je úloha s antisignálem. Jako antisignál zde vystupuje slovní spojení „přičtu trojnásobek“, které může žáky navést k tomu, aby číslo 24 dělili třemi. Pokud žák bude situaci analyzovat a vytvoří si situační model, bude postupovat například dělením čísla 24 čtyřmi, neboť $24 : 4 = 6; 6 + 3 \cdot 6 = 24$. Hledané číslo je 6. Pokud žák bude postupovat povrchově, nechá se zmást slovem „trojnásobek“ a bude počítat $24 : 3 = 8$.

3.5 Analýza dat

Data, tedy písemná žákovská řešení, byla nejdříve zpracována co do úspěšnosti řešení. Úlohy v subtestech 1, 2 a 3 byly bodovány 1 (správná odpověď) nebo 0 (nesprávná odpověď). V těchto subtestech bylo tedy možné dosáhnout maximálně 6 bodů. V úlohách, které měly více řešení, což byly úlohy vedoucí na diofantické rovnice a kombinatorické úlohy (subtesty 4 a 5), bylo bodování odlišné: V případě, že žák našel všechna řešení, dostal 2 body, v případě, že našel jen některá řešení, dostal 1 bod. V celém testu mohl tedy žák získat nejvýše 46 bodů.

Při analýze dat byla nejprve použita deskriptivní statistika, s jejíž pomocí byly zjišťovány výsledky pro celý výzkumný test a jednotlivé subtesty, četnosti bodových hodnocení jednotlivých úloh, relativní četnosti, aritmetický průměr, medián, kvartilové rozdělení a charakteristiky rozptýlení (rozptyl a směrodatná odchylka). Následně byly stanoveny korelace mezi jednotlivými metodami posuzování matematických schopností: jednak korelace mezi percentilem, který byl určen standardizovaným testem, a výsledkem ve výzkumném testu, jednak korelace mezi výsledky v testu a úrovní nadání, která byla stanovena vyučujícími. Za tím účelem byl použit Spearmanův korelační koeficient pořadí, což je statistická metoda používaná k hodnocení síly a směru monotónního vztahu mezi dvěma proměnnými, jejichž hodnoty lze seřadit podle pořadí. Tento koeficient je vhodný zejména tehdy, když nelze předpokládat lineární vztah mezi proměnnými nebo když data nevyhovují požadavkům na normalitu, k čemuž dochází například tehdy, když se pracuje s malými výběry (Hendl, 2012). Korelace byly testovány na hladině významnosti 5 %.

Kromě deskriptivního popisu výsledků výzkumného souboru byly stanoveny také hypotézy popisující vztah mezi úspěšností v každé úloze subtestu 1 a nadáním. U každé z šesti úloh byla tedy stanovena nulová hypotéza H_0 : *Úspěšnost v úloze nezávisí na nadání*. Hypotézy byly testovány na hladině významnosti 5 %. Výsledky byly statisticky zpracovány v programu *Statistica*.

V kvalitativní části byla provedena analýza řešení jednotlivých úloh, různých přístupů žáků k řešení a opakující se chyby. Byly sledovány strategie řešení úloh a jevy charakteristické pro jednotlivé úlohy,

jako je např. implikační zápis, využití distraktoru aj. Nejprve jsem vytvořila rámec kódování na základě didaktické analýzy úloh, a pak jsem tento rámec rozšířila o nové kategorie, které vzešly přímo z reálných řešení žáků. V článku budou představeny výsledky pro úlohy subtestu 1. Ostatní subtesty již byly také analyzovány a některé byly či budou publikovány (doposud Budínová, 2025).

4 Výsledky

4.1 Vyhodnocení výzkumného testu jako celku

Stručně představíme celkové výsledky výzkumného celku a poté se zaměříme na to, jak koresponduje identifikace nadání prostřednictvím standardizovaného testu a identifikace nadání prostřednictvím výzkumného testu. Dále nás bude zajímat, jak koresponduje stanovení nadání učitelem s identifikací nadání standardizovaným testem.

Jak je uvedeno výše, žáci mohli ve všech úlohách výzkumného testu získat maximálně 46 bodů. Nejvyšší získané skóre bylo 41 bodů a nejnižší 1 bod. Průměr byl 20,1 bodů a směrodatná odchylka 10,3, medián 19,0, kvartilové rozdělení: 1. kvartil (Q1): 12, 3. kvartil (Q3): 29. Procentuální úspěšnost v jednotlivých úlohách podle subtestu je uvedena v tab. 3.

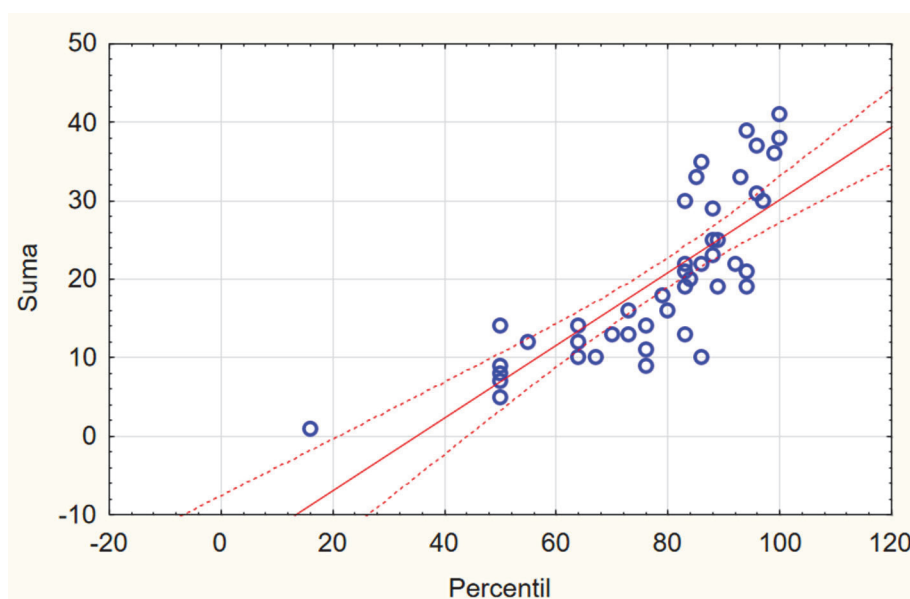
Tab. 3: Procentuální úspěšnost v úlohách subtestů

Subtest	Úloha 1	Úloha 2	Úloha 3	Úloha 4	Úloha 5	Úloha 6
1	96 %	84 %	78 %	64 %	31 %	29 %
2	40 %	58 %	42 %	53 %	27 %	29 %
3	67 %	64 %	40 %	40 %	18 %	18 %
4	82 %	58 %	47 %	41 %	33 %	21 %
5	82 %	41 %	27 %	39 %	22 %	14 %

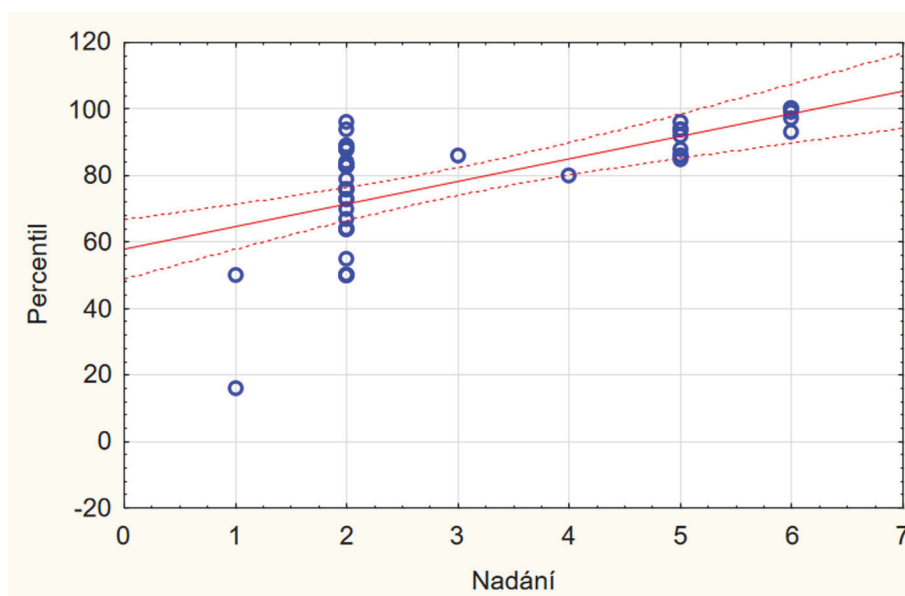
Korelace mezi percentilem, který byl určen standardizovaným testem, a výsledkem ve výzkumném testu byla statisticky významná na hladině významnosti 0,05 (percentil vs. suma bodů: Spearmanovo $R = 0,86$, $t(N - 2) = 11,07$, $p < 0,001$).

Lze tedy konstatovat, že test jako celek (s pěti subtesty a celkem 30 úlohami) stanovuje nadání poměrně spolehlivě. Z grafu na obr. 1 však vidíme skupinu žáků mezi 80. a 95. percentilem, kteří získali v testu relativně málo bodů (8 až 22 bodů z 46 bodů). Vhodným vysvětlením tohoto faktu může být výskyt náročných úloh, které byly i pro nadprůměrné či nadané žáky komplikované a snižovaly jejich úspěšnost.

Také korelace mezi percentilem, který byl stanoven standardizovaným testem, a úrovní nadání, která byla stanovena vyučujícími, byla statisticky významná na hladině významnosti 0,05 (percentil vs. nadání: Spearmanovo $R = 0,72$, $t(N - 2) = 6,87$, $p < 0,001$).



Obr. 1: Suma bodů ve výzkumném testu proti percentilu ve standardizovaném testu



Obr. 2: Percentil ve standardizovaném testu proti nadání podle učitelů

Z grafu na obr. 2 lze za prvé vyčíst, že žáci označení učiteli jako matematicky nadaní získali jako skupina dle očekávání nejvyšší skóre ve standardizovaném testu. Jako matematicky nadaný je obvykle učitelem označen žák, který v běžných školních testech dosahuje výborných výsledků. Protože matematicky nadaní žáci obvykle dobře zvládají i nestandardní úlohy, dalo se očekávat, že dosáhnou vysokého skóre i ve standardizovaném testu, který právě takové úlohy obsahoval. Za druhé, žák s dysortografií označený učitelem jako nadaný má horší výsledky než všichni žáci označení jako nadaní a také než jeden žák označený jako žák s nadprůměrnými výsledky a několik žáků s průměrnými výsledky. Za třetí, řada žáků, kteří byli učiteli označeni jako žáci s průměrnými výsledky, dosahovali ve standardizovaném testu výsledků stejných jako žáci označení jako nadaní.

Lze tedy konstatovat, že posouzení učitelů je v mnoha případech vypovídající, ale zejména u skupiny žáků s průměrnými výsledky vidíme, že na základě projevů žáka v hodinách matematiky je žák vnímán jako průměrný, avšak při řešení problémových úloh dosahuje neočekávaně dobrých výsledků.

Na základě výsledků ve standardizovaném testu TIM³⁻⁵ a ve výzkumném testu bylo ve vzorku identifikováno 20 žáků nadaných.

4.2 Nadaní žáci a výsledky úloh subtestu 1

V tomto oddíle je zjišťováno, do jaké míry lze úlohy v subtestu 1 výzkumného testu využít pro testování nadání.

Tab. 3 ukazuje, že úspěšnost úloh subtestu 1 klesá dle očekávání podle zvyšující se náročnosti úloh. První dvě úlohy jsou velmi jednoduché, žádná úloha nemá úspěšnost menší než 20 %, a nebyla tedy pro žáky příliš náročná. Pro pochopení toho, jak daná úloha testuje a zda je schopna odlišit žáky nadané a ostatní, však bude vhodné podívat se na výsledky v závislosti na matematických schopnostech žáků. Nulová hypotéza *Úspěšnost v dané úloze nezávisí na nadání* nebyla zamítnuta pro úlohy 1, 2, 3 a byla zamítnuta pro úlohy 4, 5, 6 (tab. 4). Lze tedy konstatovat, že v posledních třech úlohách testu úspěšnost žáka v dané úloze závisí na nadání.

Tab. 4: Procentuální úspěšnost úloh subtestu 1 podle nadání

	Úloha 1	Úloha 2	Úloha 3	Úloha 4	Úloha 5	Úloha 6
Nadaní (20 žáků)	100 %	95 %	85 %	95 %	55 %	60 %
Ostatní (25 žáků)	92 %	76 %	72 %	40 %	12 %	4 %
Hladina významnosti rozdílu výsledků (p)	0,076	0,065	0,143	< 0,001	0,003	0,005

Ve všech úlohách jsou nadaní žáci úspěšnější než ostatní žáci. V prvních třech úlohách však rozdíl není podstatný a není statisticky významný. Statisticky významné rozdíly na hladině významnosti 0,05 lze sledovat až u posledních tří úloh. U úloh 5 a 6 byla i u nadaných žáků relativně nižší úspěšnost, a tyto úlohy tedy pro stanovování úrovně nadání příliš vhodné nejsou. U úlohy 5 byl jedním ze zatěžujících faktorů použitý jazyk. Pojem ciferný součet i přes vysvětlení zřejmě činil řadě žáků potíže. Někteří žáci zadání zřejmě porozuměli a prováděli experiment (většinou neřízený), nedokázali z něj však vyvodit správné

závěry a úlohu nechali nedořešenou. U poslední úlohy byl zatěžujícím faktorem distraktor. Žáci, kteří jsou zvyklí používat povrchové strategie a neanalyzují dostatečně situaci, se nechali navést na nesprávný výpočet. Tuto chybu dělali také nadaní žáci, i když ne v tak velkém procentu jako žáci ostatní. V našem vzorku to konkrétně byla čtvrtina žáků nadaných a (necelé) tři čtvrtiny žáků ostatních, kteří provedli chybný výpočet $24 : 3 = 8$.

Nejvhodnější pro stanovení úrovně nadání je z těchto úloh úloha 4 (Pearsonův chí-kvadrát = 14,67, $p < 0,001$). Je to úloha, u níž žáci nemohli mít natrénovaný postup řešení a zároveň si museli vytvořit situační model. Úloha však neobsahovala další jevy, které by byly příliš zatěžující i pro nadané žáky. Rozdíl mezi nadanými a ostatními žáky se tedy nejvíce projevil ve čtvrté úloze, u níž žákům nebyl znám algoritmus, ale současně zadání nesvádělo k použití povrchových strategií řešení.

4.3 Kvalitativní analýza řešení žáků

V tomto oddíle uvádím převládající řešení žáků u jednotlivých úloh a poté časté fenomény, které bylo možné v řešeních žáků sledovat. Analýza se zaměřovala na řešení všech žáků ze vzorku. V případě, kdy u skupiny nadaných žáků převažoval určitý způsob řešení, je tento fakt zmíněn.

Úlohu 1 někteří žáci vyřešili z paměti a zapsali pouze výsledek (v 5 případech), ale nejčastěji zapsali výpočet jako na obr. 3 (ve 38 případech). Vzhledem k jednoduchosti úlohy se neobjevily žádné neočekávané jevy.

Myslím si číslo. Když k němu přičtu 8, dostanu 22. Které číslo si myslím?



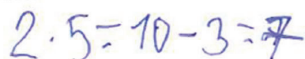
$$22 - 8 = 14$$

Obr. 3: Ukázka řešení úlohy 1

Úlohu 2 stále někteří žáci vyřešili z paměti (17 žáků), ale více žáků postup zapsalo. Nejčastěji při tom postupovali od konce (ve 24 případech), tedy úvahou $7 + 3 = 10$, $10 : 2 = 5$.

Dva kroky, které úloha vyžadovala, vedly některé žáky k implikačnímu zápisu (v 19 případech). To je zápis, ve kterém žák postupně zapisuje své na sebe navazující myšlenky, ale vznikající zápis je matematicky nesprávný. Ukázka implikačního zápisu je na obr. 4. Žák v tomto případě nezapsal ani odpověď.

Myslím si číslo. Když ho vynásobím dvěma a od výsledku odečtu 3, dostanu 7. Které číslo si myslím?

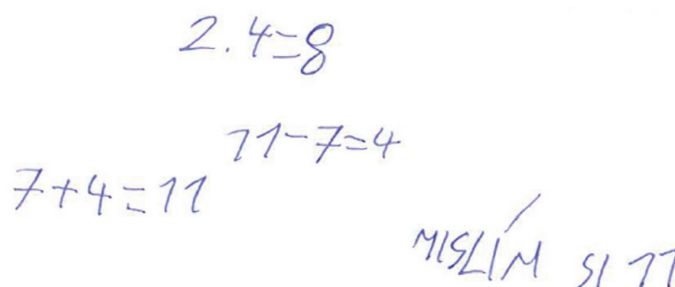


$$2 \cdot 5 = 10 - 3 = 7$$

Obr. 4: Ukázka řešení úlohy 2 s implikačním zápisem

Obdobné postupy byly u úlohy 3, ale ta měla nižší úspěšnost u obou skupin žáků. Z paměti úlohu vyřešilo 17 žáků, metodou od konce 19 žáků. Žáci častěji učinili numerickou chybu (ve 4 případech) nebo zaměnili jednu z operací (ve 4 případech). Ukázku řešení se záměnou operací vidíme na obr. 5. Žák chybně interpretoval obrat „výsledek vydělím dvěma“ a tuto část počítal jako $2 \cdot 4 = 8$. Chybějící zkouška správnosti způsobila, že chyba nebyla odstraněna.

Myslím si číslo. Když od něj odečtu 7 a výsledek vydělím dvěma, dostanu 8. Které číslo si myslím?



$$2 \cdot 4 = 8$$

$$11 - 7 = 4$$

$$7 + 4 = 11$$

Myslím si 11

Obr. 5: Chybné řešení úlohy 3 se záměnou operace

Velmi často se však objevilo řešení, kdy žák určil hledané číslo z paměti a následným zápisem pouze deklaroval správnost svých myšlenek (12 žáků), viz obr. 6.

Myslím si číslo. Když od něj odečtu 7 a výsledek vydělím dvěma, dostanu 8. Které číslo si myslím?

$$23 - 7 = 16 : 2 = 8 \quad 23$$

$$8 \cdot 2 = 16 + 7 = 23$$

Obr. 6: Správně vyřešená úloha s naznačením postupu inverzními operacemi

Implikační zápis byl u třetí úlohy použit v 10 případech.

V úloze 4, která nejlépe odlišovala žáky nadané od ostatních, byly použity nejčastěji tyto postupy: žák úlohu vyřešil z paměti a zapsal buď správný kontrolní výpočet (23 žáků), nebo jen výsledek (4 žáci), žák volil metodu pokusu a omylu (5 žáků), nebo žák nechal úlohu nevyřešenou či použil chybnou úvahu (13 žáků). Třetí a čtvrtou variantu volili nejčastěji žáci nenadaní. Ukázka správného řešení s implikačním zápisem je na obr. 7.

Myslím si číslo. Když ho vynásobím samo sebou a k získanému součinu přičtu myšlené číslo, dostanu 90. Které číslo si myslím?

$$9 \cdot 9 = 81 + 9 = 90 \quad 9$$

Obr. 7: Ukázka implikačního zápisu

Implikační zápis použilo ve čtvrté úloze 17 žáků.

Úloha byla pro řadu žáků, zejména ostatních, náročná z důvodu uvedené slovní formulace, ke které žáci obtížně hledali matematický model. Objevily se problémy s porozuměním tomu, co znamená „vynásobit číslo samo sebou“, kdy žáci často přistupovali ke sčítání dvou stejných čísel nebo k dělení dvěma jako na obr. 8.

Myslím si číslo. Když ho vynásobím samo sebou a k získanému součinu přičtu myšlené číslo, dostanu 90. Které číslo si myslím?

$$90 : 2 = 45$$

$$45 \cdot 2 = 90$$

Myslím si číslo 45.

Obr. 8: Chybné pochopení vztahů ze zadání

Úlohu 5 žáci řešili nejčastěji z paměti (napsali jen výsledek, případně s příslušným výpočtem, 11 žáků), metodou pokusu a omylu (neřízený experiment, 5 žáků), řízeným experimentem (3 žáci), nebo úlohu nechali nevyřešenou (26 žáků). Na obr. 9 je správně vyřešená úloha s implikačním zápisem. Žák své myšlenky zkrátil natolik, že dokonce dal rovnítko mezi číslo a jeho ciferný součet.

Myslím si dvojciferné číslo. Když vytvořím jeho ciferný součet (například ciferný součet čísla 25 je $2+5=7$) a vynásobím ho dvěma, dostanu myšlené číslo. Které číslo si myslím?

$$18 = 1 + 8 = 9 \cdot 2 = 18 \quad 18$$

Obr. 9: Správně vyřešená úloha 5 s implikačním zápisem

Na obr. 10 je ukázka neřízeného experimentu, při kterém žák nedokázal využít získaných výsledků k tomu, aby došel ke správnému závěru.

Myslím si dvojčíferné číslo. Když vytvořím jeho ciferný součet (například ciferný součet čísla 25 je $2+5=7$) a vynásobím ho dvěma, dostanu myšlené číslo. Které číslo si myslím?

22 $2+2=4$ $4 \cdot 2=8$ Myslím si číslo?.

13 $1+3=4$ $4 \cdot 2=8$

15 $1+5=6$ $6 \cdot 2=12$

16 $1+6=7$ $7 \cdot 2=14$

33 $3+3=6$ $6 \cdot 2=12$

99 $9+9=18$ $18 \cdot 2=36$

Obr. 10: Nedořešená úloha 5 metodou pokusu a omylu

U úlohy 6 se jako velmi silný faktor projevil distraktor. Žáci se tak rozdělili na ty, kteří si správné vztahy mezi čísly uvědomili (13 žáků), na ty, kteří se nechali svést distraktorem (28 žáků), a na ty, kteří úlohu neřešili (4 žáci). Pokud žáci použili distraktor, provedli výpočet $24 : 3 = 8$ a jako výsledek uvedli 8 (viz obr. 11). Jak již bylo uvedeno výše, na nesprávný výpočet se zadáním nechala navést čtvrtina žáků nadaných a tři čtvrtiny žáků ostatních.

Myslím si číslo. Když k němu přičtu trojnásobek tohoto čísla, dostanu 24. Které číslo si myslím?

$$24 : 3 = 8$$

8

Obr. 11: Úloha 6 vyřešená nesprávným výpočtem

Na obr. 12 je uvedeno stručné a implikační řešení matematicky nadaného žáka, který výsledek uvádí přímo ve výpočtu.

Myslím si číslo. Když k němu přičtu trojnásobek tohoto čísla, dostanu 24. Které číslo si myslím?

$$\textcircled{6} \cdot 3 = 18 + \textcircled{6} = 24$$

Obr. 12: Správně vyřešená úloha 6 s implikačním zápisem

Žáci mnohdy svůj výsledek zapsali číslem nebo slovní odpovědí, ale už jej nepodrobovali kontrole. Na obr. 13 je příklad řešení, v němž žák navrhl chybný postup řešení, který nekonfrontoval se zadáním, a v důsledku následného neprovedení zkoušky svoji chybu neodhalil.

Myslím si číslo. Když ho vynásobím dvěma a od výsledku odečtu 3, dostanu 7. Které číslo si myslím? = 17

$$2 \cdot 7 = 14 + 3 = \underline{\underline{17}}$$

Obr. 13: Ukázka chybného řešení a neprovedené zkoušky

5 Diskuse

5.1 Identifikace nadání

Zjištění této studie potvrzují, že identifikace matematického nadání ve školním prostředí je komplexní a mnohvrstevnatý proces, který nelze zúžit pouze na výsledky standardizovaných testů či posouzení učitelem (Machů et al., 2013; Szabo et al., 2024). Ačkoliv byla mezi výsledky standardizovaného testu a výkonem ve výzkumném testu nalezena silná korelace, podrobnější analýza ukázala významné individuální odchylky vztahující se k jednotlivým úlohám. Tyto rozdíly poukazují na faktory, které se pojí s testovou úlohou a které mohou výkon nadaných žáků ovlivnit – zejména náročnost úloh a typ zadaných úloh. Ačkoli problémová matematická úloha je chápána jako těžiště nejen testování, ale také vzdělávání nadaných žáků (Liljedahl et al., 2016), i v našem testování se ukázalo, že najít mezi úlohami takovou, která citlivě měří nadání, je složitý úkol. Zatímco úlohy, u nichž žáci již znají postup řešení, problémové nejsou a může je algoritmicky vyřešit i průměrný žák, úlohy s příliš zatěžujícími faktory jsou náročné i pro nadané žáky, kteří takovou úlohu nemusí, například z důvodu nedostatečné zkušenosti, zdárně vyřešit. V našem testování se jako takový nepřiměřeně náročný parametr ukázala přítomnost antisignálu (distraktoru) v 6. úloze, což je v souladu s výsledky výzkumů (Vondrová et al., 2020). Druhým zatěžujícím faktorem, který ovlivňoval výkony i nadaných žáků, byla jazyková stránka úlohy s ciferným součtem.

Významným zjištěním je fakt, že řada žáků s vysokým skóre ve výzkumném testu nebyla učiteli označena jako nadaní. To potvrzuje již dřívější upozornění (např. Hříbková, 2009), že školní hodnocení často odráží spíše výkonnost v rutinních úlohách a přizpůsobivost školnímu systému než skutečný potenciál žáka. Je zde patrný soulad s modelem Bettse a Neihart (1988), který upozorňuje na existenci „skrytých“ či podvýkonných nadaných žáků, jejichž schopnosti se ve školním prostředí nemusí plně projevit.

Dalším důležitým aspektem je samotná povaha použitých matematických úloh. Dle očekávání se ukázalo, že úlohy rutinního typu nejsou pro rozpoznání nadaných vhodné, neboť nezachycují kreativní či alternativní způsoby řešení. Kreativita je jednou z vlastností, kterou se nadaní jedinci odlišují od ostatních (Lithner, 2017; Renzulli, 1978) a kterou lze v matematice odhalovat zadáváním nevšedních problémových úloh, se kterými se jedinec dosud nesetkal. Nadaný žák by se měl v neznámé situaci snáze zorientovat a hledat alternativní způsoby řešení (Krutetskii, 1976). Zároveň však příliš komplexní problémové úlohy – pokud nejsou citlivě vybrány – mohou být natolik náročné, že při jejich řešení selhávají i nadaní žáci. To se potvrdilo i v předloženém výzkumu a je v souladu se zjištěními autorů jako Hadamard (1945), Krutetskii (1976) nebo Lithner (2017), že nadaní žáci potřebují prostor pro hlubší porozumění problému a často přemýšlejí více divergentně. Pokud však není testování navrženo tak, aby tento způsob uvažování zohlednilo (např. časově, strukturou úloh), může být výsledek zkreslený.

Z výsledků dále vyplývá, že pouze jedna z analyzovaných úloh se ukázala jako vhodná pro rozlišení mezi nadanými a ostatními žáky. Byla to právě ta, která nespouštěla automatické řešitelské strategie a zároveň nenutila žáky k povrchovému uvažování. To podporuje myšlenku, že úlohy používané pro identifikaci nadaných by měly být nejen originální a neobvyklé, ale zároveň by i u nadaných žáků měly zohledňovat různé možnosti selhání, které pramení například z malé zkušenosti nadaných žáků s problémově zadávanými úlohami.

5.2 Fenomény v písemných řešeních žáků

Kvalitativní analýza písemného řešení úloh odhalila několik jevů. Velmi častým jevem, a to i u nadaných žáků, byl implikační zápis. Tento typ zápisu ve většině případů nevede k chybě, ale poukazuje na chybné chápání ekvivalence, které může žákovi činit problémy u rovnic (Budínová, 2018). Implikační zápis se objevoval rovnoměrně u nadaných i ostatních žáků. Druhým často sledovaným jevem bylo používání povrchových strategií. Žák dle zadání zvolil určitý postup, ale nepodroboval ho hlubší analýze a nezjišťoval, zda zvolený postup odpovídal zadání. Povrchové strategie byly patrnější u skupiny ostatních žáků.

Třetím fenoménem bylo neprovádění zkoušek správnosti výsledku vzhledem k zadání, což se projevovalo napříč vzorkem a může poukazovat na nerozvinutou metakognici. Tento nedostatek platil pro obě skupiny žáků, avšak s tím rozdílem, že nadaní žáci méně často udělali chybu ve výpočtu či v úvaze, a tak se absence kontroly výsledku neprojevovала tak výrazně, jako u ostatních. Zvýšená metakognice je sice jedním z atributů nadání (Alexander et al., 1995; Machů et al., 2013), jenže i nadaný žák potřebuje dostávat k jejímu rozvoji příležitost. Naopak projev rozvinuté metakognice jsem shledávala v ochotě měnit strategie řešení, přecházet mezi strategiemi dle potřeby a tím zvyšovat možnost úspěchu. To platilo poměrně pro obě skupiny žáků, nadané i ostatní. Když žáci neznali postup řešení, velmi často vybrali metodu pokusu a omylu, aby úlohu vyřešili. Rozdíl mezi řešeními nadaných žáků a ostatních byl spíše v tom, že nadaní žáci používali přímočařejší řešení (leckdy z paměti či aritmeticky), zatímco u ostatních žáků bylo více typické experimentování či zanechání úlohy bez řešení. Přecházení mezi strategiemi se však vyskytovalo u obou skupin. Domnívám se, že to mohlo být proto, že jsou žáci z výuky zvyklí na

zadávaní nestandardních úloh, u kterých neznají algoritmus řešení. Experimentální strategie je ve škole, v níž výzkum probíhal, vnímána jako relevantní možnost řešení.

6 Limity výzkumu

Hlavním limitem výzkumu je malý počet respondentů. Všichni žáci, kteří se výzkumu zúčastnili, navíc navštěvují stejnou základní školu. Učitelé, kteří žáky vyučují, byli tři, tedy alespoň v tomto ohledu byla zachována určitá diverzita. Ovšem charakter školy, která je otevřena vzdělávání nadaných žáků, mohl výsledky testování ovlivnit. Pokud bychom chtěli získat spolehlivější data, bylo by nutné výzkum zopakovat na podstatně větším počtu žáků z různých typů škol.

Dalším z omezení výzkumu bylo rozdělení žáků na nadané a ostatní, které si vyžádalo stanovení jasněho kritéria. Ačkoli bylo toto rozdělení založeno na odborné literatuře, dlouhodobém pozorování žáků a výsledcích standardizovaného testu, je třeba si uvědomit, že pojetí nadání může být různé. Zvolená kritéria tak nemusejí plně zachycovat všechny jeho dimenze a výsledné rozdělení nemusí být zcela vypovídající.

Limitující byl i výběr typu úloh, tedy algebraických a kombinatorických úloh. Ačkoli jsou úlohy z uvedených oblastí školské matematiky vhodné pro rozvoj i testování nadaných žáků z toho úhlu pohledu, že vyžadují prokázat orientaci ve vztazích uvedených v zadání, pro jejich řešení je mnohdy potřebná kreativita a schopnost volit alternativní způsoby řešení, nebo najít všechna řešení, přesto nepokrývají celou škálu dovedností, kterou se nadaní liší od ostatních. Jiné typy úloh by mohly přinést jiné závěry.

7 Závěr

Cílem výzkumu bylo určení vztahu mezi výsledkem v daném výzkumném testu a percentilovým hodnocením získaným na základě jiného standardizovaného testu, a dále mezi výsledkem ve výzkumném testu a posouzením úrovně nadání vyučujícím matematiky. Cílem výzkumu bylo také zjistit, do jaké míry jednotlivé úlohy testu přispívají k identifikaci nadaných žáků a jak dokážou tyto žáky odlišit od jejich vrstevníků.

Výsledky analýz ukazují, že percentil žáků silně souvisí jak s jejich výkonem ve výzkumném testu, tak s učitelským hodnocením matematického nadání. Korelace mezi percentilem a výsledkem testu byla velmi vysoká (Spearmanovo $R = 0,860$; $p < 0,001$), což potvrzuje, že percentil spolehlivě predikuje úspěšnost žáků. Silný vztah byl zjištěn také mezi percentilem a učitelským soudem (Spearmanovo $R = 0,720$; $p < 0,001$), což naznačuje, že učitelé ve velké míře dokáží odhadnout schopnosti žáků na základě jejich výkonu.

Přesto bodové grafy odhalily určité nesoulady. V případě vztahu mezi percentilem a výsledkem ve výzkumném testu byla patrná skupina žáků v rozmezí 80.–95. percentilu, kteří v testu získali pouze 8–22 bodů (z maximálních 46). Tento rozdíl lze pravděpodobně vysvětlit charakterem výzkumného testu, jehož úlohy byly náročné i pro nadané žáky. Podobně se i u korelace mezi percentilem a učitelským soudem ukázala skupina žáků s vysokým percentilem, kteří však byli učiteli hodnoceni jako průměrní. Tyto případy naznačují, že ačkoli je vztah obecně silný, mohou existovat situace, kdy učitelský úsudek či konstrukce testu vedou k odlišným výsledkům.

Oba zjištěné vztahy jsou nejen statisticky, ale i věcně významné, a to i přes omezení vyplývající z velikosti vzorku ($N = 45$). Je proto vhodné ověřit je na větším souboru žáků a současně se zaměřit na faktory, které mohou vysvětlovat zmíněné nesoulady – ať už jde o náročnost úloh či další charakteristiky žáků ovlivňující učitelský úsudek. Tyto výsledky ukazují, že objektivní měření výkonu a učitelský úsudek se vzájemně doplňují, avšak každý z těchto zdrojů může podléhat specifickým omezením.

Pro každou úlohu byla testována hypotéza *Úspěšnost v úloze nezávisí na nadání*. Jako citlivá na testování nadání se ukázala jediná úloha ze sledovaného subtestu, a to byla úloha 4, která neumožnila algoritmické řešení, ale zároveň neobsahovala příliš zatěžující faktory pro nadané žáky. V písemných řešeních žáků jsem se setkávala s povrchovými způsoby řešení, absencí kontroly výsledku a implikačními zápisy. Povrchové přístupy k řešení se týkaly více skupiny žáků ostatních. Chybějící kontrola výsledku platila pro obě skupiny, nadaní žáci měli však výhodu v tom, že méně často chybovali a neprovedení zkoušky se tedy na výsledku neprojevovalo. Implikační zápis se vyskytoval u obou skupin žáků. Na druhou stranu se ve vzorku žáků projevila ochota přecházet mezi strategiemi v případě, že žák nedokázal řešení najít sofistikovanější metodou. To platilo pro obě skupiny žáků, ovšem s tím rozdílem, že ve skupině nadaných byla větší část těch, kteří řešení určili z paměti, a ve skupině ostatních naopak větší část těch, kteří úlohu nechali nevyřešenou.

Vyhledávání matematických talentů v průběhu základní školy je důležité zejména proto, aby se nadaní žáků mohlo včas začít rozvíjet, a to ideálně na úrovni školy, nikoli jen mimo ni. Předložený výzkum znovu

upozornil na to, že identifikace nadání je komplexní proces, do kterého by mělo vstupovat nejen posuzování výkonnosti žáka na základě testů, ale také dlouhodobé posuzování projevů žáka z různých hledisek.

Acknowledgment

Děkuji Marii Budíkové z Přírodovědecké fakulty Masarykovy univerzity, která mi pomáhala při vyhodnocování dat a zasvětila mě do práce s programem Statistica.

Literatura

- Adetula, L. O. (1990). Language factor: Does it affect children's performance on word problems? *Educational Studies in Mathematics*, 21, 351–365.
- Alexander, J. M., Carr, M., & Schwanenflugel, P. J. (1995). Development of metacognition in gifted children: Directions for future research. *Developmental Review*, 15(1), 1–37. <https://doi.org/10.1006/drev.1995.1001>
- Betts, G. T., & Neihart, M. (1988). Profiles of the gifted and talented. *Gifted Child Quarterly*, 32(2), 248–253. <https://doi.org/10.1177/001698628803200202>
- Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. McKay.
- Boonen, A. J., de Koning, B. B., Jolles, J., & Van der Schoot, M. (2016). Word problem solving in contemporary math education: A plea for reading comprehension skills training. *Frontiers in Psychology*, 7, Article 191. <https://doi.org/10.3389/fpsyg.2016.00191>
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (2004). *How people learn*. National Academy Press.
- Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation and understanding* (pp. 65–116). Erlbaum.
- Budínová, I. (2018). *Přístupy nadaných žáků 1. a 2. stupně základní školy k řešení některých typů úloh v matematice*. Muni Press. <https://doi.org/10.5817/CZ.MUNI.M210-9216-2018>
- Budínová, I. (2025). Word problems with Diophantine equations solved by gifted 5th grade elementary school pupils. In J. Novotná & H. Moraová (Eds.), *International symposium elementary mathematics teaching. Elementary mathematics: Building strong foundations* (pp. 86–94). Charles University, Faculty of Education.
- Cígler, H., Jabůrek, M., Straka, H., & Portešová, Š. (2017). *Test pro identifikaci nadaných žáků v matematice TIM³⁻⁵*. Masarykova univerzita. <https://www.nadanedeti.cz/testy-matematicky-test-tim>
- Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology*, 20(4), 405–438. [https://doi.org/10.1016/0010-0285\(88\)90011-4](https://doi.org/10.1016/0010-0285(88)90011-4)
- Forčík, V. & Forčíková, J. (2007). *Nadané dítě a rozvoj jeho schopností*. Portál.
- ČŠI (2022). *Podpora vzdělávání nadaných a mimořádně nadaných žáků v základních a středních školách*. <https://www.csicr.cz/cz/Aktuality/Tematicka-zprava-%E2%80%933-Podpora-vzdelavani-nadanych-a-m>
- Gagné, F. (2005). From gifts to talents. The DMGT as a developmental model. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 98–119). Cambridge University Press. <https://doi.org/10.1017/CBO9780511610455.008>
- Gardner, H. (2006). *Multiple intelligences: New horizons in theory and practice*. Perseus Books.
- Gutiérrez, A., Benedicto, C., Jaime, A., & Arbona, E. (2018). The cognitive demand of a gifted student's answers to geometric pattern problems: Analysis of key moments in a pre-algebra teaching sequence. In F. M. Singer (Ed.), *Mathematical creativity and mathematical giftedness (ICME-13 Monographs)* (pp. 169–198). Springer. https://doi.org/10.1007/978-3-319-73156-8_7
- Hadamard, J. (1945). *An essay on the psychology of invention in the mathematical field*. Dover Publication.
- Havigerová, J. M. (2011). *Pět pohledů na nadání*. Grada.
- Hendl, J. (2012). *Přehled statistických metod: analýza a metaanalýza dat*. Portál.
- Hříbková, L. (2009). *Nadání a nadaný*. Grada.
- Jabůrek, M., Cígler, H., Valešová, T., & Portešová, Š. (2022). What is the basis of teacher judgment of student cognitive abilities and academic achievement and what affects its accuracy? *Contemporary Educational Psychology*, 69, 1–16. <https://doi.org/10.1016/j.cedpsych.2022.102068>
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, 41(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

- Krutetskii, V. A. (1976). *The psychology of mathematical abilities in schoolchildren*. University of Chicago Press.
- Liljedahl, P., Santos-Trigo, M., Malaspina, U., & Bruder, R. (2016). *Problem solving in mathematics education. ICME-13 Topical Surveys*. Springer. https://doi.org/10.1007/978-3-319-40730-2_1
- Lithner, J. (2017). Principles for designing mathematical tasks that enhance imitative and creative reasoning. *ZDM*, 49(6), 937–949. <https://doi.org/10.1007/s11858-017-0867-3>
- Machů, E., Kočvarová, I., Císlerová, T., & Vápeník, P. (2013). *Kvalita školy z hlediska péče o nadané žáky*. Univerzita Tomáše Bati ve Zlíně.
- McBee, M. T., & Makel, M. C. (2019). The quantitative implications of definitions of giftedness. *AERA Open*, 5(1). <https://doi.org/10.1177/2332858419831007>
- Ministerstvo školství, mládeže a tělovýchovy. (2024). *Statistická ročenka školství – výkonové ukazatele 2023/24*. <https://statis.msmt.cz/rocenka/rocenka.asp>
- Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra-word- problem comprehension and its implications for the design of learning environments. *Cognition and Instruction*, 9, 329–389. https://doi.org/10.1207/s1532690xc0904_2
- Portešová, Š. (2011). *Rozumově nadané děti s dyslexií*. Portál.
- Portešová, Š., Poledňová, I., Macek, D., Růžička, M., Straka, O., & Urmanová, J. (2014). *Rozumově nadaní studenti s poruchou učení*. Masarykova univerzita.
- Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan*, 60(3), 180–184.
- Renzulli, J. S. (2005). *Equity, excellence, and economy in a system for identifying students in gifted education: A guidebook* (RM 05208). University of Connecticut, National Research Center on the Gifted and Talented.
- Sternberg, R. J., & Williams, W. M. (2002). *Educational psychology*. Allyn & Bacon.
- Straka, O. (2021). *Jak měřit metakognici (nejen) u nadaných dětí*. Masarykova univerzita. <https://doi.org/10.5817/CZ.MUNI.M210-9905-2021>
- Straker, A. (1980). Identification of mathematically gifted pupils. *Mathematics in School*, 9(4), 4–8.
- Szabo, A., Tillnert, A. S., & Mattsson, J. (2024). Displaying gifted students' mathematical reasoning during problem solving: Challenges and possibilities. *The Mathematics Enthusiast*, 21(1), 179–202. <https://doi.org/10.54870/1551-3440.1623>
- Tannenbaum, A. J., & Baldwin, L. J. (1983). Giftedness and learning disability: A paradoxical combination. In L. H. Fox, L. Brody & D. Tobin (Eds.), *Learning disabled/gifted children: Identification and programming* (pp. 11–36). University Park Press.
- Vondrová, N. (2020). *Didaktika matematiky jako nástroj zvládnutí kritických míst*. Pedagogická fakulta, Univerzita Karlova.
- Vondrová, N., Havlíčková, R., Hirschová, M., Chvát, M., Novotná, J., Páchová, A., Smetáčková, I., Šmejkalová, M., & Tůmová, V. (2019). *Matematická slovní úloha. Mezi matematikou, jazykem a psychologií*. Karolinum.
- Warne, R. T. (2012). History and development of above-level testing of the gifted. *Roeper Review*, 34(3), 183–193. <https://doi.org/10.1080/02783193.2012.686425>

Příloha

Subtest 1: Úlohy typu *Myslím si číslo*

1. Myslím si číslo. Když k němu přičtu 8, dostanu 22. Které číslo si myslím?
2. Myslím si číslo. Když ho vynásobím dvěma a od výsledku odečtu 3, dostanu 7. Které číslo si myslím?
3. Myslím si číslo. Když od něj odečtu 7 a výsledek vydělím dvěma, dostanu 8. Které číslo si myslím?
4. Myslím si číslo. Když ho vynásobím samo sebou a k získanému součinu přičtu myšlené číslo, dostanu 90. Které číslo si myslím?
5. Myslím si dvojčíferné číslo. Když vytvořím jeho ciferný součet (například ciferný součet čísla 25 je $2 + 5 = 7$) a vynásobím ho dvěma, dostanu myšlené číslo. Které číslo si myslím?
6. Myslím si číslo. Když k němu přičtu trojnásobek tohoto čísla, dostanu 24. Které číslo si myslím?

Subtest 2: Dělení celku na nestejně části

1. Součet dvou čísel je 100, jejich rozdíl je 32. O která čísla se jedná?
2. Maminka s tatínkem mají dohromady 70 let. Tatínek je o 4 roky starší než maminka. Kolik let je mamince a kolik tatínkovi?
3. Julča má v levé a pravé kapse dohromady 51 korun. V levé kapse má o 7 korun méně než v pravé kapse. Kolik korun má v pravé kapse?
4. Rozdělte 100 kostek do tří krabic tak, aby v modré krabici bylo dvakrát víc kostek než v červené a v zelené o 20 kostek méně než v červené krabici. Kolik kostek bude v každé krabici?
5. Obvod trojúhelníku je 87 cm. Strana b je o 15 cm kratší než strana a , strana c je o 12 cm delší než strana b . Vypočítej délky jednotlivých stran.
6. 35 litrů benzínu se má rozlít do čtyř kanystrů tak, aby ve třetím kanystru bylo o pět litrů méně než v prvním, ve čtvrtém kanystru o 10 litrů více než ve třetím a v druhém kanystru polovina toho, co v prvním. Kolik litrů benzínu bude v jednotlivých kanystrech?

Subtest 3: Úlohy vedoucí na soustavy rovnic

1. Dvě housky a dva rohlíky stojí dohromady 22 Kč. Dvě housky a čtyři rohlíky stojí dohromady 28 Kč. Urči, kolik korun stojí rohlík a kolik houska.
2. Tatínek kupoval tři autíčka, červené, modré a zelené. Modré stálo dvakrát více než červené, zelené stálo tolik, co červené a modré dohromady. Všechna autíčka stála dohromady 120 Kč. Kolik korun stálo každé autíčko?
3. Sešit a tužka stojí dohromady 39 Kč. Dvě tužky a tři sešity stojí dohromady 102 Kč. Kolik korun stojí sešit a kolik tužka?
4. Lucka chová kočku a pejska. Kočka a pes váží dohromady 24 kg, kočka a Lucka váží dohromady 41 kg, Lucka a pes dohromady 47 kg. Kolik kilogramů váží každý z nich?
5. Všichni členové čtyřčlenné rodiny váží dohromady 176 kg. David a Honza váží dohromady 45 kg, Honza a maminka dohromady 84 kg, Honza a tatínek 99 kg. Kolik kilogramů váží Honza?
6. Dědeček rozdává vnukům oříšky. Kdyby jim dával po 10 oříscích, 4 oříšky mu budou chybět. Kdyby rozdával po 8 oříscích, 6 oříšků mu zbude. Kolik má dědeček vnuků a kolik má oříšků?

Subtest 4: Diofantické rovnice

1. Mirek měl 14 koleček a stavěl z nich autíčka a trojkolky. Kolik autíček a kolik trojkolek mohl vytvořit, když využil všechna kolečka?
2. Kuba měl v pokladničce pouze dvoukorunové a pětikorunové mince. Kolik kterých mincí může mít, jestliže má v pokladničce právě 37 Kč? Pokus se najít všechna řešení.
3. Motýli a pavouci mají dohromady 76 nohou. Kolik může být motýlů a kolik pavouků? (Motýl má 6 nohou a pavouk 8 nohou). Pokus se najít všechna řešení.
4. Laura si ukládá do pokladničky jen dvacetikorunové a padesátikorunové mince. Kolika způsoby může zaplatit 430 Kč?
5. Aleš měl 23 koleček a chtěl stavět autíčka, trojkolky a koloběžky. Chtěl využít všechna kolečka a od každého vozidla chtěl mít alespoň jeden kus. Kolik měl možností, jak vozidla vyrobit?
6. Na farmě chovají koně, kozy a kachny. Dohromady mají 64 hlav. Kachen je dvakrát více než koz, koní je nejméně. Kolik mohou mít koní, koz a kachen?

Subtest 5: Kombinatorika

1. Z číslic 2, 3, 6, 7 zapiš všechna dvojčífná čísla. Číslice se v zápisu čísla nesmí opakovat.
2. V rovině je 5 bodů. Každé dva z těchto bodů spojíme úsečkou. Kolik takových úseček existuje?
3. Pět kamarádů běhalo na plesu v hadovi. Kolika způsoby mohli být v hadovi uspořádání?

4. Kolik existuje čtyřciferných čísel, která jsou sestavena z číslic 1, 3, 4, 6, 0? Číslice se v zápisu čísla nesmí opakovat.
5. Na večírku bylo 12 osob a podali si ruku každý s každým. Kolik podání ruky to bylo?
6. Kolik různých slov (i těch, která nadávají smysl) lze sestavit z písmen slova OPERA?

Do various groups involved in physics education appreciate the same aspects of physics demonstrations? – A methodological approach

✉ Alexandr Nikitin^{1,*}, ✉ Marie Snětinová¹

¹ Faculty of Mathematics and Physics, Charles University, V Holešovičkách 2, 180 00 Prague, Czech Republic; alexandr.nikitin@matfyz.cz

This paper presents a novel methodological approach to examining how different groups involved in upper secondary physics education perceive lecture demonstrations. The research utilizes a video-based, mixed-methods design that integrates high-inference rating scales and open-ended qualitative questions. This captures both holistic and analytical evaluations of demonstration quality. The paper focuses on the psychometric properties of the quantitative instrument and the alignment of ratings across four groups: secondary school students, in-service teachers, pre-service physics teachers, and teacher trainers. Initial findings suggest minimal statistically significant differences in how these groups evaluate physics demonstrations, indicating a potential universality in their perceptions. The methodological approach described in this paper offers a framework for assessing experiment focused lecture demonstrations. By providing insights into perceptions of teaching practices during demonstrations, this paper contributes to improving the design and delivery of physics demonstrations that engage diverse audiences and foster conceptual understanding.

Key words:
physics demonstrations,
mixed methods design,
upper-secondary
education, video-study,
comparative analysis.

Received 12/2024

Revised 8/2025

Accepted 8/2025

1 Introduction

Experiments are fundamental to both physics as a scientific discipline and physics education (Owen et al., toward more student-centred learning approaches, physics demonstrations continue to play a crucial role in educational settings. Previous research highlights both the benefits and limitations of demonstrations in fostering students' interest and understanding. For instance, some studies report that science demonstrations can raise students' interest in a given field (Lin et al., 2013) and increase students' understanding (Basheer et al., 2017), while others suggest that lecture demonstrations have a negligible impact on students' learning outcomes (Crouch et al., 2004). Despite these mixed findings, demonstrations remain amongst the most common forms of experimentation in many physics classrooms (Seidel et al., 2006), making them a valuable subject of research attention.

Our Department of Physics Education has been conducting physics lecture demonstrations (DEMOS) for upper secondary school students (ISCED 3) for over 30 years. Each year, the DEMOS engage approximately 5 000 students. In 2017/2018, a questionnaire-based study using the Intrinsic Motivation Inventory (Ryan, 1982) examined students' perceptions of these demonstrations (Kácovský & Snětinová, 2021). The results revealed significant variation in students' engagement with different physics topics, which led to a follow-up video-study aimed at identifying specific performance aspects that influence these perceptions.

To address these variations, we previously developed a categorical system¹ for analysing lecturer behaviour during demonstrations, focusing on audience interaction and the use of audiovisual technology (Nikitin, 2021). The results revealed that the various topics performed by different lecturers are very diverse regarding the mentioned aspects (Nikitin et al., 2022). However, the comparison with the data from the Intrinsic Motivation Inventory questionnaire did not provide a satisfactory explanation of the variations in students' perception. This quantitative analysis showed some similarities in the interaction patterns during topics that received lower ratings. Ill-perceived performances were generally less interactive, with the interaction being less varied than in the well-perceived ones. Yet some of the DEMOS did not follow this tendency and even contradicted it. This indicates that there are important aspects of these performances that were not sufficiently covered by the conducted quantitative analysis.

In response, this study adopts a new methodological approach to broaden and deepen our analysis of physics demonstrations. We incorporate an expert evaluation method using high-inference rating scales and engage diverse groups (referred to as *raters* in this paper) related to physics education – including secondary school teachers, students, pre-service teachers, and teacher trainers – to assess demonstration quality. This paper presents a detailed description of the newly developed methodology, offering a replicable framework for future research in science education. Identifying aspects of demonstrations that are valued by different stakeholder groups can help pinpoint the various parameters of the performances that

¹Applicable to all sorts of lecturer-centred science shows.

influence students' perception. These insights may support the development of more effective teaching practices in physics education.

The paper is structured as follows. Section 2 reviews prior research on physics lecture demonstrations and introduces the specific demonstrations (DEMOs) studied in this work. Section 3 outlines the overall research design, with a detailed description of the video-based evaluation tool used to capture perceptions of demonstration quality. Section 4 focuses on the methodological approach, including the validation of the tool's psychometric properties and the statistical techniques employed to process the quantitative data. Section 5 presents illustrative results from the comparison of rater groups, highlighting the practical application of the methodology. Finally, Section 6 offers conclusions and discusses the implications of this study for future research and teaching practices in physics education.

The primary focus of this paper is the detailed description and validation of the video-based methodology used to evaluate perceptions of physics demonstrations. Comparative findings from the different rater groups are included to demonstrate the practical application of the method and to illustrate its potential for providing meaningful insights into demonstration quality.

2 Research context

2.1 Lecture demonstrations

Lecture demonstrations are widely used in science education as an engaging instructional tool. Their effectiveness in promoting student learning and conceptual understanding has therefore been extensively researched and debated.

Ample evidence across multiple studies supports the use of lecture demonstrations in facilitating cognitive and affective learning outcomes. Austin and Sullivan (2019), Basheer et al. (2017), and Breckler et al. (2013) all reported that demonstrations led to significant improvements in conceptual understanding compared to the initial understanding, academic achievement, and retention of concepts. Crouch et al. (2004) argued that passive observation of demonstrations has a negligible effect, whereas involving student predictions and interaction enhances learning outcomes. This finding is further supported by other research (Manivannan & Meltzer, 2001; Milner-Bolotin et al., 2007; Thornton & Sokoloff, 1998). Furthermore, the control group (without demonstrations) in the study by Breckler et al. (2013) caught up to the experimental group (with demonstrations) after a month, suggesting only a temporary advantage gained from watching demonstrations.

Di Stefano (1996) and Milne and Otieno (2007) highlighted the positive impact of demonstrations on student engagement, interest, and emotional energy, fostering a conducive learning environment. According to Káčovský and Snětinová (2021), these affective gains were particularly evident among students intending to study physics or who felt competent in the subject.

While demonstrations offer potential benefits, their effectiveness depends on various factors identified across multiple studies. Chang and Shieh (2018), Miller et al. (2013), Neo and Yap (2009), and Roth et al. (1997) emphasized the importance of students' prior knowledge, opportunities for prediction and discussion, and explicit instructional guidance in facilitating meaningful learning from demonstrations.

Various studies proposed structured frameworks and strategies to optimize demonstration implementation. Examples include interactive lecture demonstrations (Sokoloff & Thornton, 1997; Zimrot & Ashkenazi, 2007), the survey-question-experiment-recite-reflect-review approach (Chamely-Wiik et al., 2014), using demonstrations as contextual roadmaps (Buncick et al., 2001), ensuring accurate student observation (Miller, 2013), and limiting demonstration duration for better focus (Walton, 2002).

Despite the positive outcomes reported, some studies revealed contrasting findings or limitations. Odom and Bell (2015) found a negative association between teacher demonstrations and student achievement, suggesting demonstrations alone may be insufficient for developing scientific understanding. Thijs and Bosch (1995) observed no significant differences in learning outcomes between teacher demonstrations and small-group practical work. However, in small-group practicals, girls showed a tendency to underperform compared to boys. Rose (2018) noted challenges in delivering effective live demonstrations and suggested that supplementing them with videos accessible outside class time may be more efficient.

Additionally, Sharma et al. (2010) reported learning gains from *interactive lecture demonstrations* that were lower than previously claimed, although still substantial. This finding emphasizes the need for realistic expectations and further research. Moll and Milner-Bolotin (2009) suggest that *interactive lecture demonstrations* have the potential to improve academic achievement. However, their effectiveness depends on implementation strategies, feedback mechanisms, and alignment with assessment practices.

The studies collectively highlight the potential benefits of demonstrations in enhancing student engagement, visualization, and understanding. At the same time, they also emphasize the importance of instructional design, implementation strategies, and consideration of student characteristics and prior

knowledge. Effective lecture demonstrations should be interactive, encourage accurate observation, and be combined with other active learning strategies. Future research should continue to explore ways to optimize the use of demonstrations and investigate their long-term impact on student learning.

2.2 Physics demonstrations for upper secondary school students

The Department of Physics Education at MFF, Charles University, has a long-standing tradition of organizing *Physics demonstrations for upper secondary school students* (DEMOs) (Káčovský & Snětinová, 2021). These lecture demonstrations consist of a thoughtfully selected series of physics experiments accompanied by theoretical explanations. At the time of this research, DEMOs offered seven specialized monothematic performances: Acoustics, Electricity and magnetism, Ionising radiation, Mechanics, Optics, Thermodynamics and molecular physics, and Electromagnetic waves.

DEMOs are held in a university lecture hall and are led by one or two lecturers. Each session lasts 75 minutes and typically draws 60–90 students from multiple schools. Student participation in the experiments is generally limited, with only a few volunteers directly involved. However, focus on fostering student understanding is strong, as lecturers provide both basic and advanced explanations.

One video recording of each performance within DEMOs was obtained during the 2017/2018 school year. To avoid disrupting the usual flow of the sessions, the recordings were taken from a video room located behind the lecture hall. The camera focused on the lecturers' performance, experimentation and presentation. After each session, the lecturers were asked whether the performance had proceeded normally or if it should be recorded again. This option was used only once, when the participants arrived late and the session would otherwise have been significantly shortened.

2.3 Video as a common tool for educational inquiry

Video-based methodologies are widely used in educational research to analyse teaching practices. These studies typically involve systematic coding of classroom interactions and teacher behaviours within recorded lessons. Coders often assign codes from predefined frameworks (e.g., Roth et al., 2006) or rate behaviours using structured scales, such as the Likert scale (e.g., Dalehefte et al., 2009), to evaluate the intensity of specific actions. An alternative approach involves scales with explicit category definitions for behaviour evaluation, such as the assessment of teachers' curriculum knowledge (Wang et al., 2023). Coding systems vary in their level of interpretation, ranging from high-inference methods, which require subjective judgment, to low-inference methods that rely on more objective classifications (Dalehefte et al., 2009). These methods facilitate the analysis of both the observable and the contextual aspects of teaching. Some studies even examine multiple dimensions of behaviour within the same video segment.

Video-based methodologies have proven effective across various research designs. Seidel and Prenzel (2006) used time sampling to analyse physics lessons and demonstrated stable teaching patterns through a coding framework with high inter-coder reliability. Jewitt (2012) emphasized the capacity of video to capture multimodal classroom interactions, allowing for detailed analysis of gestures, expressions, and speech. However, he also noted challenges such as camera effects and ethical concerns. Zhang et al. (2011) highlighted the benefits of self-produced videos in teacher reflection, contrasting them with published videos that model best practices, although technical limitations remained an obstacle. Similarly, Vondrová and Žalská (2018) found that while pre-service teachers could identify mathematical phenomena in videos, their interpretive skills showed limited progress.

Studies also support the role of video in teacher professional development. Simpson et al. (2018) found that guided observation programs improved teachers' focus on student engagement and reasoning. However, interpretation skills still required further training. Lebak (2023) demonstrated that video-based pedagogical action research fosters systematic reflection, helping teachers address instructional challenges. Together, these findings underscore the potential of video for enhancing teacher awareness and instructional improvement, provided that structured guidance and sustained practice are included. In recognition of this impact, we have incorporated a dedicated video study into our research design.

The methodologies outlined above align closely with the aims of this research, particularly the use of structured scales and coding frameworks to analyse teacher behaviours. Building on these approaches, this study adopts a multi-dimensional analysis that leverages video data to capture both the observable and contextual aspects of teaching. By integrating video reflection into a broader research design, this study aims to address existing gaps in interpretive training and contribute to the refinement of video-based methodologies.

3 Overall research design

From a methodological perspective, the research design can be described as an observational video-study using high-inference rating scales for expert evaluation. The observation in our research is open, non-participatory, partially structured and conducted in a natural setting (although captured on video).

To comprehensively address the aspects that may influence students' perceptions, we evaluated the recordings from two perspectives – as a *whole topic* and as many monothematic *short sections* into which these performances are divided.

The research includes four groups of raters (21 raters each) related to upper-secondary school physics education – upper-secondary school students, upper-secondary school teachers, pre-service physics teachers and physics teacher trainers.

3.1 Assessing whole topic and short sections

One video recording of each topic, approximately 75 mins long, was analysed. These recordings are referred to as *whole topics*, and one unique set of rating scales was developed for their assessment. Each *whole topic* was evaluated by twelve distinct raters.

Furthermore, the recording of each whole topic was divided into several monothematic short videos, typically about fourteen per topic. Here, monothematic refers to a short video centred on a single topic or physics phenomenon, with an average length of about six minutes. These are referred to as *short sections*, and a specific set of scales was developed to assess them.

Of the short sections, 55 were classified as *mainly experimental*, containing only minimal theoretical explanation. An additional 29 sections were identified as *mixed*, combining experiments with theory. Finally, 16 sections were designated as *mainly theoretical*, involving only minimal experimental activity. Each of these sections was assessed by at least nine distinct raters.

3.2 Rating scales

Two sets of rating scales were developed, each tailored to the nature of the performances being studied and to the specifics of evaluating either a *short section* (one set of scales), or a *whole topic* (the other set of scales). The following sections list the scales for both types of recordings, beginning with short sections.

Short section scales:

Atmosphere in the auditorium: Assesses how well the lecturer maintains a focused and engaging atmosphere in the audience, balancing attention, eye contact, humour, and age appropriateness.

Experiment clarity: Evaluates how clearly the lecturer conveys the purpose and results of experiments, making them accessible and logically connected to theoretical points.

Visibility: Measures the visibility of demonstrations and equipment for all audience members, including diagrams or camera use if needed.

Speech clarity: Assesses how understandable the lecturer's explanations are, ensuring that scientific terms and complex ideas are conveyed clearly to a lay audience.

Overall impression of the lecturer's performance: Summarizes the lecturer's effectiveness, confidence, and ability to engage the audience through the presentation.

The designed scales consist of five points, with detailed descriptions provided for the 1st, 3rd, and 5th point. The *Experiment clarity* and *Visibility* scales also contain the point **N**, denoting “not relevant” or “not judgeable” (e.g., evaluating visibility of an experiment in a mainly theoretical section). *Visibility* scale is provided here as an example:

Scale: Visibility

5 Demonstrated or measured phenomena are clearly visible, provably observable. The equipment is sufficiently large and visible to the audience, even from the back rows, or a camera is appropriately used. Diagrams and drawings are sufficiently visible.

3 Some deficiencies reduce the visibility of the experiment – the effect is less noticeable or not equally visible to all viewers.

1 The experiment is poorly visible to some viewers; the equipment is too small.

N Not relevant for this section.

Each scale is accompanied by a written commentary related to the aspects measured. This commentary is mandatory if the rater selects a non-extreme point of the scale (points **2**, **3**, **4** or **N**), in which case the rater is required to justify their choice.

Whole show scales:

Introduction and establishing contact: Evaluates the lecturer's ability to quickly establish rapport with the audience and create a relaxed, interactive environment in the first five minutes.

Interaction with the audience: Evaluates the lecturer's efforts to engage the audience consistently through questions, discussions, and responses to inquiries.

Atmosphere in the auditorium: Measures the overall engagement and atmosphere created by the lecturer during the show.

Utilization of motivation: Assesses the lecturer's use of motivational techniques (like surprise or problem questions) to maintain interest and emphasize relevance of the content.

Overall logical structure: Evaluates the logical coherence and flow of the presentation, particularly the connection between theory and experiments.

Overall subjective impression: Summarizes personal responses to the lecturer's style, engagement level, and perceived value of the performance, using both a rating scale and open-ended questions.

The whole show scales are designed in the same way as the short section scales (see the example of *Visibility* above), except for *Overall subjective impression*.

Unlike the other scales, the *Overall subjective impression* consists of a series of scales and questions intended for two purposes: (a) to estimate the rater's impression of the performance with an emphasis on their subjective perception, and (b) to gather information about the rater's attitude towards experiments and their evaluation strictness. The second goal is achieved through ten statements rated on a four-point Likert-type scales (with the neutral point being purposely omitted). For example:

- I felt immersed in the plot of the performance.
- The lecturer's style of presentation is very interesting to me.
- I enjoyed the performance.

These scales are followed by four open-ended questions:

- The moment that interested me the most during the DEMOs (e.g. specific experiments, presentation methods, explanations, lecturer's reactions, ...)
- I think the performance could be improved by...
- What I appreciate the most about the presenter...
- Is there anything you would like to add to the performance that you could not express in the previous items?

The *Overall subjective impression* concludes the observation sheet and serves as a final reflection point for the raters.

3.3 Raters, rater groups and allotment of videos to raters

Collectively, 84 raters, evenly distributed amongst four different groups of respondents involved in Czech physics instruction, participated in this research. These groups are *in-service physics teachers* (I), *upper-secondary school students* (S), *pre-service physics teachers* (P), and *physics teacher trainers* (T). One of the objectives of this study is to ascertain whether these groups involved in physics education value the same aspects of lecture demonstrations.

Each rater evaluated eleven *short sections* and one *whole topic* using the previously described scales. The short sections were systematically allotted to the raters according to the following rules:

- evaluation of the *whole topic* excluded evaluation of *sections* from that particular performance;
- the distribution of *sections* assigned to each rater reflected the distribution of all sections (6/11 *mainly experimental*, 3/11 *mixed*, and 2/11 *mainly theoretical*);
- thematic diversity was ensured (*sections* originated from several topics of DEMOs);
- there was a sufficient overlap both among raters and among sections.

Raters underwent a training session (with a recording available afterwards), and received codebooks describing the fundamental principles of video evaluation and detailed explanations of the scales used.

Data from *in-service physics teachers* and *upper-secondary school students* was gathered during summer and autumn of 2022, while data from *pre-service physics teachers* and *physics teacher trainers* was gathered during summer and autumn of 2023.

4 Methodological approach

Due to the complex nature of the designed research tool, several methodological considerations need to be addressed. Firstly, we verify that the attitude statements display reasonable psychometric properties, as they are used to improve the objectivity of the respondent's answers. We also examine psychometric properties of the individual responses to both the short section and whole show scales. Finally, we describe the procedure for adjusting rater's responses according to their attitude towards experimenting and present its results.

4.1 Psychometric analysis of attitude statements

As previously mentioned, the *overall subjective impression of the show* covers ten items with 4-point Likert scales (*disagree – rather disagree – rather agree – agree*). These items aim to estimate respondents' strictness and their attitude toward experimenting, as both are important factors that may influence their ratings. Table 1 presents the complete list of these items.

Table 1: Attitude statements used to estimate respondent's attitude towards experimenting

item code	reverse worded	item wording
p1	yes	The lecturer's style of demonstrating experiments and explanations does not sit well with me.
p2		I was intrigued by the performance the whole time.
p3		I consider the time spent watching and evaluating the performance to be meaningfully spent.
p4	yes	I was bored while watching the performance.
p5		I felt immersed in the show.
p6		The lecturer's style of presentation is very interesting to me.
p7		I enjoyed the show.
p8	yes	At times, I lost my attention during the performance.
p9	yes	The performance was not interesting to me.
p10	yes	I did not find watching and judging the performance useful.

Responses to the reverse worded items were recoded so that more positive responses were assigned higher values scaling from 1 to 4. Therefore, the increasing value reflects both a respondent's better attitude towards experimentation and a respondent who is more benevolent in their ratings. We refer to this feature as *respondent's bias* (or simply *bias*).

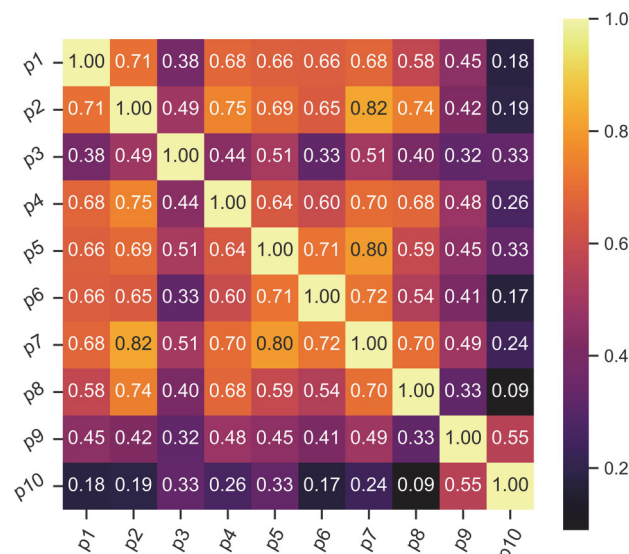
Inspecting parameters of the distribution of responses to these items in Table 2 reveals that the answers are negatively skewed, with responses concentrating around the more positive values of the scales.

Figure 1 shows a heatmap of Pearson's correlations between these items. Items p3, p9 and p10 show weaker correlations with the remaining items. Analysis of reliability (Table 2) confirms this pattern: dropping item p10 would increase the internal consistency of the scale. Subsequently, removing item p10 identifies item p9 for removal and removing item p9 in turn identifies item p3. Eliminating these three items results in a scale exhibiting internal consistency of 0.934 in Cronbach's alpha.

Any further analysis of the attitude statements used to estimate *respondent's bias* is based only on items p1, p2 and p4–p8.

Table 2: Analysis of reliability of attitude statements (marked by their item code p1–p10)

item	mean	SD	min	Q1	Q2	Q3	item-rest correlation	Cronbach's α
scale	3.30	0.652					–	0.914
If item dropped								
p1	3.58	0.746	1	3	4	4	0.747	0.903
p2	3.08	0.903	1	3	3	4	0.832	0.896
p3	3.48	0.683	1	3	4	4	0.537	0.913
p4	3.33	0.892	1	3	4	4	0.789	0.899
p5	3.06	0.943	1	2	3	4	0.809	0.897
p6	3.25	0.898	1	3	3	4	0.714	0.903
p7	3.35	0.841	1	3	4	4	0.864	0.895
p8	2.79	1.114	1	2	3	4	0.694	0.907
p9	3.51	0.826	1	3	4	4	0.557	0.912
p10	3.58	0.762	1	3	4	4	0.319	0.924

**Fig. 1:** Correlation heatmap of the attitude statements (marked by their item code p1–p10)

4.2 Descriptives of the short sections and whole shows scales

All the scales used for both – short sections and whole shows – consist of 5 points with higher values corresponding to more positive ratings. The mean and median (Q2) for the whole show scales in Table 3 suggest that the responses are generally positive. Because the median is already equal to the highest rating on the scales, the 3rd quartile (Q3) and maximum are equal to it and therefore omitted from the table.

Table 3: Whole show scale descriptives and reliability analysis

<i>N</i> = 85	mean	SD	min	Q1	Q2	item-rest correlation	Cronbach's α
scale	4.45	0.666				–	0.845
If item dropped							
introduction	4.40	1.01	1	4	5	0.678	0.811
interaction	4.33	0.97	1	4	5	0.804	0.767
atmosphere	4.39	0.90	2	4	5	0.748	0.785
motivation	4.55	0.72	3	4	5	0.756	0.792
logical structure	4.71	0.55	3	5	5	0.317	0.881

The fact that more than 50% of responses correspond to the highest rating provides evidence of the overall quality of DEMOs. However, it also provides a methodological challenge, as the responses do not allow for clear differentiation among the shows with good ratings.

Regarding reliability analysis of these scales, the *overall logical structure of the show* correlates the least with the remaining scales, and its removal increases Cronbach's α to 0.88. This result is not surprising, as this scale measures qualitatively different and more objective aspect than the others, which mainly capture subjective experiences or perceptions, which can vary widely across individuals based on personal preferences and emotional responses.

The high internal consistency suggests that a combination of the remaining four scales can reasonably be used as a measure of the show's quality. Further inspection of the heatmap in Figure 2 shows that these scales also correlate well with the attitude scales (apart from *overall logical structure of the performance*).

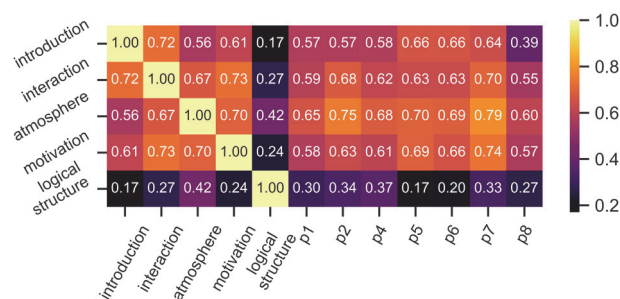


Fig. 2: Correlation heatmap of the whole show scales with the attitude statements p1–p8

Table 4: Short section scales descriptives

	<i>N</i>	<i>mean</i>	<i>SD</i>	<i>min</i>	<i>Q1</i>	<i>Q2</i>
atmosphere	937	4.38	0.92	1	4	5
experiment clarity	787	4.56	0.86	1	4	5
visibility	806	4.34	1.01	1	4	5
speech clarity	937	4.64	0.73	1	5	5
overall impression	937	4.41	0.85	1	4	5

As shown in Table 4, the scales for short sections are skewed towards the positive scale ratings as well. More than 50% of responses represent the highest values of the scales. Since the scales *visibility* and *experiment clarity* allow for the choice *N* (non-relevant), these are treated as “missings” in the quantitative analysis, which explains the lower number of responses.

The correlation matrix in Figure 3 shows that the *short section scales* correlate significantly less at the level of individual responses than the *whole show scales*. This outcome is expected, as the short sections are significantly more varied regarding quality of the assessed aspects. Since the short section scales are never analysed as a single aggregated scale, this does not present a problem.

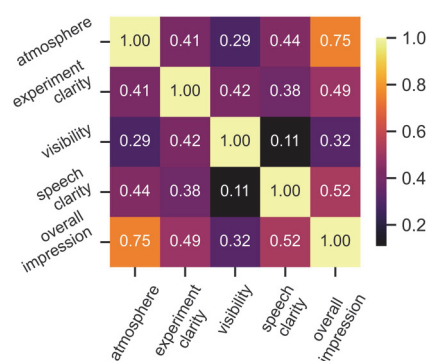


Fig. 3: Short section scales correlation heatmap

4.3 Response level correction

We assume that each respondent strives to be as objective as possible. However, the way they use the scales differs based on their individual *bias*. A person that has better attitude towards experimenting in physics (higher *bias*) tend to be more benevolent in their ratings. Therefore, a correction of the numerical

scale levels is performed separately for each rater and applied to all their responses. This correction was carried out as follows:

- 1) Each respondent was assigned an average score of their responses to the attitude statements – an estimate of their *bias*.
- 2) Each respondent was also assigned an average score of their rating of the *whole show*.²
- 3) For each whole show, a linear regression of the respondent's average rating on their bias was estimated using weighted ordinary least squares with 1 000 bootstraps.
- 4) Each respondent was then assigned a residual from this model. The reasoning is that respondents above the regression line are more positive in their responses than they should be (according to their bias), and thus use the scales more benevolently, while those below the regression line are more critical. The histogram of these residuals shown in Figure 4 reveals that these corrections are relatively small for majority of respondents.
- 5) Each response for each respondent is lowered by this value (responses of those using the scales more benevolently are lowered, and vice versa).
- 6) A scale-index estimate is computed as the mean of these responses with corrected levels, the standard error of the mean is used as estimation error.

This approach considers both the multilevel research design and respondent *bias*.

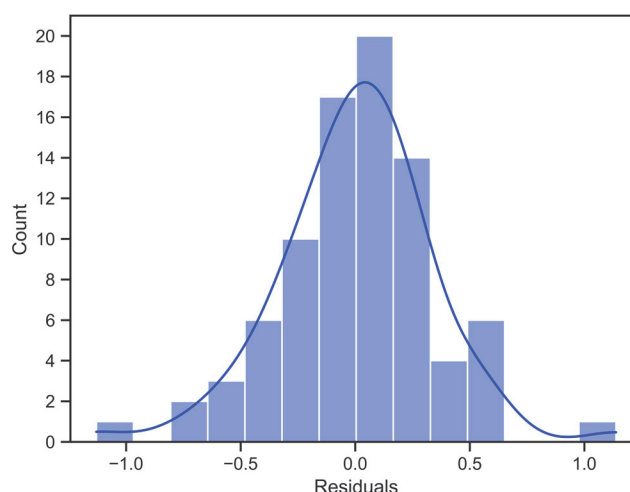


Fig. 4: Histogram of residuals used for response-level correction (with a kernel density estimation)

In simpler terms, if a respondent indicates a strong bias but their responses to *whole show scales* do not reflect it (e.g. they are too benevolent given their bias), their responses are adjusted according to the “general consensus about the whole show” as represented by the linear model. Thus, the correction does not aim to eliminate *respondent's bias* entirely, but rather to estimate how each respondent uses the designed rating scales.

5 Illustrative results – comparison of rater groups

The theory of incomplete block designs (Dey, 2010) is used to compare various rater groups.

In essence, the responses to the individual *short section scales* with performed *response level correction* are analysed using a mixed-effects linear model, with individual short sections as fixed factors³ and individual respondents as random factors.⁴ Estimation of the random factors of this model provides an estimate of each respondents' effect on the ratings of the particular short section scale.

According to Vonesh et al. (1996), R^2 for mixed-effects models can be categorised into two types: marginal R^2 and conditional R^2 . Marginal R^2 relates to the variance explained by the fixed factors, while conditional R^2 relates to the variance explained by both the fixed and random factors (Nakagawa

²Excluding the scale *overall logical structure of the performance* as discussed above.

³*Treatments* in the language of block designs.

⁴*Blocks* in the language of block designs.

et al., 2013). Table 5 states these goodness-of-fit statistics for the mixed-effects models for each scale. The results show that including respondents as random effects in the models significantly improves their explanatory capabilities, with the full models explaining about 90% of variance in the responses. This proves, that the response level correction described in Subchapter 4.3 does not remove the influence of individual raters; rather, it serves as a ‘calibration’ of how the raters interpret the definitions of the rating scales.

Table 5: Goodness-of-fit of the mixed-effects models for individual short section scales

	overall impression	atmosphere	speech clarity	speech clarity	visibility
R^2_{marg}	0.504	0.529	0.423	0.431	0.469
R^2_{cond}	0.923	0.926	0.911	0.898	0.912

Descriptive statistics of the respondent effects estimated from these models are presented in Table 6 for each respondent group separately. The table shows that the means for *in-service teachers* (I), and *physics teacher trainers* (T) are negative, while the means for *upper-secondary students* (S) and *pre-service teachers* (P) are positive. This suggests that I and T tend to be more critical towards the DEMOs than S and P. A brief look at the boxplots of the respondent effects in Figure 5 further shows that the group distributions overlap significantly.

Table 6: Descriptives of the estimated respondent effects on the short scale ratings split by the respondent group (I – in-service physics teachers; P – pre-service physics teachers; S – upper secondary school students; T – physics teacher trainers)

scale	respondent group	mean	SEM	min	Q1	Q2	Q3	max
overall impression	I	-0.03	0.11	-1.50	-0.13	0.02	0.29	0.95
	P	0.11	0.08	-0.67	-0.03	0.14	0.38	0.69
	S	0.06	0.10	-0.68	-0.24	0.06	0.36	0.82
	T	-0.14	0.12	-1.47	-0.41	-0.09	0.08	0.86
atmosphere	I	-0.05	0.11	-1.38	-0.21	0.16	0.34	0.56
	P	0.13	0.08	-0.84	0.00	0.21	0.41	0.57
	S	0.09	0.12	-0.91	-0.21	0.09	0.44	1.13
	T	-0.17	0.14	-1.58	-0.36	-0.06	0.10	1.19
speech clarity	I	-0.11	0.09	-1.14	-0.26	-0.01	0.28	0.43
	P	0.12	0.07	-0.60	-0.06	0.11	0.38	0.70
	S	0.12	0.10	-0.49	-0.30	0.10	0.27	1.15
	T	-0.13	0.09	-1.12	-0.37	-0.07	0.10	0.66
speech clarity	I	-0.10	0.12	-1.71	-0.19	0.00	0.17	0.61
	P	0.12	0.09	-0.95	0.05	0.20	0.33	0.62
	S	0.17	0.11	-0.86	-0.10	0.18	0.62	0.87
	T	-0.20	0.12	-1.32	-0.38	-0.19	-0.01	0.65
visibility	I	-0.04	0.11	-1.25	-0.26	0.03	0.30	0.59
	P	0.21	0.08	-0.74	0.06	0.17	0.40	0.83
	S	0.02	0.10	-1.28	-0.24	0.12	0.33	0.81
	T	-0.18	0.10	-1.28	-0.37	-0.02	0.15	0.47

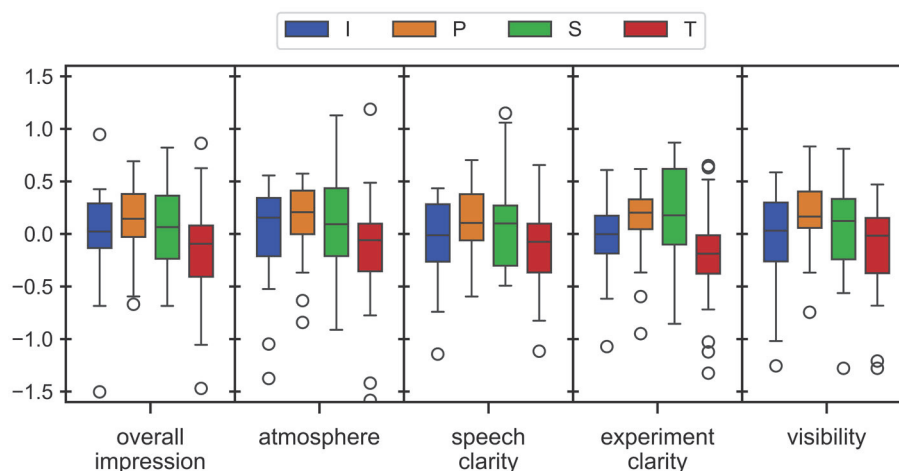


Fig. 5: Boxplots of the estimated respondent effects split by the respondent group (I – in-service physics teachers; P – pre-service physics teachers; S – upper secondary school students; T – physics teacher trainers)

The results of Welch's one-way ANOVAs in Table 7 confirm that the distributions overlap significantly, as only *visibility* shows statistically significant differences ($p < 0.05$) between the respondent groups. According to the Games-Howell post hoc test, the only significant difference in *visibility* is between T and P, with P being more positive in their ratings (mean difference 0.39, Cohen's $d \approx 1.24$).

Table 7: One-way Welch's ANOVA for the effect of respondent groups on individual short section scales

scale	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
overall impression	1.16	3	44.0	0.335
atmosphere	1.34	3	43.8	0.272
speech clarity	2.47	3	44.1	0.074
experiment clarity	2.60	3	44.1	0.064
visibility	3.19	3	44.1	0.033

There are practically no significant differences between the quantitative responses of different rater groups. Although *physics teacher trainers* appear somewhat more critical, as one might expect, and *pre-service teachers* somewhat more benevolent, these differences are statistically insignificant, with the sole exception of *visibility*.

6 Conclusions

Our research methodology is unique in several key aspects:

1. Comprehensive performance evaluation:

Unlike approaches that focus solely on selected key moments, we assess entire performances by systematically segmenting videos and evaluating each segment without exception. This eliminates the potential bias caused by selectively choosing specific parts for analysis.

2. Systematic segmentation and consistency:

The videos are divided into segments according to a systematic framework, ensuring uniformity and objectivity in segment selection. Every segment is assessed using the same predefined scales, regardless of its content. This standardization allows for direct comparisons between different parts of the video and ensures consistency in data collection.

3. Detailed and explicit evaluation scales:

Our five-point scales include detailed descriptions for three anchor points, expressed in several sentences rather than just a few words. This level of detail reduces ambiguity and makes the scales more explicit compared to less defined metrics or those relying heavily on subjective coder interpretations.

4. Qualitative insights:

Raters are encouraged to supplement numerical scores with written justifications. This allows for a deeper reflection and adds qualitative depth to the data collected, extending beyond purely numerical analysis.

5. Emphasis on systematic, detailed, and standardized assessment:

Our approach prioritizes a methodical and highly detailed evaluation process, creating a standardized framework that promotes more objective and accurate assessments of video performances. It should be noted, however, that this approach may require greater time investment and coordination from raters.

Finally, it is worth noting that our study does not rely on professional perspectives, as even the students participating in the research are not experts. This allows for a broader applicability of our methodology beyond expert-level evaluations, making it accessible and relevant to a wider audience.

This novel methodology distinguishes itself from existing approaches by offering greater accuracy and objectivity in video analysis for educational research, though it requires a higher investment of resources. Our findings indicate that the proposed methodology, together with the evaluation tool for assessing lecture demonstrations, successfully integrates quantitative and qualitative approaches. Initial analysis of the quantitative data confirms the appropriateness of the selected rating scales, demonstrating satisfactory psychometric properties. However, the validity of the qualitative component remains to be examined in further research.

The findings also suggest that various groups of raters (students, physics teachers, pre-service physics teachers, and physics teacher trainers) generally value similar aspects of DEMOs. This points to a degree of consistency in preferences across different roles in physics education. While teacher trainers tend to be more critical in their assessments than students, the differences between the groups are statistically insignificant in most cases. The only significant differences appeared in the visibility of DEMOs, where pre-service teachers were more positive in their ratings compared to teacher trainers.

The minimal differences observed across rater groups likely reflect a shared understanding of what makes lecture demonstrations effective – qualities such as clarity, visibility, and conceptual relevance are widely appreciated in physics education, regardless of role or experience level. This alignment may result from common educational experiences and professional norms that emphasize these features. While teacher trainers tend to be slightly more critical, likely due to their greater pedagogical expertise, the overall consistency indicates that effective demonstrations share broadly recognized characteristics.

To interpret this cautiously, analysing raters' open-ended comments (not covered in this paper) could help explain why more experienced individuals in physics education tend to be stricter. For instance, teacher trainers may detect physics inaccuracies that upper secondary students miss, or they may focus more strongly on evaluating the lecturer's pedagogical content knowledge.

Looking ahead, we aim to investigate how these aspects influence students' perceptions of physics demonstrations by linking the collected data with previously published results from the Intrinsic Motivation Inventory questionnaire (Kácovský & Snětinová, 2021). This approach will help identify which specific parameters of demonstrations contribute most to positive student evaluations, thereby enabling us to optimize the performance of DEMOs.

Acknowledgment

The present work was supported by the Specific University Research Project SVV 260 828 and by Charles University in Prague, project GA UK No. 62223.

References

- Austin, S. R. P., & Sullivan, M. (2019). How are we performing? Evidence for the value of science shows. *International Journal of Science Education, Part B*, 9(1), 1–12. <https://doi.org/10.1080/21548455.2018.1532620>
- Basheer, A., Hugerat, M., Kortam, N., & Hofstein, A. (2017). The effectiveness of teachers' use of demonstrations for enhancing students' understanding of and attitudes to learning the oxidation-reduction concept. *Eurasia Journal of Mathematics Science and Technology Education*, 13(3), 555–570. <https://doi.org/10.12973/eurasia.2017.00632a>
- Breckler, J. L., Christensen, T., & Sun, W. (2013). Using a physics experiment in a lecture setting to engage biology students with the concepts of poiseuille's law. *CBE—Life Sciences Education*, 12(2), 262–273. <https://doi.org/10.1187/cbe.12-08-0129>
- Buncick, M. C., Betts, P. G., & Horgan, D. D. (2001). Using demonstrations as a contextual road map: Enhancing course continuity and promoting active engagement in introductory college physics. *International Journal of Science Education*, 23(12), 1237–1255. <https://doi.org/10.1080/09500690010025030>
- Chamely-Wiik, D. M., Haky, J. E., Louda, D. W., & Romance, N. (2014). SQER3: An instructional framework for using scientific inquiry to design classroom demonstrations. *Journal of Chemical Education*, 91(3), 329–335. <https://doi.org/10.1021/ed300689n>
- Chang, W., & Shieh, R. S. (2018). A study of the conceptual comprehension of electric circuits that engineer freshmen display. *European Journal of Physics*, 39(4), 045705. <https://doi.org/10.1088/1361-6404/aab6e1>
- Crouch, C., Fagen, A. P., Callan, J. P., & Mazur, E. (2004). Classroom demonstrations: Learning tools or entertainment? *American Journal of Physics*, 72(6), 835–838. <https://doi.org/10.1119/1.1707018>
- Dalehefte, I. M., Rimmele, R., Prenzel, M., Seidel, T., Labudde, P., & Herweg, C. (2009). Observing instruction “next-door”: a video study about science teaching and learning in Germany and Switzerland. In T. Janík & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 83–99). Waxmann.
- Dey, A. (2010). *Incomplete block designs*. World Scientific.
- Di Stefano, R. (1996). Preliminary IUPP results: Student reactions to in-class demonstrations and to the presentation of coherent themes. *American Journal of Physics*, 64(1), 58–68. <https://doi.org/10.1119/1.18293>
- Jewitt, C. (2012). *An introduction to using video for research (National Centre for Research Methods Working Paper 03/12)*. Institute of Education.

- Káčovský, P., & Snětinová, M. (2021). Physics demonstrations: who are the students appreciating them? *International Journal of Science Education*, 43(4), 529–551. <https://doi.org/10.1080/09500693.2020.1871526>
- Lebak, K. A. (2023). Utilizing video-based pedagogical action research to transform teacher practice in elementary, middle and high school classrooms. *Journal of Inquiry and Action in Education*, 11(2). Retrieved from <https://digitalcommons.buffalostate.edu/jiae/vol11/iss2/2>
- Lin, H.-s., Hong, Z.-R., & Chen, Y.-C. (2013). Exploring the development of college students' situational interest in learning science. *International Journal of Science Education*, 35(13), 2152–2173. <https://doi.org/10.1080/09500693.2013.818261>
- Manivannan, K., & Meltzer, D. (2001). *Use of in-class physics demonstrations in highly interactive format*. Paper presented at Physics education research conference 2001, Rochester, New York. <https://doi.org/10.1119/perc.2001.pr.011>
- Miller, K. (2013). Use demonstrations to teach, not just entertain. *The Physics Teacher*, 51(9), 570–571. <https://doi.org/10.1119/1.4830081>
- Miller, K., Lasry, N., Chu, K., & Mazur, E. (2013). Role of physics lecture demonstrations in conceptual learning. *Physical Review Special Topics – Physics Education Research*, 9(2), 020113. <https://doi.org/10.1103/PhysRevSTPER.9.020113>
- Milne, C., & Otieno, T. (2007). Understanding engagement: Science demonstrations and emotional energy. *Science Education*, 91(4), 523–553. <https://doi.org/10.1002/sce.20203>
- Milner-Bolotin, M., Kotlicki, A., & Rieger, G. (2007). Can students learn from lecture demonstrations? *Journal of College Science Teaching*, 36(4), 45–49. <https://www.jstor.org/stable/42992942>
- Moll, R. F., & Milner-Bolotin, M. (2009). The effect of interactive lecture experiments on student academic achievement and attitudes towards physics. *Canadian Journal of Physics*, 87(8), 917–924. <https://doi.org/10.1139/P09-048>
- Nakagawa, S., Schielzeth, H., & O'Hara, R. B. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in ecology and evolution*, 4(2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Neo, C. S., & Yap, K. C. (2009). The effect of classroom demonstrations based on conceptual change instruction on students' understanding of electromagnetism and electromagnetic induction. In M. Kim, S. W. Hwang & A.-L. Tan (Eds.), *Proceedings of the International Science Education Conference 2009* (pp. 1346–1386). Singapore: National Institute of Education.
- Nikitin, A. (2021). *Fyzikální pokusy pro střední školy – videostudie* [Masters, Charles University]. <http://hdl.handle.net/20.500.11956/127719>
- Nikitin, A., Káčovský, P., & Snětinová, M. (2022). Looking for parameters of students-lecturer interaction that influence how students perceive physics demonstrations. In J. Ondruška, L. Valovičová & L. Zelenický (Eds.), *Didactic Transfer of Physics Knowledge Through Distance Education, AIP Conference Proceedings*, 2458(1). AIP Publishing LLC. <https://doi.org/10.1063/5.0078259>
- Odom, A. L., & Bell, C. V. (2015). Associations of middle school student science achievement and attitudes about science with student-reported frequency of teacher lecture demonstrations and student-centered learning. *International Journal of Environmental and Science Education*, 10(1), 87–97.
- Owen, S., Dickson, D., Stanisstreet, M., & Boyes, E. (2008). Teaching physics: Students' attitudes towards different learning activities. *Research in Science & Technological Education*, 26(2), 113–128. <https://doi.org/10.1080/02635140802036734>
- Rose, T. M. (2018). Lessons learned using a demonstration in a large classroom of pharmacy students. *American Journal of Pharmaceutical Education*, 82(9), 6413. <https://doi.org/10.5688/ajpe6413>
- Roth, W.-M., McRobbie, C. J., Lucas, K. B., & Boutonné, S. (1997). Why may students fail to learn from demonstrations? A social practice perspective on learning in physics. *Journal of Research in Science Teaching*, 34(5), 509–533. [https://doi.org/10.1002/\(SICI\)1098-2736\(199705\)34:5<509::AID-TEA6>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1098-2736(199705)34:5<509::AID-TEA6>3.0.CO;2-U)
- Roth, K. J., Druker, S. L., Garnier, H. E., Lemmens, M., Chen, C., Kawanaka, T., Rasmussen, D., Trubacova, S., Warvi, D., Okamoto, Y., Stigler, J., & Gallimore, R. (2006). *Teaching science in five countries: Results from the TIMSS 1999 video study. Statistical analysis report. NCES 2006-011*. ED Pubs.
- Ryan, R. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43(3), 450–461. <https://doi.org/10.1037//0022-3514.43.3.450>
- Seidel, T., & Prenzel, M. (2006). Stability of teaching patterns in physics instruction: Findings from a video study. *Learning and Instruction*, 16(3), 228–240. <https://doi.org/10.1016/j.learninstruc.2006.03.002>

- Seidel, T., Prenzel, M., Rimmel, R., Dalehefte, I. M., Herweg, C., Kobarg, M., & Schwindt, K. (2006). Blicke auf den physikunterricht. Ergebnisse der IPN videostudie. *Zeitschrift für Pädagogik*, 52(6), 799–821.
- Sharma, M., Johnston, I., Helen, J., Varvell, K., Gordon, R., Andrew, H., Stewart, C., Ian, C., & Thornton, R. (2010). Use of interactive lecture demonstrations: A ten year study. *Physical Review Special Topics. Physics Education Research*, 6, 020119. <https://doi.org/10.1103/PhysRevSTPER.6.020119>
- Simpson, A., Vondrová, N., & Žalská, J. (2018). Sources of shifts in pre-service teachers' patterns of attention: The roles of teaching experience and of observational experience. *Journal of Mathematics Teacher Education*, 21(6), 607–630. <https://doi.org/10.1007/s10857-017-9370-6>
- Sokoloff, D. R., & Thornton, R. K. (1997). Using interactive lecture demonstrations to create an active learning environment. *The Physics Teacher*, 35(6), 340–347. <https://doi.org/10.1119/1.2344715>
- Thijs, G. D., & Bosch, G. M. (1995). Cognitive effects of science experiments focusing on students' preconceptions of force: a comparison of demonstrations and small-group practicals. *International Journal of Science Education*, 17(3), 311–323. <https://doi.org/10.1080/0950069950170304>
- Thornton, R. K., & Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics*, 66(4), 338–352. <https://doi.org/10.1119/1.18863>
- Vondrová, N., & Žalská, J. (2018). Ability to notice mathematics specific phenomena: What exactly do student teachers attend to? *Orbis scholae*, 9(2), 77–101. <https://doi.org/10.14712/23363177.2015.81>
- Vonesh, E. F., Chinchilli, V. M., & Pu, K. (1996). Goodness-of-Fit in generalized nonlinear mixed-effects models. *Biometrics*, 52(2), 572–587. <https://doi.org/10.2307/2532896>
- Walton, P. H. (2002). On the use of chemical demonstrations in lectures. *University Chemistry Education*, 6(1), 22–27.
- Wang, J., Wang, Y., Wipfli, K., Thacker, B., & Hart, S. (2023). Investigating learning assistants' use of questioning in online courses about introductory physics. *Physical Review Physics Education Research*, 19(1), 010113. <https://doi.org/10.1103/PhysRevPhysEducRes.19.010113>
- Zhang, M., Lundeborg, M., Koehler, M. J., & Eberhardt, J. (2011). Understanding affordances and challenges of three types of video for teacher professional development. *Teaching and Teacher Education*, 27(2), 454–462. <https://doi.org/10.1016/j.tate.2010.09.015>
- Zimrot, R., & Ashkenazi, G. (2007). Interactive Lecture Demonstrations: a tool for exploring and enhancing conceptual change. *Chemistry Education Research and Practice*, 8(2), 197–211. <https://doi.org/10.1039/B6RP90030E>

The development of self-assessment accuracy in mathematical word problems: A study of primary school learners

 Irena Smetáčková¹,  Anna Páchová^{1,*}

¹ Faculty of Education, Charles University, Magdalény Rettigová 4, Prague, Czech Republic; Anna.pachova@pedf.cuni.cz

Self-assessment, as a critical component of self-regulated learning, plays a significant role in improving learning achievement. This study focuses on the accuracy of self-assessment in the context of mathematical word problems among primary school students. The evolution of self-assessment accuracy was examined between the second and third grades. The impact of cognitive abilities, mathematical knowledge, and teacher evaluations was investigated. Data were collected from a sample of 542 primary school pupils. Our findings reveal that students demonstrated greater accuracy in self-assessing their performance in mathematical word problems in Year 2 compared to Year 3. This self-assessment accuracy varied based on cognitive abilities and the pedagogical environments. These findings suggest the importance of creating adequate classroom conditions to enhance self-assessment accuracy.

Key words:

self-assessment, learning, metacognition, self-regulation, primary school, mathematics.

Received 9/2025

Revised 11/2025

Accepted 12/2025

1 Introduction

The learning process is a very complex cognitive process. In an effort to enhance the effectiveness of learning, scientific attention has turned to self-regulation in recent decades (Panadero, 2017). Zimmerman's Cyclical Phases Model of self-regulated learning (2000) comprises three phases: forethought, performance, and self-reflection. In the self-reflection phase, students evaluate their task performance and determine their success or failure. This assessment then influences their approach and performance in future tasks.

From the above, it follows that self-assessment is a crucial component of self-regulated learning. Moreover, it is hypothesised, that the relationship between self-regulated learning, metacognition and self-efficacy is mediated by self-assessment (Cera et al., 2013; Schraw et al., 2006). Self-assessment also plays a crucial role in supporting students' learning by actively involving them in assessing their own progress and improve their performance (Ross, 2006). However, self-assessment supports learning and contributes to the development of both metacognition and self-efficacy only when students are able to evaluate their performance accurately – judging weak performance negatively and strong performance positively (Barana et al., 2022; Brown & Harris, 2014).

Although numerous studies have examined self-assessment accuracy (Barana et al., 2022; Brown et al., 2015; Bradshaw, 2001), the existing evidence is heavily concentrated at the secondary and tertiary educational levels. Far less is known about how young learners in elementary school monitor their performance, despite the fact that metacognitive and self-regulatory abilities begin to develop early and show substantial variation at this stage. Existing research with younger children is fragmented and provides inconclusive evidence about how accurately they can judge their own performance. In this article, we are particularly interested in investigating self-assessment in the context of solving word problems among primary school pupils, as this is one of the most challenging aspects of the mathematics curriculum.

1.1 Self-assessment and learning process

Self-assessment is the evaluation of one's own performance, knowledge, or skills (Barana et al., 2022). It involves comparing one's actual performance against a standard or set goals. Klenowski (1995) defined self-assessment as “the evaluation or judgment of ‘the worth’ of one's performance and the identification of one's strengths and weaknesses with a view to improving one's learning outcomes” (p. 146). Andrade (2019) emphasizes the significance of the assessment process and its role as feedback within the framework of formative assessment. Self-assessment is a process focused on evaluating learning outcomes to provide learners with insights into their strengths and weaknesses. The main goal of self-assessment is to promote self-awareness, helping learners gauge their progress, identify areas for development, and set objectives for improving their performance.

Self-assessment is intrinsically linked to self-regulated learning, particularly as a crucial component of the self-reflection phase in Zimmerman's model (2000). With respect to Vygotsky's tradition of mediated learning (Taber, 2025), based on Bandura's social cognitive theory (2014) and experimental paradigms of cognitive psychology (de Bruin & van Gog, 2012; Andrade, 2019), learning is considered to be the result of the interaction of personal, environmental and behavioral factors and thus efficiency of learning increases due to self-regulation. A self-regulated learner is able to set goals and choose appropriate

strategies to achieve them, make effort, monitor learning progress, evaluate achievement, and reflect on their emotional reactions (Pintrich, 2000; Siegesmund, 2017; Zimmerman & Schunk, 2011). As shown by Zimmerman (2000) and others, self-regulated learning involves a number of interacting cognitive, metacognitive (Bakar & Ismail, 2020; Georgiades, 2004), and motivational components, including self-efficacy (Bandura, 2014; Liu-Ambrose et al., 2010).

There are many theories exploring the relationships between these concepts. There is undoubtedly a strong connection between self-assessment and metacognition. Self-assessment involves monitoring and reflecting on one's own work, which are essentially metacognitive processes (Yan, 2020). From this perspective, self-assessment can be considered as part of metacognition. If students have well-developed metacognition and self-assessment, they are able to monitor and evaluate their performance effectively. However, causality can also operate in the opposite direction, where the source of influence determines the relationship to ability or performance. For example, Siegesmund (2017) argues that self-assessment can strengthen students' metacognition and positively influence their self-regulation. At the same time, if students improve the accuracy of their self-assessment, this contributes to the development of metacognition and self-regulation. Thus, self-assessment serves as both a prerequisite and a consequence.

Other studies have also highlighted the relationship between self-assessment and self-efficacy. One explanation lies in the connection through self-regulated learning models (Pintrich, 2000; Zimmerman, 2000). Accurate self-assessment helps learners evaluate their performance, leading to better self-regulation and higher self-efficacy through successful learning experiences. Panadero et al. (2017) demonstrated this by analyzing 19 studies involving 2 305 students, revealing the importance of self-assessment interventions for the development of self-efficacy and self-esteem.

From the discussion above, it is evident that self-assessment is a crucial concept that is strongly connected with self-regulated learning. Moreover, improving self-assessment can enhance related concepts such as metacognition and self-efficacy.

1.2 Accuracy¹ of self-assessment

However, self-assessment aids the learning process and the only if students can accurately evaluate their performance. This means that students should recognize poor performance as such and assess it negatively, and similarly, recognize and assess strong performance positively (Barana et al., 2022).

Numerous studies have shown that learners often make inaccurate self-assessments, often overestimating or underestimating their own abilities. Psychologists refer to this as the Dunning-Kruger effect (Kruger & Dunning, 1999). Moreover, this effect is often more pronounced in lower-achieving students who have difficulty recognizing their poor performance, its causes, and how to improve (Dunning et al., 2004). Inaccurate self-assessment acts as a barrier to self-regulated learning and contributes to poor academic performance.

The accuracy of self-assessment may be influenced by cognitive, motivational, emotional factors, individual characteristics, item characteristics and classroom environment characteristics (Brown et al., 2015). Reduced accuracy in self-assessment can arise from learners' tendencies to be unrealistically optimistic about their abilities, deficiencies in skills and abilities, unclear assessment criteria, and classroom norms that encourage overestimation (Dunning et al., 2004). Highly difficult tasks introduce greater uncertainty into self-assessment, whereas easier tasks allow pupils to judge their performance with higher confidence (Bradshaw, 2001). This means that the difficulty of the task correlates with the accuracy of self-assessment, regardless of individual metacognitive competence (Barana, 2022). In the case of word problems, assessing one's own performance is even more complex, as accuracy depends on the combination of mathematical competence and reading comprehension, which interact in non-trivial ways (Schleppegrell, 2007).

Accuracy in self-assessment shows clear developmental patterns (Bradshaw, 2001). Brown and Harris (2014) found that accuracy varies with both age and performance level, with younger and less proficient elementary pupils displaying lower accuracy. Age therefore plays a central role. Young children's cognitive abilities—particularly working memory, inhibition, and self-regulation—are still developing (Zelazo & Carlson, 2012), which limits their capacity to evaluate their performance accurately. They also have less mature emotional regulation and awareness, making it more difficult for them to reflect honestly, whereas adolescents, with more advanced cognitive and emotional capacities, tend to produce more accurate self-assessments (Harter, 2011).

According to Harter (2011), average levels of self-evaluation, and consequently self-assessment, tend to decline from early childhood into adolescence as children develop social comparison skills and adopt

¹Some authors use the accuracy of self-assessment, others the consistency (Andrade, 2019). We stay in the term accuracy because we compare whether the student's performance corresponds with his/her self-report on it. The consistency is usually used in the wider context of comparison with teacher's assessments.

a more critical view of themselves. Piagetian theory further suggests that younger children's moral and evaluative reasoning is heteronomous—anchored in authority and rules—which leads them to perceive self-assessment more as obedience to teacher expectations than as autonomous reflection. Children also rely heavily on teachers' judgments as their primary source of evaluative feedback (Eelder, 2010). With increasing age, pupils develop a more differentiated understanding of their abilities and a stronger capacity for self-regulation, while younger children often display simpler and overly optimistic self-evaluations. However, some longitudinal evidence, such as that from Orth et al. (2021), challenges the notion of a sharp transition in self-efficacy or self-esteem during early adolescence. More longitudinal research—both short- and long-term—is therefore needed.

Bradshaw (2001) also reports associations between self-assessment and cognitive ability, though this relationship is not linear. Accuracy in self-assessment appears to depend on an interplay of cognitive abilities, metacognitive skills, domain-specific knowledge, and task complexity. Importantly, intervention research (Kajamies et al., 2010) shows that training and feedback can substantially improve accuracy, even among individuals with varying cognitive profiles.

Teachers play a critical role in shaping pupils' learning, including their self-assessment practices. They influence students not only through positive expectations and instructional support but also through the nature of learning challenges and classroom interactions. Li and Rubie-Davies (2015) [24] demonstrate that teacher optimism can translate into measurable gains in academic progress. Certain pedagogical practices—especially formative assessment—can enhance pupils' self-assessment and self-regulation (Nicol & Macfarlane-Dick, 2006). Factors supporting the development of self-regulated learners include the use of multiple representations, rich mathematical tasks, productive classroom discourse, scaffolding of strategic behaviour, and adaptive instructional support (Pape et al., 2003). Kramarski and Revach (2009) further argue that teachers need to be capable of self-regulation themselves in order to foster student-centred learning, a condition that has been linked to improved self-regulation among pupils (Perry et al., 2006).

1.3 The relationship between self-assessment and performance in mathematics

In this section, we examine the relationship between self-assessment and performance among elementary school pupils. Self-assessment has received growing attention in both learning theory and educational practice, as it represents a reflective process through which pupils evaluate their work against explicit criteria and identify opportunities for improvement. As a formative practice, it has been shown to enhance learning, support pupils in managing academic demands, and contribute to school success while reducing the risk of school failure (Broadfoot, 2021; Andrade, 2019). Because empirical studies focusing specifically on younger learners remain relatively scarce, we also draw on findings from research with older students.

Mathematical word problems offer a particularly meaningful context for studying self-assessment. They require pupils to engage in higher-order thinking and complex problem-solving, draw on multidisciplinary knowledge, apply mathematical concepts to real-world situations, and reveal conceptual understanding rather than procedural recall (Rendl et al., 2013; Kajamies et al., 2010; Pape et al., 2003). Because solving word problems places considerable demands on metacognitive monitoring and self-efficacy, self-assessment may serve as an important mechanism supporting pupils' performance in this domain.

Findings from Czech research on young learners provide additional support for this argument. Chytrý, Říčan and Živná (2019) showed that metacognitive skills are significantly associated with mathematical performance across different educational approaches and that pupils with more developed metacognition tend to achieve higher mathematical outcomes. The relationship also appears bidirectional: engaging in word problems can strengthen metacognitive skills and self-efficacy, while these skills subsequently support more effective problem solving. In a qualitative study, Tachie (2019) demonstrated that the use of metacognitive strategies—including task analysis, planning, monitoring, checking, reflection, and self-assessment—helped 8th- and 9th-grade learners solve mathematical problems.

Findings from younger pupils also support this relationship. Research with 311 primary school students showed that those who successfully solved mathematical problems demonstrated higher levels of prediction and self-assessment accuracy than less successful pupils (Nováková, 2024). Consistent with this, the study by Chytrý et al. (2019) indicates that metacognitive characteristics can differentiate mathematical performance even among relatively young learners, suggesting that metacognition is not merely an advanced skill but a central component of early mathematical development.

Although Nováková's (2024) study points to a positive link between metacognition and performance, the broader literature reveals considerable variability in the strength of this relationship across studies and age groups. In their meta-analysis, Brown and Harris (2013) found that relationships between self-assessment and performance varied widely, with effect sizes ranging from 0.04 to 1.62 (Cohen's *d*). This variability was not determined by the type of self-assessment but by its complexity. One possible explanation is the formative nature of assessment. For example, Black, Harrison, Lee, Marshall, and Wiliam

(2004) demonstrated a strong relationship between formative assessment (including self-assessment) and achievement among 11- to 15-year-olds.

Feedback may also influence the strength of the relationship. Sitzmann, Ely, Brown and Bauer (2010) found that correlations between self-assessment and learning were stronger in courses that included feedback than in those without it. Although their meta-analysis focused on adults, evidence suggests that similar mechanisms operate in children. Elder (2010), for instance, showed that both younger (1st grade) and older (4th/5th grade) pupils rely on evaluations provided by others, especially teachers, though older pupils draw on a broader range of sources when forming self-assessments.

One of the few intervention studies targeting elementary pupils (grades 5–6) is the work by Ross et al. (2002), who examined the effects of self-evaluation training on mathematics achievement. Their 12-week program, which included a range of activities aimed at improving self-assessment, led to improved mathematical problem-solving in the experimental group compared to the control group.

Taken together, these findings suggest that comprehensive self-assessment training supported by feedback can lead to more accurate self-evaluations and improved problem-solving performance. However, further research is needed to examine the sources of self-assessment accuracy in younger children in lower grades. Only then will it be possible to develop effective self-assessment interventions tailored to this age group.

As demonstrated by prior research, self-assessment is closely linked to a range of psychological factors that play a pivotal role in school learning. Nevertheless, establishing self-assessment practices in the classroom may be challenging. Individual variability in cognitive abilities and developmental maturity must be taken into account. In light of these considerations, the present study focuses on the accuracy of self-assessment in mathematics across two developmental periods. Mathematical word problems were selected because of their recognized complexity within the curriculum (Rendl et al., 2013). These problems not only reflect real-life applications but also have a documented connection to self-assessment, as discussed earlier. The study seeks to explore changes in self-assessment accuracy in the context of mathematical word problems and examine its relationship with cognitive ability.

The research aims to investigate whether pupils can accurately evaluate their performance in test items, particularly in the challenging domain of mathematical word problems. Our study is guided by the following research questions: (1) How does self-assessment accuracy in solving mathematical word problems develop from Year 2 to Year 3 among elementary school pupils? (2) Is there a relationship between self-assessment accuracy and cognitive ability in solving mathematical word problems?

2 Methodology

This study forms an integral part of two larger research projects that synergistically complement one another. The first project is dedicated to enhancing pupils' problem-solving strategies in mathematical word problems, an area perceived as challenging by elementary school pupils. Regrettably, no longitudinal, comprehensive data from primary school pupils were available for this research. The second project concentrates on strategies employed by teachers to mitigate the risk of school failure. This study encompasses data from 29 primary school classes, gathered through a variety of methods, including tests, standardized psychological instruments, and teacher questionnaires. In this paper, we present the results of our study, which utilizes data collected in the second research project and analyses it within the framework of the objectives established in the first research project.

2.1 Measurements

Mathematical test and word problems

The main tool used in the study was a test on mathematics. They have been developed in several steps following Downing's recommendations (Downing, 2006). The content was based on the Framework Educational Programme for Basic Education (FEP BE) that sets out the expected learning outcomes. The structure of the test, including the format, types of questions, and scoring criteria, was reviewed by a group of educational experts. Test items are systematically generated, reflecting the content and cognitive skills to be assessed. The items were pretested to enhance their quality, considering clarity, relevance, and fairness. The same test specification was used for in Year 2 and Year 3 with respect to increasing knowledge and cognitive skills. Thus, both test results can be compared. The psychometric parameters of both tests were good. Pupils solved tests individually, but in the presence of other classmates, their own teacher, and a researcher whom they already knew. The situation was relatively familiar to them and thus it was not a source of great stress. The tests had no time limit. Due to the COVID-19 situation,

the administration of the first test was delayed, resulting in a reduced level of difficulty compared to the second test.

The mathematics test contained 7 tasks with a total score of 100 points. Pupils' answers in tests were recorded in electronic form and then evaluated. The sum of the points for the individual tasks resulted in the total score. Each task was generated to measure a certain educational outcome requested by the Framework Educational Programme for Basic Education (FEP BE). The type of task was either typical or untypical for school lessons in researched classes, however, for the purpose of the study, only typical items were used. The maximum score for each task was set based both on the levels in which the task was typical for school lesson and cognitive challenging. The analysis of self-assessment accuracy across the entire test was presented in a previous article (Švamberk Šauerová & Smetáčková, 2022). However, due to limited comparability between the tests, a closer examination of individual items with similar content was necessary.

The tests in Year 2 and Year 3 contained of one school-typical mathematical word problem (M2 in Year 2 and M3 in Year 3) with similar structures. In both cases, it was complex multi-step problems in which pupils needed to solve individual phases through addition and subtraction. The difference between the tasks was both in the number of steps and in the size of the numbers with which the mathematical operations were performed. The wording of both mathematical word problems is given below.

The task M2 for Year 2 was: *There are 11 girls in the class, 4 more boys than girls. There are ____ boys. There are ____ fewer girls than boys. There are ____ children in the class.* The task diagnoses the pupil's ability to connect two situations described in the addition task, which requires reading comprehension, and the pupil's ability to add two small numbers: $11 + 4$ and $11 + 15$. It is a typical word problem with the comparison operator ("4 more than"). The first subtask was deliberately easier, as it was intended to allow pupils to experience success and motivate them to solve further problems. The context of the task is familiar to the pupils and is part of their life experience. The text of the task is simple, the words 'more' and 'total' correspond to the operation of addition which leads to the correct result. The number range corresponds to the beginning of Year 2 when the number range starts to expand from 20 to 100. The two numbers in the problem (11 and 4) express numbers and are easily modeled using some manipulatives. The problem is compound – the calculation of the first part of the problem is needed for the second part. Between the two parts is sandwiched a statement for the pupil to complete, which tests whether the pupil is aware of the link between the relations more and less.

The task M3 for Year 3 was: *Zdenka got 403 points in the computer game, Tereza got 118 points more than Zdenka, David scored 20 points less than Tereza. Tereza scored ____ points. Zdenka and Tereza scored together ____ points. David scored ____ more points than Zdenka. David scored ____ points.* The task, like the previous one, diagnoses the pupil's ability to connect the situations described in the task with additive operations. The key is understanding the text, that is, the discovery of the links between the number of points of each child, and the ability to perform additive operations in the domain of natural numbers up to one thousand. The task context belongs to pupils' lives, but not every pupil needs to have experience with computer games. The number domain corresponds to Year 3. The numbers in the problem represent a number but are too large to use manipulatives. The text of the problem is rather short but contains a lot of information. The problem is complex and requires chaining of thought operations. Completing the first two statements is easy. For the first, the comparison operator (118 more) is added to the state (403 points). The second requires adding two forms, one of which is in the problem statement, and the other is written in the first statement. The completion of these first two statements is identical in idea to that in Problem 1 of the Year 2 test, only the number field is adapted to Year 3. Completing the third statement is more challenging as it requires finding the comparison operator. The pupil can arrive at it in two ways, the first of which is easier. Either the pupil completes the fourth statement first (the comparison operator "20 points less" subtracts from the state "521 points") and then calculates the comparison operator from the two known states (501 and 403). Or the pupil works only with comparison operators (Tereza 118 points more than Zdenka, David 20 points less than Tereza). The number field corresponds to Year 3. The numbers in the problem express a number, but they are too large to use manipulatives.

For each item, after solving it, the pupils were asked the question: *How do you think you managed to solve the problem?* Pupils indicated using circling one of three options: thumbs up (good performance), thumbs horizontally (uncertain or average performance), or thumbs down (poor performance). For each task, including the mathematical complex word problem an accuracy index of the pupil's self-assessment was created, and their sum resulted in an accuracy index for the word problem item and for the math test overall.

Cognitive abilities

Besides tests on mathematics, pupils in Year 2 and Year 3, also completed an independent standardized test measuring cognitive abilities. Nonverbal reasoning was tested by Coloured Progressive Matrices (CPM) in Year 2, respectively by Standard Progressive Matrices (SPM) in Year 3. In each year, the cognitive test was assigned several weeks before the test. The cognitive test was used in a one-off manner and thus may have been influenced by situational factors.

Teachers evaluations of individual pupils

Therefore, we used the teacher's assessment of cognitive abilities as a supplementary contextual data source based on long-term experience with the pupils in school tasks, to provide additional background information for Research Question 2 rather than as a separate research question.

The measure used was a questionnaire in which teachers were asked to make expert judgments about each pupil on 32 items relating to different aspects of school achievement. The following five items related to the assessment of cognitive ability: *He/she can concentrate for 10–15 minutes on one activity; He/she has a good memory, remembers easily; He/she is a logical thinker; He/she is inquisitive, likes to learn new things; He/she immediately understands the task and the teacher's instructions*. Each statement was rated on a scale of 1 (statement about the child is true) to 3 (statement about the child is not true).

The analytical unit was the pupil. For each pupil, characteristics were analyzed as assessed by two tests and within them for one complex mathematical word problem as well as his/her achievements were ascertained with two independent standardized achievement measures.

2.2 Sample

The study included 657 pupils, of whom 49% were boys and 51% were girls. Complete longitudinal data were obtained from 542 pupils. The pupils belonged to 29 primary school classes. The classes were selected at the end of Year 1 and subsequently tested in Year 2 and Year 3. Classes were included based on the following criteria: a) an average-size school and class with a typical proportion of pupils with special educational needs according to national statistics; b) no specific or alternative educational programme; c) agreement from teachers, parents, and pupils to participate in a three-year research project involving extensive data collection through psychological tests, interviews, and classroom observations; d) an experienced teacher, whose practice was consistently rated as effective by school leadership, colleagues, parents, and the Czech School Inspectorate. Although these indicators offer a reasonable proxy for effective teaching, they do not encompass all dimensions of teacher quality, which is a limitation of the study. Class sizes ranged from 15 to 29 pupils, with an average of 23, which corresponds to the national mean in the Czech Republic.

3 Results

Descriptive statistical results of mathematical word problem-solving during Year 2 and Year 3 are presented in Table 1. This comprehensive analysis encapsulates crucial metrics encompassing the mean success rate, average self-assessment scores, and the accuracy of these self-assessments.

Table 1: Scores and self-assessment in mathematical word problems in Year 2 (M2) and Year 3 (M3)

	<i>N</i>	Minimum	Maximum	Mean	Std. Deviation
M2 score	624	0	13	11.18	3.05
M3 score	608	0	12	8.16	4.20
M2 self-assessment	599	1	3	1.25	0.49
M3 self-assessment	591	1	3	1.67	0.63

Both mathematical word problems were proficiently addressed, with an average score of 11.2 out of a maximum of 13 points in M2, and 8.2 out of 12 points in M3. While 62.7% of Year 2 pupils accurately solved the M2 problem, a substantial 27.9% responded partially correctly, and only 9.4% answered incorrectly. However, in Year 3, the scenario underwent a significant transformation, with only 31.3% solving Problem M3 correctly, 48.8% responding partially correctly, and 20% offering incorrect solutions. Notably, despite an overall dominance of high performance, there was a substantial decline in the success rate between Year 2 and Year 3 ($p < 0.001$). The M2 test was less challenging due to its delayed administration caused by the COVID-19 situation. Therefore, the poorer performance on the

M3 test does not necessarily indicate a decline in children's abilities. These differences in performance prompted further consideration of the structural characteristics of the two tasks.

It is therefore important to note that the apparent decline in pupils' accuracy between M2 and M3 may partly reflect methodological characteristics of the two tests. The third-grade task (M3) was not only more cognitively demanding but also included larger numbers and a more complex wording structure. Therefore, the difference in performance could result from the task's higher difficulty rather than a true decrease in self-assessment accuracy.

Since performance and self-assessment accuracy may follow different developmental trajectories, both outcomes were examined separately. The accuracy decreased as well. In M2, 58% of pupils estimated their achievement adequately, while in M3 it was only 32%. The paired *t*-test shows that this result is statistically significant ($p < 0.001$) and it is a high effect as measured² by Cohen's $d = 0.761$. We would reach the same conclusion in the case of the self-assessment for these two tests, with $p < 0.001$. However, this is no longer a large effect, as Cohen's $d = 0.018$. This divergence highlights that changes in performance and changes in the accuracy of self-assessment cannot be interpreted as parallel phenomena.

Based on the first research question, the analysis concerned whether the accuracy of self-assessment evolves over time. Given the observed differences in both performance levels and accuracy between M2 and M3, it was necessary to examine self-assessment patterns in a more controlled way. To ensure that the analysis is not burdened by different success rates in Task 2 and Task 3, we looked at accuracy separately for the group of pupils who gave a completely correct or incorrect answer in each task. The results are presented in Table 2. The table shows that in M2, almost twice as many pupils considered their solution correct in a situation where they solved the problem incorrectly in comparison with M3 (63,3% vs. 36.4%). In contrast only slightly more than half of the successful pupils in M3 considered their solution to be correct.

Table 2: Comparison of the self-assessment accuracy and the achievement

Achievement in Task	Self-Assessment	M2		M3	
		Frequency	Percent	Frequency	Percent
0 point (wrong answer)	Correct answer	136	63.3%	147	36.4%
	Not sure or average performance	69	32.1%	211	52.2%
	Wrong answer	10	4.7%	46	11.4%
	Total	215	100%	404	100%
1 point (correct answer)	Correct answer	330	85.9%	103	55.1%
	Not sure or average performance	49	12.8%	77	41.2%
	Wrong answer	5	1.3%	7	3.7%
	Total	384	100%	187	100%

In addition, we analysed an independent subgroup of pupils who solved both tasks correctly ($N = 143$). The average accuracy of the self-assessment (on a scale of -1 ; 1) for item M2 was -0.063 ($SD = 0.20$) and for item M3 -0.21 ($SD = 0.27$). This difference was significant ($p < 0.001$) and showed that the accuracy of self-assessment declined over time among same children who are correct solvers. Pupils are more rigorous in assessing their own performance in test tasks and there is an increasing tendency towards self-deprecation. This pattern can be explained methodologically: M2 was generally easier, allowing most pupils to succeed and thus to correctly evaluate their success. In contrast, M3's complexity may have increased uncertainty, leading to more mismatches between actual and perceived performance.

According to the comparison of the accuracy of self-assessment in Year 2 and Year 3, we divided the pupils into several groups. The accuracy of the self-assessment measured by the Bias procedure takes values from -1 to $+1$, where negative values indicate underestimation (the pupil's estimate was worse than the actual performance), positive values indicate overestimation (the estimate was better than the actual performance), and a value of 0 indicates agreement between the estimate and the actual performance. We compared whether the pupil's self-assessment was adequate (accurate), overestimating or underestimating on both the first and second measures. The proportions of each variation in the accuracy of self-assessment are shown in Table 3.

The results show that the accuracy of pupil self-assessment is quite low. Those who are stably capable of accurate self-assessment make up only $1/5$. If we also include the subset of pupils who have moved towards accurate self-assessment in at least the second measurement, the proportion rises to just under $1/3$. A similar proportion of pupils tend to overestimate their performance and a slightly higher proportion tend to underestimate their performance.

²0.2–0.5: small effect; 0.5–0.8: medium effect; 0.8 or more: large effect (Cohen, 1988)

Table 3: Number of pupils according to the self-assessment accuracy over time

Accuracy of Self-Assessment		Number of pupils	Percent
M2	M3		
Accurate	Accurate	121	22.4%
Overestimation	Overestimation	57	10.5%
Underestimation	Underestimation	48	8.9%
Accurate or Underestimation	Overestimation	115	21.3%
Accurate or Overestimation	Underestimation	152	28.1%
Overestimation or underestimation	Accurate	48	8.9%

For further analyses, we included only three groups of pupils whose self-assessment accuracy was stable across both measures. These groups were (1) pupils who stably underestimated themselves (their self-assessment was below their real performance; Underestimation/ Underestimation in Table 3), (2) pupils whose self-assessment was stably accurate (their self-assessment matched their performance on both tasks; Accurate/Accurate in Table 3), and (3) pupils who overestimated themselves (their self-assessment overestimated their real performance; Overestimation/Overestimation in the Table 3).

The groups were compared in the cognitive ability measured by Coloured Progressive Matrices (CPM) in Year 2, respectively by Standard Progressive Matrices (SPM) in Year 3. The highest score in the CPM might be 36 and in the SPM 60. An additional variable was used in the teacher's evaluation of pupils' cognitive skills. The score for the dimension "Cognitive skills" was computed from five items and presented as an average on a scale from 1 to 3. The means in all three variables for groups according to the development of self-assessment accuracy are shown in Table 4.

Table 4: Cognitive abilities of three groups based on the self-assessment accuracy

	Underestimated pupils (<i>N</i> = 48)		Accurately estimated pupils (<i>N</i> = 121)		Overestimated pupils (<i>N</i> = 57)	
	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>
Teacher's Evaluation	1.64	0.60	1.32	0.46	1.64	0.56
CPM (Year 2)	27.06	5.63	29.90	3.60	26.91	4.39
SPM (Year 3)	37.68	8.20	42.11	6.88	35.50	8.72

Within the comparison of individual groups, a one-factor analysis of variance was used.³ The determined values are in Table 5.

Table 5: ANOVA – Cognitive abilities of three groups based on the self-assessment accuracy

Areas monitored	ANOVA	Effect size – ω^2
Teacher's Evaluation	$F = 10.964; p < 0.001^{***}$	$\omega^2 = 0.042$
Post-hoc analysis	$1 > 5 (p = 0.001)^{***},$ $5 < 9 (p < 0.001)^{***}$	$\omega_{\text{Lower Index}}^2 = 0.001$ $\omega_{\text{Upper Index}}^2 = 0.071$
Raven 1	$F = 12.818; p < 0.001^{***}$	$\omega^2 = 0.058$
Post-hoc analysis	$1 < 5 (p < 0.001)^{***},$ $5 > 9 (p < 0.001)^{***}$	$\omega_{\text{Lower Index}}^2 = 0.009$ $\omega_{\text{Upper Index}}^2 = 0.091$
Raven 2	$F = 15.729; p < 0.001^{***}$	$\omega^2 = 0.057$
Post-hoc analysis	$1 < 5 (p < 0.001)^{***},$ $5 > 9 (p < 0.001)^{***}$	$\omega_{\text{Lower Index}}^2 = 0.008$ $\omega_{\text{Upper Index}}^2 = 0.089$

Explanations: $^*(p < 0.1)$; $^{**}(p < 0.05)$; $^{***}(p < 0.001)$

The analysis of variance shows that statistically significant differences at the one percent significance level are found in all areas. The post-hoc analysis conducted through the Scheffeny test shows that the 5 groups always differ from the same groups. While in terms of the teacher's perspective, the values are significantly lower than the other two groups, in terms of the Raven's test the values are significantly higher for both year groups. In terms of substantive significance, these are not very significant effects.

Pupils who were consistently accurate in word problems had significantly higher cognitive ability (as measured by the standardized test and as judged by the teacher) than both other groups. In contrast, the cognitive abilities of pupils who overestimated themselves were the weakest.

³Data were assessed for normality using the Shapiro-Wilk test and equality of variances using Brown-Forsythe test. When necessary data were transform to fit assumptions of normality and homogeneity of variance prior the analysis.

In addition to cognitive ability, which is an individual characteristic, it is necessary to ask to what extent the accuracy of self-assessment varies from class to class. In our research, 29 classes of the same grade were included. Comparison of the proportion of each subgroup of pupils according to the accuracy of self-assessment confirmed strong differences between classes. The number of children with consistently accurate self-assessments in word problems varied across classes from 7% to 39%. Some classes had a strong group of underestimating pupils (up to 22%), while others had a group of overestimating pupils (up to 28%). In some classes, the accuracy of self-assessment was unstable, with up to 1/3 of pupils moving from overestimating to underestimating or in opposite direction between Year 2 and Year 3.

4 Discussion

Our study examines patterns of self-assessment accuracy in mathematical word problem-solving among elementary school pupils. We focus on changes in accuracy across grade levels, the influence of cognitive ability, and the role of classroom environments. The research addresses two questions: (1) How does self-assessment accuracy in solving mathematical word problems develop from Year 2 to Year 3? (2) Is there a relationship between self-assessment accuracy and cognitive ability in this domain? Because the Year 2 test was less challenging due to COVID-19 delays, children performed better in Year 2 than in Year 3. This decline in performance was accompanied by a noticeable drop in self-assessment accuracy. Although average accuracy remained relatively stable, older pupils in Year 3 were more likely to believe they had solved the task incorrectly, suggesting a reduced proportion of children capable of accurately judging their performance. Methodologically, these differences may be driven not only by development but also by variation in task difficulty and the limited number of items. With only one complex word problem per grade, random factors (e.g., text comprehension, familiarity with context) could disproportionately influence accuracy. Future studies should therefore employ multiple tasks of varying difficulty levels.

Although our observations cover only two school years, the results align with Harter's (2011) assertion that self-assessment tends to decline as children move from early to middle childhood. This shift often reflects a more critical self-view, which may lead to performance underestimation. At the same time, it is important to consider potential teacher influences—third-grade teachers may apply stricter evaluation criteria than second-grade teachers—an issue explored later in this article.

We also examined the relationship between self-assessment accuracy and cognitive ability. Using standardized tests and teacher evaluations, we found substantial differences in cognitive skills across accuracy groups. This finding supports conclusions by Bradshaw (2001) and Dunning, Heath, and Suls (2004), who identified cognitive ability as a key factor underlying reduced accuracy in self-assessment.

Our study revealed variation in self-assessment accuracy across classrooms. Beyond age and cognitive ability, teaching practices and classroom climate played central roles. Interestingly, the frequency of self-assessment activities did not predict accuracy. This suggests that not only the presence but also the form and quality of self-assessment practices matter. Teachers' communication, instructional strategies, and assessment methods were identified as influential factors, consistent with findings by Brookhart et al. (2004) and Andrade and Valtcheva (2009). More targeted research is needed to understand these dynamics.

Self-assessment is an essential component of self-regulated learning, which is increasingly recognized as foundational for academic success and lifelong learning. Accurate self-assessment enables pupils to reflect on performance, set realistic goals, and assume responsibility for their learning. Our contribution lies in showing that self-assessment accuracy does not automatically improve with age or exposure to self-assessment activities. In fact, accuracy may decline when cognitive demands increase or when classroom conditions become more complex. This challenges the assumption that developmental progression alone enhances metacognitive skills and underscores the need for intentional pedagogical strategies.

Our findings illustrate the dynamic nature of self-assessment and its variation across pupils with different cognitive abilities. Teachers should be mindful of these differences and adjust their approaches accordingly. Simply expecting children to develop more precise self-assessments as they mature is insufficient. Developmental shifts may introduce new challenges, such as heightened sensitivity to self-esteem threats or increased peer comparison, which can impede accuracy. Developing precise self-assessment therefore requires carefully structured practices tailored to pupils' cognitive and metacognitive abilities (Siegesmund, 2017). Pupils with weaker abilities may require targeted interventions to strengthen their capacity for evaluating their performance.

Even in classrooms where teachers implement self-assessment tools, our study highlights inconsistencies in the proportion of pupils who assess themselves accurately. These discrepancies may stem from uniform, one-size-fits-all approaches using a single self-report scale. Such approaches fail to account for pupils' individual tendencies to overestimate or underestimate. Our findings emphasize the need for

more nuanced and individualized self-assessment methods. Teachers should adapt strategies and tools to pupils' characteristic self-evaluation patterns, which may be linked to personality and self-esteem. While our study does not provide an exhaustive set of effective techniques, it clearly signals the need for further research in this direction.

Several limitations should be considered when interpreting the findings. The comparability of the two assessment points was influenced by a delay in administering the first test, which resulted in a slightly lower difficulty level than intended. Although tasks M2 and M3 had comparable structures, the broader number range and greater linguistic complexity of M3 may have contributed to differences in performance and accuracy unrelated to development. The small number of test items and participating classes also reduces the generalizability of the findings. Although the same children were tested at both time points, the design does not constitute a fully controlled longitudinal study, as differences in task difficulty may influence interpretations of developmental change. Testing in familiar classrooms reduced stress but may have introduced subtle social influences. Finally, the absence of standardized measures of metacognition and the limited range of tasks likely constrained our analyses. Future research would benefit from longitudinal designs, use of multiple task types, and more controlled conditions.

Future work should employ a more differentiated set of assessment tasks varying in linguistic and numerical complexity. Multiple items at each measurement point would allow for more reliable estimates of accuracy and help distinguish task-specific effects from genuine developmental trends. Incorporating standardized metacognition instruments would also strengthen analyses. A longitudinal design following the same cohort over several years would further clarify how cognitive ability, task difficulty, and instructional context shape the accuracy of pupils' self-evaluations over time.

5 Conclusions

In conclusion, our research aligns with contemporary educational paradigms that emphasize learner agency, self-regulation, and the development of accurate self-assessment. Our findings contribute to the growing body of evidence that self-assessment is not a static skill but one requiring intentional support and differentiation. The study reveals that pupils' self-assessment does not inevitably improve with age; instead, it may decline. This poses a significant educational challenge and highlights the need for pedagogical approaches designed to counteract this decline and increase the proportion of pupils capable of accurate self-evaluation.

Acknowledgment

This study was financially supported by the Ministry of Education, Youth and Sports of the Czech Republic through the grant “*Teachers' understanding of the causes of school failure and the effectiveness of educational interventions*” (Project No. CZ.02.3.68/0.0/0.0/19_076/0016390). The research was further supported by the Technology Agency of the Czech Republic within the project “*Supporting the integration of mathematical, reading and language literacy in primary school pupils*” (Project No. TL03000469), carried out at the Faculty of Education, Charles University.

References

- Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4, 87. <https://doi.org/10.3389/feduc.2019.00087>
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*, 48(1), 12–19. <https://doi.org/10.1080/00405840802577544>
- Bakar, M. A. A., & Ismail, N. (2020). Metacognitive learning strategies in mathematics classroom intervention: A review of implementation and operational design aspect. *International Electronic Journal of Mathematics Education*, 15(1), em0555. <https://doi.org/10.29333/iejme/5937>
- Bandura, A. (2014). Social-cognitive theory. In R. B. Ewen, *An introduction to theories of personality* (pp. 341–360). Psychology Press.
- Barana, A., Boetti, G., & Marchisio, M. (2022). Self-assessment in the development of mathematical problem-solving skills. *Education Sciences*, 12(2), 81. <https://doi.org/10.3390/educsci12020081>
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Research into practice: Formative assessment for learning. In E. F. Redish & M. Vicentini (Eds.), *Research on Physics Education* (pp. 91–102). IOS Press. <https://doi.org/10.3254/978-1-61499-012-3-91>

- Bradshaw, B. K. (2001). Do pupils effectively monitor their comprehension? *Reading Horizons: A Journal of Literacy and Language Arts*, 41(3), 2. Retrieved from https://scholarworks.wmich.edu/reading_horizons/vol41/iss3/2
- Broadfoot, P. (2021). *The sociology of assessment: Comparative and policy perspectives. The selected works of Patricia Broadfoot*. Routledge. <https://doi.org/10.4324/9780429262609>
- Brookhart, S. M., Andolina, M., Zuza, M., & Furman, R. (2004). Minute math: An action research study of student self-assessment. *Educational Studies in Mathematics*, 57(2), 213–227. <https://doi.org/10.1023/B:EDUC.0000049293.55249.d4>
- Brown, G. T., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: Directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, 22(4), 444–457. <https://doi.org/10.1080/0969594X.2014.996523>
- Brown, G. T., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 367–393). Sage.
- Brown, G., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research*, 2(1), 22–30. <https://doi.org/10.14786/flr.v2i1.24>
- Cera, R., Mancini, M., & Antonietti, A. (2013). Relationships between metacognition, self-efficacy and self-regulation in learning. *Journal of Educational, Cultural and Psychological Studies (ECPS Journal)*, 4(7), 115–141. <https://doi.org/10.7358/ecps-2013-007-cera>
- Chytrý, V., Říčan, J., & Živná, D. (2019). Matematická výkonnost a metakognice žáků základních škol běžných, základních škol Montessori a žáků vyučovaných podle Hejného metody. *Studia paedagogica*, 24(1), 107–133. <https://doi.org/10.5817/SP2019-1-5>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- De Bruin, A. B., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction*, 22(4), 245–252. <https://doi.org/10.1016/j.learninstruc.2012.01.003>
- Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 3–25). Lawrence Erlbaum Associates Publishers.
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5(3), 69–106. <https://doi.org/10.1111/j.1529-1006.2004.00018.x>
- Elder, A. D. (2010). Children's self-assessment of their school work in elementary school. *Education 3–13*, 38(1), 5–11. <https://doi.org/10.1080/03004270802602044>
- Georghiades, P. (2004). From the general to the situated: Three decades of metacognition. *International Journal of Science Education*, 26(3), 365–383. <https://doi.org/10.1080/0950069032000119401>
- Harter, S. (2011). The development of self-representations. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (2nd ed., pp. 229–245). Guilford Press.
- Kajamies, A., Vauras, M., & Kinnunen, R. (2010). Instructing low-achievers in mathematical word problem solving. *Scandinavian Journal of Educational Research*, 54(4), 335–355. <https://doi.org/10.1080/00313831.2010.493341>
- Klenowski, V. (1995). Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England. *Assessment in Education: Principles, Policy & Practice*, 2(2), 145–163. <https://doi.org/10.1080/0969594950020203>
- Kramarski, B., & Revach, T. (2009). The challenge of self-regulated learning in mathematics teachers' professional training. *Educational Studies in Mathematics*, 72, 379–399. <https://doi.org/10.1007/s10649-009-9204-2>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Li, Z., & Rubie-Davies, C. M. (2017). Teachers matter: Expectancy effects in Chinese university English-as-a-foreign-language classrooms. *Studies in Higher Education*, 42(11), 2042–2060. <https://doi.org/10.1080/03075079.2015.1130692>
- Liu-Ambrose, T., Davis, J. C., Nagamatsu, L. S., Hsu, C. L., Katarynych, L. A., & Khan, K. M. (2010). Changes in executive functions and self-efficacy are independently associated with improved usual gait speed in older women. *BMC geriatrics*, 10(1), 25. <https://doi.org/10.1186/1471-2318-10-25>

- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Nováková, E. (2024). Metacognitive skills of pupils in primary mathematics education. *Journal of Elementary Education*, 17(2), 223–239. <https://doi.org/10.18690/rei.3274>
- Orth, U., Dapp, L. C., Erol, R. Y., Krauss, S., & Luciano, E. C. (2021). Development of domain-specific self-evaluations: A meta-analysis of longitudinal studies. *Journal of Personality and Social Psychology*, 120(1), 145–172. <https://doi.org/10.1037/pspp0000378>
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, 422. <https://doi.org/10.3389/fpsyg.2017.00422>
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74–98. <https://doi.org/10.1016/j.edurev.2017.08.004>
- Pape, S. J., Bell, C. V., & Yetkin, I. E. (2003). Developing mathematical thinking and self-regulated learning: A teaching experiment in a seventh-grade mathematics classroom. *Educational Studies in Mathematics*, 53(3), 179–202. <https://doi.org/10.1023/A:1026062121857>
- Perry, N. E., Phillips, L., & Hutchinson, L. (2006). Mentoring student teachers to support self-regulated learning. *The Elementary School Journal*, 106(3), 237–254. <https://doi.org/10.1086/501485>
- Pintrich, P. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 452–502). Academic Press.
- Rendl, M., Vondrová, N., Hříbková, L., Jirotková, D., Kloboučková, J., Kvasz, L., & Žalská, J. (2013). *Kritická místa matematiky na základní škole očima učitelů* [Critical places of primary mathematics in the eyes of teachers]. Univerzita Karlova v Praze, Pedagogická fakulta.
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research, and Evaluation*, 11(1), 10. <https://doi.org/10.7275/9wph-vv65>
- Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self-evaluation in grade 5–6 mathematics: Effects on problem-solving achievement. *Educational Assessment*, 8(1), 43–59. https://doi.org/10.1207/S15326977EA0801_03
- Šauerová Švamberk, M., & Smetáčková, I. (2022). Accuracy of pupils' self-assessment. *EduPort*, 6(2), 13–25. <https://doi.org/10.21062/edp.2022.009>
- Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading & Writing Quarterly*, 23(2), 139–159. <https://doi.org/10.1080/10573560601158461>
- Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, 36(1–2), 111–139. <https://doi.org/10.1007/s11165-005-3917-8>
- Siegesmund, A. (2017). Using self-assessment to develop metacognition and self-regulated learners. *FEMS Microbiology Letters*, 364(11), fnx096. <https://doi.org/10.1093/femsle/fnx096>
- Sitzmann, T., Ely, K., Brown, K. G., & Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education*, 9(2), 169–191. <https://doi.org/10.5465/AMLE.2010.51428542>
- Taber, K. S. (2025). Mediated learning leading development: The social development theory of Lev Vygotsky. In B. Akpan & T. Kennedy (Eds.), *Science education in theory and practice: An introductory guide to learning theory* (2nd ed., pp. 275–292). Springer.
- Tachie, S. A. (2019). Meta-cognitive skills and strategies application: How this helps learners in mathematics problem-solving. *Eurasia Journal of Mathematics, Science and Technology Education*, 15(5), em1702. <https://doi.org/10.29333/ejmste/105364>
- Yan, Z. (2020). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education*, 45(2), 224–238. <https://doi.org/10.1080/02602938.2019.1629390>
- Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives*, 6(4), 354–360. <https://doi.org/10.1111/j.1750-8606.2012.00246.x>
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82–91. <https://doi.org/10.1006/ceps.1999.1016>
- Zimmerman, B. J., & Schunk, D. H. (2011). Self-regulated learning and performance: An introduction and an overview. In D. H. Schunk & B. Zimmerman (Eds.), *Handbook of self-regulation of learning and performance* (pp. 15–26). Routledge.

Scientia in educatione

*Vědecký recenzovaný časopis pro oborové didaktiky
přirodovědných předmětů a matematiky
Scientific Journal for Science and Mathematics Educational Research*

Vydává nakladatelství Karolinum – <http://www.scied.cz>

Vedoucí redaktorka (Pedagogická fakulta, Univerzita Karlova)
prof. RNDr. Naďa Vondrová, Ph.D.

Redakce (Univerzita Karlova)

doc. RNDr. Svatava Janoušková, Ph.D.
RNDr. Petr Kolář, Ph.D.
prof. RNDr. Jarmila Novotná, CSc.
RNDr. Lenka Pavlasová, Ph.D.
doc. PhDr. Martin Rusek, Ph.D.

Mezinárodní redakční rada

Dr. John Carroll (Nottingham Trent University, Great Britain)
prof. RNDr. Hana Čtrnáctová, CSc. (Univerzita Karlova)
assoc. prof. Robert Harry Evans (University of Copenhagen, Denmark)
RNDr. Eva Hejnová, Ph.D. (Univerzita J. E. Purkyně, Ústí nad Labem)
doc. PhDr. Alena Hošpesová, Ph.D. (Jihočeská univerzita v Českých Budějovicích)
doc. PhDr. Martin Chvál, Ph.D. (Univerzita Karlova)
prof. Dr. Rainer Kaenders (Rheinische Friedrich-Wilhelms-Uni. Bonn, Germany)
RNDr. Alena Kopáčková, Ph.D. (Technická univerzita v Liberci)
PhDr. Magdalena Krátká, Ph.D. (Univerzita J. E. Purkyně, Ústí nad Labem)
prof. RNDr. Ladislav Kvasz, DSc. (Univerzita Karlova)
prof. Dr. Martin Lindner (Martin Luther University Halle-Wittenberg, Germany)
dr. hab. Małgorzata Nodzyńska (Uniwersytet Pedagogiczny, Krakow, Poland)
dr. Samet Okumus (Recep Tayyip Erdogan University, Turkey)
prof. Dr. Gorazd Planinšič, Ph.D. (Univerza v Ljubljani, Slovinsko)
doc. RNDr. Jarmila Robová, CSc. (Univerzita Karlova)
prof. Bernard Sarrazy (Université Bordeaux, France)
dr. hab. prof. UR Ewa Swoboda (Uniwersytet Rzeszowski, Poland)
doc. RNDr. Petr Šmejkal, Ph.D. (Univerzita Karlova)
prof. Dr. Andrej Šorgo (University in Maribor, Slovenia)
doc. RNDr. Vasilis Teodoridis, Ph.D. (Univerzita Karlova)

Adresa redakce

Pedagogická fakulta, Univerzita Karlova, Magdalény Rettigové 4, 116 39 Praha 1
e-mail: scied@pedf.cuni.cz

Pokyny pro autory jsou uvedeny na

<http://ojs.pedf.cuni.cz/index.php/scied/about/submissions#authorGuidelines>.

Sazbu v systému L^AT_EX zpracoval Ing. Miloš Brejcha, Vydavatelský servis, Plzeň.
Logo navrhl Ivan Špírk.

Redaktorka a jazyková korektorka Marie-Anna Kociánová