

Obsah

Alena Hošpesová, Jarmila Novotná, Man Ching Esther Chan, David Clarke Peer-feedback as a part of collaborative problem-solving.....	2
Jaroslav Říčan, Jiří Škoda, Viktorie Hermanová, Barbora Lanková Komparace kvality tzv. teacher made testů s didaktickými testy a jejich vliv na úspěšnost žáků: případová studie	18
Martin Šrámek, Milada Teplá Vlastnosti úloh z obecné chemie vyplývající z analýzy přijímacího řízení na Přírodovědeckou fakultu Univerzity Karlovy.....	32

Peer-feedback as a part of collaborative problem-solving

Vrstevnícká zpětná vazba jako součást spolupráce skupiny při řešení problému

Alena Hošpesová^{1,*}, Jarmila Novotná², Man Ching Esther Chan³, David Clarke³

¹ University of South Bohemia, Faculty of Education, Jeronýmova 10, 371 15 České Budějovice, Czech Republic; hospes@pf.jcu.cz

² Charles University, Faculty of Education, Magdalény Rettigové 4, 116 39 Praha, Czech Republic

³ Melbourne Graduate School of Education, Ground floor, 234 Queensberry Street, Melbourne, VIC 3010, Australia

The paper presents a study within an international project titled *The Social Unit of Learning*. This study examines the ways in which students interact with each other when solving an open mathematical problem in small groups of four. Attention is focused on peer feedback, especially on its form and how it affects the process of solving the problem. Our analysis suggests that feedback often appears in the group discussion and takes the form of both positive reaction (confirmation; question asking for clarification, request for clarification, eliciting) and negative reaction (rejection, explained or unexplained objection, doubt, question involving doubt). The analysis also identified a communication pattern that generalizes the course of the group discussion and factors influencing the quality of peer feedback related to the content structure of the problem (thematic, conceptual and competence layer).

Key words:
collaborative
problem-solving, peer
feedback, mathematics
education, 12 to
13-year-old students.

Received 12/2020

Revised 5/2021

Accepted 9/2021

Článek prezentuje výsledky výzkumné studie uskutečněné v rámci mezinárodního projektu nazvaného *The Social Unit of Learning*. Studie zkoumala způsoby interakce žáků při řešení otevřené matematické úlohy ve čtyřčlenných skupinách. Pozornost se zaměřila na identifikování forem vrstevnícké zpětné vazby a jejího vlivu na proces řešení úlohy. Z analýzy vyplynulo, že se zpětná vazba objevuje v diskuzi skupiny často a má formu pozitivních (potvrzení; požádání o vysvětlení) i negativních reakcí (odmítnutí, vysvětlené nebo nevysvětlené námitky, pochybnosti). Analýza také identifikovala komunikační vzorce v diskuzi ve skupině a faktory ovlivňující kvalitu vrstevnícké zpětné vazby vztahující se k obsahové struktuře problému (tematické, konceptuální a kompetenční vrstvě).

Klíčová slova:
řešení úloh, vrstevnícká
zpětná vazba, výuka
matematiky, základní
škola.

Zasláno 12/2020

Revidováno 5/2021

Přijato 9/2021

1 Introduction, rationale

Learning is a social matter, which implies that knowledge and experience is gained through interaction, communication and sharing with others. Collaborative learning has been used as a pedagogical approach to enable students to solve challenging learning tasks (Damon & Phelps, 1989). It has been found to have strong effects on a range of dependent variables such as achievement, socialization, motivation, personal self-development (Gillies, 2016; Francisco, 2013; Isohäätä et al., 2018), and problem-solving skills (Lohman & Finkelstein, 2000). In the classroom, the promotion of collaborative learning facilitates students to take an active role in their learning through interaction with a group of classmates. The teacher gives instructions to the group of students in accordance with the aim of the lesson and further acts as a consultant and facilitator as the students process and complete the assigned tasks. In an ideal group situation, it is assumed that each student understands his/her role in the work and that the tasks assigned to the groups provide each student with opportunities to express themselves (Cohen, 1994; Cohen & Lotan, 2014).

The Social Unit of Learning Project investigated the social phenomena that characterise learning processes in a mathematics classroom (Chan, Clarke & Cao, 2018). The project used the *Science of Learning Research Classroom* in Australia and equivalent facilities in China to record audiovisually the interactions of whole classes of students and their teacher as the students engage in purposefully developed mathematical activities. In each filming session, multiple cameras and audio inputs were used to capture, in fine-grained detail, teacher-student and student-student interactions and speech throughout the session. A multi-theoretic research design (Chan & Clarke, 2019) adopted by the project involved the construction of a complex data set composed of video records and other supplementary data. The design allows an analytical team to examine different aspects of student collaborative problem-solving. An international multi-disciplinary research team (combining education, cognitive and emotive psychology, learning analytics, and neuroscience perspectives) was recruited to develop analytical frames for coding the data, including analysis of the negotiative foci of student exchange (Chan & Clarke, 2017); dialogic talk between students (Díez-Palomar et al., 2021); student agency (Nieminen et al., 2021); student motivating desires (Tuohilampi, 2018); student interactivity (Chan & Sfard, 2020); behavioural indicators of student

engagement (Chan et al., 2020); student shared cognition (Clarke & Chan, 2020); and material resources (Moate et al., 2021). The researchers each constructed distinct data sets according to different theoretical perspectives applied to the same set of video records and other supplementary data. The multi-theoretic research design allowed the research team to juxtapose different interpretive accounts reflecting different theoretical positions in order to compare and contrast the capacity of different theories to characterise different aspects of the complex classroom setting; to examine their assumptions and implications, as well as their strengths and limitations.

When the authors of this paper started their cooperation, they decided to perform a didactic micro-analysis of communication within a group of students in the process of problem-solving. The research team originally thought that we would perform the analysis from the point of view of peer assessment. However, when working with the data, we found that the recorded reactions of students to suggestions of a classmate often do not have the nature of assessment. They are often a quick response to different types of messages. That's why we started using the term *feedback*, which is more concise. The results from the analysis are valuable for understanding the dynamics of working in groups and for gaining a better understanding of feedback mechanisms by peers when solving a problem. This can be of importance for teachers when preparing didactical situations based on group work to optimize social interactions and promote student learning.

2 Theoretical Background

2.1 Collaborative learning

Collaborative learning is an educational approach which aims to integrate academic and social learning experience in classroom activities (Gillies, 2016). Collaborative learning is not only arranging students into groups. Johnson and his colleagues (Johnson, Johnson & Holubec, 1994) stressed the social dimension in the successful incorporation of collaborative learning in the classroom through: positive interdependence, individual and group accountability, promotive face to face interaction, and the development of the students' interpersonal and small group skills, and group processing. Students must work in groups to complete tasks collectively toward academic goals. Unlike individual learning, which can be competitive in nature, students learning collaboratively can capitalize on one another's resources and skills (asking one another for information, evaluating one another's ideas, monitoring one another's work, and so on) (Chiu, 2008). Furthermore, the teacher's role changes from giving information to facilitating students' learning (Johnson & Johnson, 1984). Everyone succeeds when the group succeeds. Ross and Smyth (1995) described successful collaborative learning tasks as intellectually demanding, creative, open-ended, and involving higher order thinking. Collaborative learning has also been linked to increased levels of student satisfaction (Leikin & Zaslavsky, 1999, p. 244).

Hesse and colleagues defined collaboration "as the activity of working together towards a common goal" (Hesse et al., 2015, p. 33). They mentioned several elements included in this definition:

- communication (the exchange of ideas with the objective to optimise understanding by recipients accompanied by a division of labour);
- responsive contributions to planning and problem analysis;
- active and insightful participation.

Collaboration is dependent on factors such as a readiness to participate, mutual understanding, and the ability to manage interpersonal conflicts.

Collaborative learning as defined by Damon and Phelps (1989) can be seen as closely linked, if not synonymous with collaborative problem-solving. According to Hesse et al. (2015, p. 38), "collaborative problem-solving means approaching a problem responsively by working together and exchanging ideas." They emphasized its usefulness when dealing with complex problems, and a range of social and cognitive skills that students can acquire as part of the activity (e.g., perspective taking, social regulation, resource management, goal setting, among others). Research comparing collaborative learning with traditional instruction-based teaching showed that students who engaged in collaborative learning learned significantly more and enjoyed the classes more (Johnson, Johnson & Holubec, 1994). Leikin and Zaslavsky (1999, p. 245) examined the specifics of students' activities during successful collaborative learning in mathematics education. They indicated that the implementation of the exchange of knowledge promoted students' active explorations and lead naturally to an increase in students' mathematical communications. Students' attitudes towards such an exchange-of-knowledge method were positive, and their achievements in the experimental method were at least as good as those of students learning in the conventional way. An investigation of the types of help those students received while learning showed that verbal explanation was the predominant type of help received by the students.

When analysing the communication of students participating in the *Social Unit of Learning Project*, we realised that one noteworthy part of the collaborative problem-solving activity is that the students provided their schoolmates with “feedback” during the activity, which sometimes was and sometimes was not evaluative in nature. This stimulated our curiosity regarding the ways in which students’ collaborative learning activities constituted useful feedback for each other’s mathematical learning. Our objective is to examine how the discussion in the group supports understanding and shifts or progresses the problem-solving procedure.

To further explain, feedback can be considered as a process of collecting information about people’s reactions to a certain message (Žantovská, 2015). In this sense, feedback is used in many areas of human life, and it is not a concept specific to the field of education. In education, feedback is defined as “information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one’s performance or understanding” (Hattie & Timperley, 2007, p. 81). Feedback is seen as integral to assessment that supports learning, that is, formative assessment (Black & Wiliam, 2009). In the following section, therefore, we start from the explanation of what we understand about formative assessment, and discuss the differences in assessment carried out by teachers and by classmates.

2.2 Formative assessment, self-assessment, and peer feedback

Assessment can be defined as the “process of reasoning from evidence” (Pellegrino, Chudowsky & Glaser, 2001, p. 43), and can be seen as a natural part of human activities. In educational settings, assessment can be distinguished in terms of the purpose for which the assessment is directed. *Summative assessment* “[describes] learning achieved at a certain time for the purposes of reporting to parents, other teachers, the students themselves and, in summary form, to other interested parties such as school governors or school boards” (Harlen & James, 1997, p. 370). *Formative assessment* informs teachers about “students’ existing ideas and skills, and . . . the point reached in development and the necessary next steps to take” (Harlen & James, 1997, p. 369). It involves the processes of data collection and interpretation that teachers use to make decision about the following: What have the students learned? What is student’s learning objective? How can students be supported and assisted during their learning?

Formative assessment has its basics in educational evaluation practices. Researchers in the United States, the United Kingdom, Australia, and New Zealand have contributed to the shift from a focus on formative evaluation to formative assessment or assessment for learning and brought this issue into the spotlight for teacher education programs, teacher professional development, and educational research. Formative assessment has been slowly introduced into the school practice in different parts of the world (OECD, 2005), including the Czech Republic (Žlábková & Rokos, 2013). In upper secondary education practice, different methods of formative assessment are used, for example: marking-grading, structured classroom dialogue, immediate teacher feedback as a response to a student’s problem-solving (on-the-fly assessment) or less frequently written teachers’ comments or formative peer assessment.

From the perspective of this research, the notion of formative peer-assessment is particularly relevant, because of the emphasis on peer interaction in the course of collaborative problem-solving. Peer-assessment is an educational approach for classmates to judge the level, value, or worth of the products or learning outcomes of their equal-status peers. Boud and Falchikov stated that (2007, p. 132):

Peer assessment requires students to provide either feedback or grades (or both) to their peers on a product or a performance, based on the criteria of excellence for that product or event which students may have been involved in determining.

Peer assessment and the assessment made by the teacher, differ considerably in the purpose for which the data are collected and interpreted. In the process of assessment, the teacher has to make a number of decisions about: (a) the data to be collected (e.g., whether systematically or ad hoc), (b) interpretation of the assessment results, (c) communication about the results, and (d) building further decisions based on the results (Black et al., 2004). Peer assessment does not cover the full range of these intentions. The student reflects on the pieces of information obtained from the classmate(s), interprets them, and communicates about them with the classmate(s). The student may compare the classmate’s solution with their own solution as a means of providing feedback. They interpret their classmate’s solution intuitively, mostly without the intention of supporting someone else’s long-term learning.

Despite differences between assessment by teachers or by peers, Slavík (2003) emphasized the importance of peer assessment which the students themselves use, manage, which they understand to the necessary extent and which they can explain or possibly defend. Slavík (ibid) theorised that autonomous assessment, which is in fact the aim of school education, can develop and deepen through self-assessment and through assessment of others’ performance (most likely of classmates, i.e., peer assessment) where students learn to reflect on their work. Boud et al. (2010) stressed that students and teachers in higher

education become responsible partners in learning and assessment. Students should progressively take responsibility for assessment and feedback processes and for that reason they should develop and demonstrate the ability to judge the quality of their own work and the work of others.

Peer-feedback can be effective in supporting student learning for several reasons:

- the mutual exchange of ideas and advice is realized in students' language, where the exchange is formulated in verbal phrases that students commonly use (Black et al., 2004, p. 14);
- if the students do not understand the explanation, it is easier for them to ask their classmate rather than address the teacher in such a situation (students may find the teacher's feedback too complicated, so they prefer peer language) (ibid);
- research and experience show that students can learn from each other even though their intellectual level is different; they share their assessment with each other and can thus better master the subject matter themselves (Sluismans, 2002).

These three reasons served as the basis for formulating the research questions.

2.3 Research questions

The general objective of our research was to verify the contribution of providing and receiving feedback from classmates in the process of collaborative problem-solving. We elaborate this objective in terms of two research questions:

What forms did peer feedback take during the communication between a group of students when solving a problem together?

In what way did the students' peer feedback influence the problem-solving process?

3 Data and methodology

3.1 Data collection

The data used for the analyses presented in this paper were from the *Social Unit of Learning Project* and were collected in the research classroom facility, the *Science of Learning Research Classroom*, at the University of Melbourne, Australia. The research classroom is a university teaching space with built-in video cameras and audio recording equipment for capturing the activities of teachers and students in the room. In the *Social Unit of Learning Project*, Year 7 students (12 to 13 years old) with their usual teacher were brought into the research classroom to participate in purposefully designed activities in mathematics. The students undertook a sequence of tasks individually, in pairs, and in small groups (typically four students). The conditions for student collaborative problem-solving and learning were manipulated so that student and teacher contributions to the learning process can be distinguished. For more details, see (Chan, Clarke & Cao, 2018).

All students' (and the teacher's) classroom speech were video- and audio-recorded and all students' written work was copied and scanned. The organization of the research classroom enabled the recording of the interactions of each student group. The video recordings of the group discussion were transcribed.

The students completed a mathematics problem as part of a larger project that investigated student collaborative problem-solving (for more details, see Chan, Clarke & Cao, 2018). Since the focus of the project was on student-student interaction, the teacher was given explicit instructions to minimise intervention in the students' problem-solving activity. In this paper, video recordings of the collaborative problem-solving activities of two groups of four students were analysed (Group Number 3 and 5). The groups were chosen on the basis of their written solution which appeared to be distinctly different compared to the other groups (see Chan & Clarke, 2017). These two groups then became the focal groups for the application of different analytical perspectives (see Section 1).

3.2 Solved problem and its analysis

The following open-ended problem (Fred's Apartment) was given to students verbally by the teacher, visually on the student worksheet, and was projected on the screen in front of the classroom:

Fred's apartment has five rooms. The total area is 60 square metres.

1. *Draw a plan of Fred's apartment.*
2. *Label each room, and show the dimensions (length and width) of all rooms.*

The students were given 20 minutes to complete the open problem. Silver (1995) described several types of open problems and explained that they are “susceptible to different interpretations or to different acceptable answers” (p. 68). This definition fits the Fred’s Apartment problem. Although the educational goal of the activity was not explicitly stated, in the recorded lesson, we can deduce that the students should deepen their problem-solving competencies. The problem was purposefully designed so that students need to draw from knowledge outside of the classroom and determine the task constraints themselves in order to complete the task.

In analysing the structure of educational content, Slavík et al. (2017, p. 242) distinguished several layers of educational content of a task – see the diagram in Fig. 1.

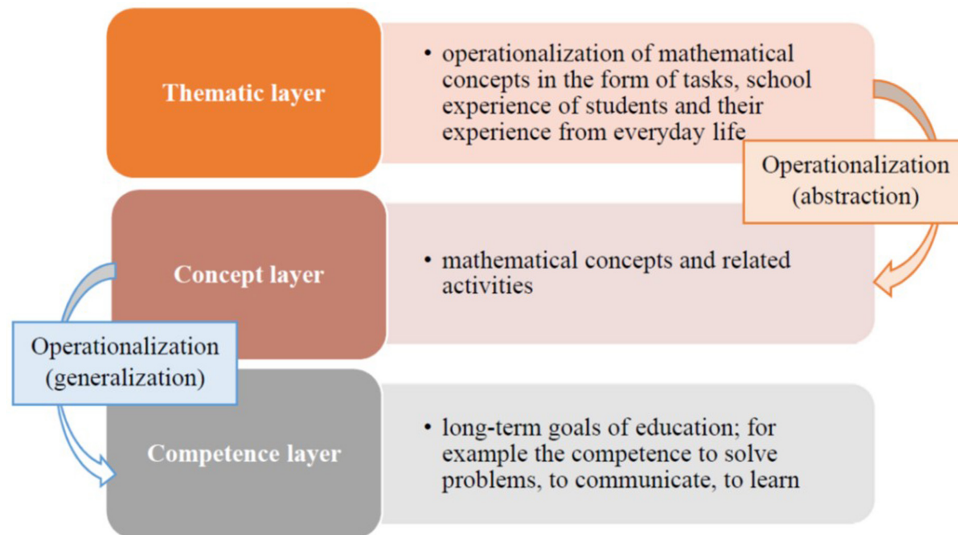


Fig. 1: Diagram of the analysis of the structure of educational content (adapted according Slavík et al., 2017, p. 342)

Analysing the problem of Fred’s Apartment, we start from the conceptual layer of the problem. We identify which key concepts students have to master in order to be able to solve it. The central concept in the task is the concept of area (how it is measured, calculated and what it means). Related to this is the shape and dimensions of the shape in the plane, which we know has the area of 60 m². In addition, this shape should be divided into five sections, for which students choose the shape and dimensions.

When designing the shape of the apartment, its dimensions and division into parts, students rely on experiences from everyday life in terms of what they know about living spaces – they move to the thematic layer of the problem.

To understand the following analyses, we include a brief summary of how the groups proceeded to solve the task.

Group 3 (two girls: Audrey and Katie; two boys: Pedram and Poya) started solving the problem by considering the size of the rooms. The students first solved this problem arithmetically – how 60 square metres can be divided into 5 parts. The discussion was confused by the fact that some students were talking about the area of the rooms, others about their dimensions. From the beginning, this group tried to decide whether the corridor should be part of the apartment. They repeatedly returned to the issue and referred to various practical experiences (for example the size of the room, the accessibility of the flat from outside). The group also addressed the issue of door placement and movement around the apartment. Finally, in the last two minutes, a student (Katie) drew a plan (Fig. 2) of the apartment on her own. The others gave her no feedback.

Group 5 (two girls: Anna and Pandit; two boys: John and Arman) started solving the problem by discussing the scale of their drawing. The discussion was rather lively but it focused on the form of the flat/rooms, their sizes, number of rooms etc. Group 5 was reminded by the teacher that their task is to find the best solution, not the only existing solution. The discussion in Group 5 covered the size of rooms if rooms on the plan were proposed, the number of rooms and their types (kitchen, toilette, living room, lounge room, missing bedroom, workplace, ...). All the time, the discussion focused more on practical aspects than a mathematical solution. When Group 5 started to construct a plan, several advantages and disadvantages of different proposals were raised. Social relationships in the group penetrated through

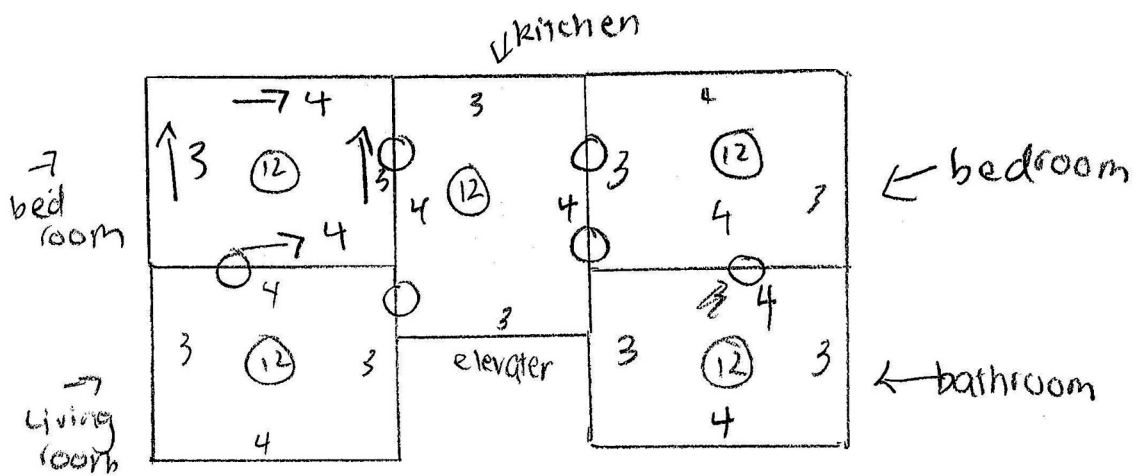


Fig. 2: The solution to the problem drawn by Group 3 (Reprinted from Chan & Clarke, 2017, p. 957, with permission)

the whole activity. There were different attitudes towards mathematics in Group 5 (Tuohilampi, 2018). Finally, the plan in Figure 3 was drawn but even at the end, not all group members were fully happy with the plan. The group did not mention a corridor, but it is included in the plan. The area of 60 m^2 covers the whole flat including the corridor and sanitary facilities.

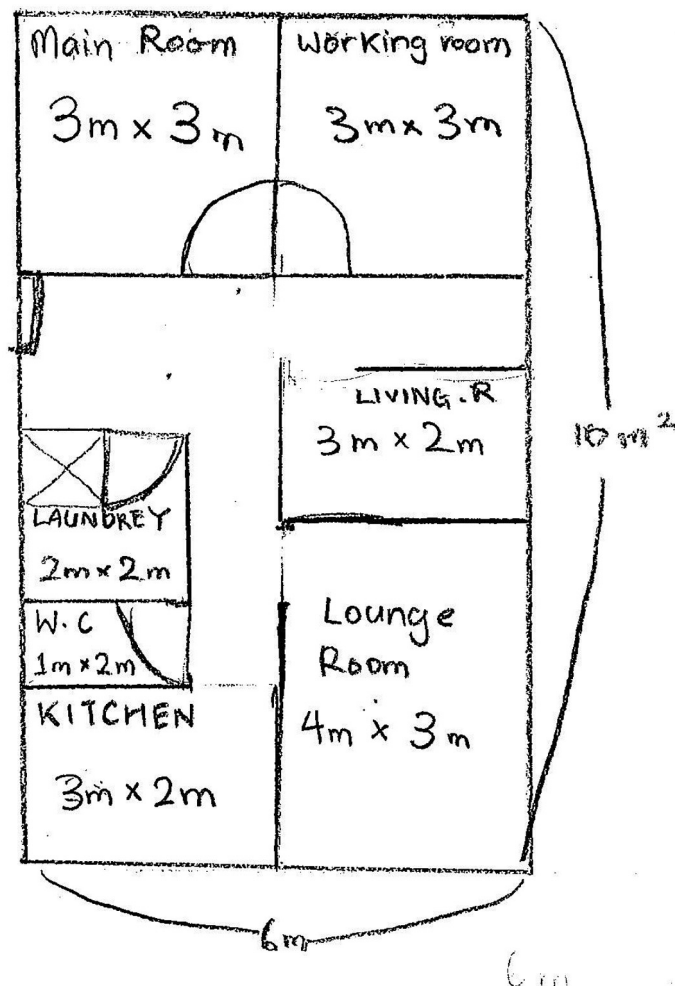


Fig. 3: The solution to the problem drawn by Group 5 (Reprinted from Chan & Clarke, 2017, p. 958, with permission)

3.3 Search for the aim of the research

Hošpesová and Novotná were offered the opportunity to analyze data obtained by Chan and Clark (members of the project team of *Social Unit of Learning Project*) from their research perspective in 2018. At that time they were focusing on formative assessment, namely peer assessment in mathematics. They realized that certain parts of communication of pupils in a group can be regarded as formative peer feedback. The team of authors of this paper decided to conduct a descriptive case study (Hendl, 2005) and formulated the first research question: What forms did peer feedback take during the communication between a group of students when solving a problem together?

In searching for an answer to this question, the team increasingly found that moving towards solving a problem was complicated for the students in some cases. The second research question was formulated: In what ways did the students' peer feedback influence the problem-solving process?

3.4 Determining how to analyse the data for peer feedback

The research design of the current study was qualitative. This was due to the nature of the research questions and the data we analysed. The research had the character of a case study (Hendl, 2005) examining the communication of members of a group of students aimed at solving a mathematical problem. We were inspired by grounded theory research design (Strauss & Corbin, 1990) in our analytical process. Through coding and analytical processes, we identified peer feedback as the main phenomenon for supporting the solution process.

Data analysis was conducted in two delimited stages. The search for the answer to the first research question started by open coding. At first, we worked independently. The research team members processed the transcripts of students' communication in both groups using open coding. The research team exchanged conclusions from analysis during the personal meetings and via Skype communication and looked for consensus in interpretations.

In the next phase, the research team members analysed the transcripts of both groups of students again independently using the list of codes that were agreed upon.

At the joint meeting, we sought re-occurring codes to search for communication patterns in the groups' discussion.

During the analytical procedure we examined segments of communication. Each segment contains a sequence of group members' statements discussing an issue initiated by a group member. This is consistent with Clarke's (2001) definition of a *negotiative event* as "an utterance sequence constituting a social interaction with a single identifiable purpose" (Chan & Clarke, 2017, p. 959). We concentrated on those negotiative events in communication related to problem-solving. We omitted the events which had the nature of a "social" conversation that was not task-focused. The following steps were then followed:

- Selection of negotiative events to be analysed.
- Analysis of students' interaction.
- Comparison of analyses of negotiative events and looking for more general phenomena characteristic for collaborative learning when solving this problem.

In the discussion of Group 3 we distinguished 18 events, and in Group 5 there were 16 events.

4 Results

4.1 What forms did peer feedback take during the communication between a group of students when solving a problem together?

In the list of codes created during open coding, we identified various topics, such as: who carried out the peer feedback, what is the form of peer feedback, how does the assessed person react, and what is assessed. In the following discussion, we concluded that the feedback provided within the student group appeared to be a key factor in influencing learning. Based on the team's experiences, we constructed the initial version of codes labelling different forms of peer feedback and the students' reactions to the feedback.

Forms of peer feedback:

- Agreement (with or without any explanation)
- Objection (without explanation, explained)
- Question
- New proposal

- Praise (of the assessed person, of the assessed solution)
- Doubt (statement, question addressed to schoolmates)
- Change in the solving process
- Self-critique (of mathematical procedure, of the own consideration)

Reactions of the assessed person:

- Changes the result
- Clarifies or refines the original idea
- Restarts the solving process from the beginning (the same or a different one)
- Returns to the previous step
- Stops working/communicating

These codes allowed us to describe the on-going communication in the group. However, they did not indicate much about the individual students' contribution to the solution of the problem.

4.2 “Messages” in peer-feedback

As we went through the transcripts repeatedly, we realized that in the communication of a group of students' utterances were of a similar nature. During ongoing comparison of events and a search for a pattern, it was revealed that there was a similarity with the ESRU model. The ESRU model was discovered by Ruiz-Primo and Furtak (2006) during the analysis of a teacher's feedback on a student's idea that was unexpected. The acronym for this model is derived from the initials of the names of four consecutive steps in providing immediate feedback:

1. The first step is to invoke an interaction (called “E” from *eliciting*).
2. The second step is the *student's response* (called “S”). Here the student responds to the teacher's question, opinion, or objection from the previous step.
3. The third step is the *recognition* (marking “R”) of the student's reaction or the teacher's response to the student's reaction.
4. The fourth and probably the most difficult step for teachers is the *use* of the information obtained (marking “U”).

We believed that coding, which followed the steps of this model, would help us find the answer to the second research question. We entered the second stage of the research. We used only four codes from this model in the coding process: eliciting, response, recognition, use. Again, we worked independently at first and compared our conclusions in our joint “sessions”. In addition to searching for a pattern, we also paid attention to the content of students' communication, to be able to answer the second research question.

The communication, however, was naturally not as “straightforward” as in the communication controlled by the teacher.

1. In a group of students, the discussion of an idea (negotiative event) started with *eliciting* provided by one of the students. Eliciting took various forms, for example the student turned the attention of classmates to some information in the assignment, proposed an idea (next step(s) of the solution), or simply the student asked a question.
2. One (or more) student(s) gave *feedback* to the proposal. As feedback, we will interpret all the reactions of classmates to the proposal. This reaction took the form of an admission, a question, a request for clarification. It could also be a rejection.
3. The one who elicited the discussion *responded* to the feedback from his/her classmates by trying to modify or refine his/her idea as required. Sometimes he/she gave up and suggested a new idea. Sometimes he/she just repeated the original idea. We did not label this reaction as recognizing, because it has a different nature than the recognition given by the teacher. We label this step as *reaction*.
4. If the one who elicited the discussion pushed through the idea, it is used in the next discussion, or in solving the problem or task. Here it seems apt to stay with the original designation: *use*.

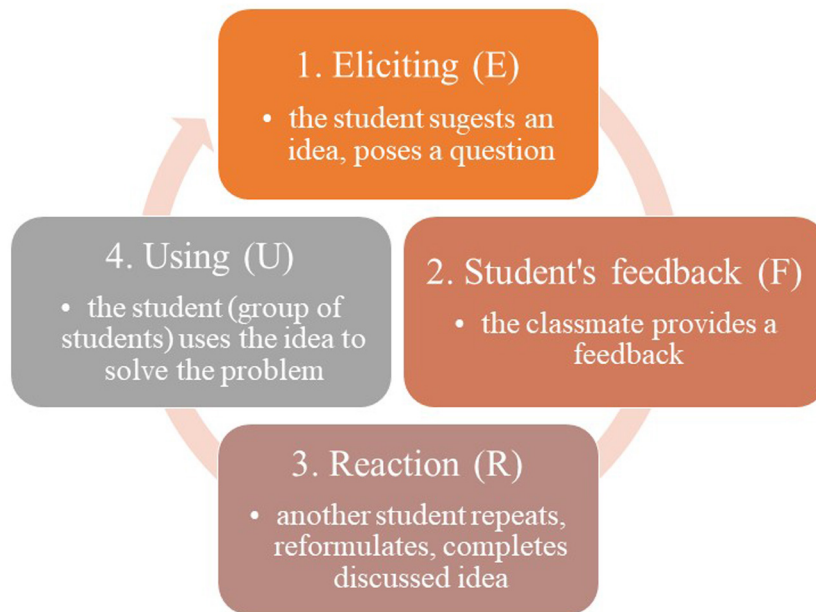


Fig. 4: Model EFRU and its components (schema adapted for this study)

The structure of communication can be described as an EFRU model and it is schematically illustrated in Fig. 4.

In practice, of course, some steps were found to be repeated. Also, in our data, there were several times where the student discussion did not show a clear intention (is the idea applicable to solving the problem or not) and the group reacted to the eliciting response of another student in the group. The group also repeatedly returned to some ideas (e.g., whether a corridor can be considered as a room or not – see below) (schematically elaborated in Fig. 5).

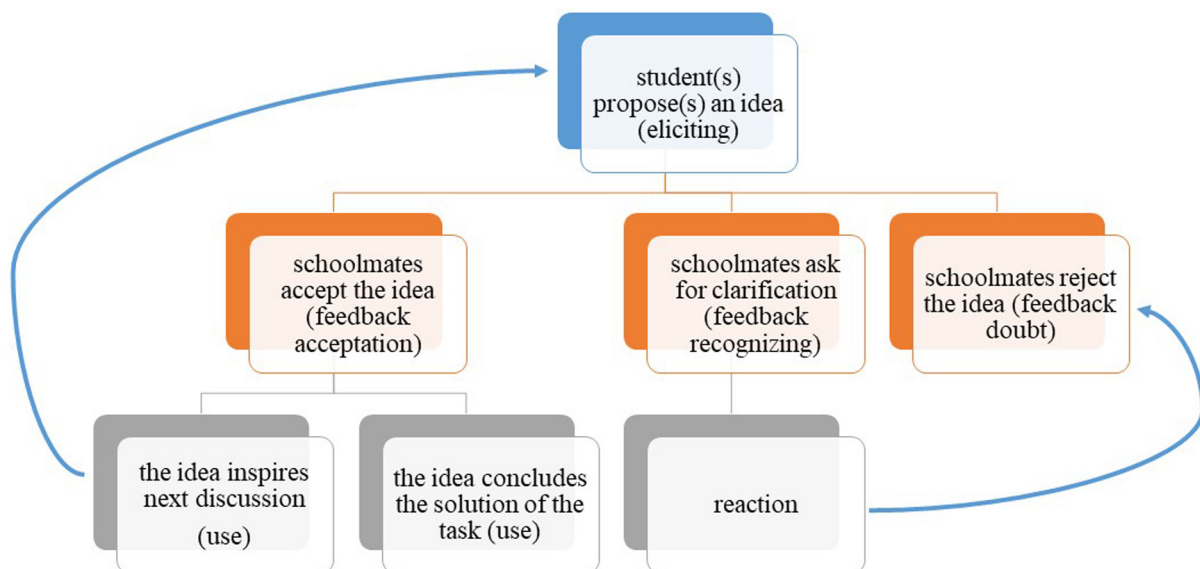


Fig. 5: Application of the EFRU model in peer feedback during group discussion (our own scheme)

The above model is illustrated in the example communication of Group 3 (Transcript 1). Here you can see the cyclic character of the communication (how some of the steps were repeated).

In the performed analyses of students' discussions, the EFRU model appeared to be a suitable tool to deal with the complexity of the student discussion enabling us to take into account the significance of individual utterances for the solution of the problem.

Tab. 1: Transcript Excerpt 1: The communication when clarifying the problem “corridor” (group 3)

Time	Dialogue	Code
101 [36:14.11]	Pedram: But how large will the rooms be?	Eliciting (question)
102 [36:16.24]	Poya: Twelve metres squared.	Eliciting (answer)
103 [36:18.08]	Pedram: Twelve metres squared so there’s no corridor. (Over speaking)	Feedback (objection)
104 [36:22.00]	Poya: Well make them eleven metres squared.	Reaction (changes original idea)
105 [36:23.10]	Pedram: If there are 12 metres square and there are five rooms. . . Dude listen, if there are 12 metres square rooms and there are five rooms, that’s the whole 60 metres	Feedback (explanation of the objection)
106 [36:33.06]	Poya: Okay. Maybe 11 metres squared.	Reaction (changes original idea)
107	Pedram: So there is only five metres.	Feedback (objection)
108 [36:40.00]	Poya: Six, really, six metres squared corridor.	Reaction (clarification)
109 [36:40.12]	Katie: There’s only five rooms.	Feedback (doubt)
. . .		
135 [38:09.11]	Pedram: Sixty metres, make each room five metres. I mean five – each room five metres?	Eliciting
136 [38:21.00]	Poya: Ten metres.	Feedback (new proposal)
137 [38:23.02]	Pedram: Five square metres. Are you smart?	Reaction (explanation)
138 [38:26.21]	Poya: Maybe.	Feedback (doubt)
139 [38:27.11]	Pedram: Five metres, you can’t really step in the room because it’s five metres with the walls. Not the floor, you idiot.	Reaction (recognizes of own mistake)

4.3 In what way did the students’ peer feedback influence the problem-solving process?

The feedback that the students provided to each other is very frequent (see example in Transcript 1), but brief, and sometimes formulated imprecisely. However, the students seem to understand the meaning of each other. For example: at the beginning of their exchange in Transcript 1 (turn 105) Pedram points out that if the five rooms are each 12 metres squared, that they will take up the total area of the apartment (60 square metres) with no space for any corridors. Poya suggests an adjustment by saying: “Okay. Maybe 11 metres squared.” Despite the brief statement, Pedram seems to understand what Poya means. Without explanation, he says the size of the corridor will be five (square) metres (turn 107). However, the correct unit is not used. Poya’s reaction (“Six, really, six metres squared corridor” turn 108) is confusing and Katie’s feedback (“There’s only five rooms.”) recalls the task instructions and, in our interpretation, shows that Katie probably doesn’t understand her classmates. The group returns to the idea of dividing the area of an apartment into five rooms of the same size a little later by Pedram’s proposal (turn 135), which is, however, incomprehensible. He seems to have confused the number of rooms (five) with the area of the room. Poya continues with his original idea, proposing to reduce the size of the rooms to 10 metres (squared) (turn 136). Pedram repeats his original idea (turn 137). Poya does not insist on his idea, and expresses doubt about himself being smart (turn 138). Pedram realizes that their solution is not realistic. The negotiative event can be summarized as: the boys clarified their thoughts, they actively participated in solving the problem. Poya’s idea might have led the group to solve the task, but in the end, he was uncertain about his classmates’ reactions and abandoned the idea.

In the second stage of the research we tried to reconstruct the reasoning of individual students on the basis of what they had said in the discussions. Although we are aware that we may have interpreted the students’ intentions incorrectly, we believe we have managed to identify the moments when the group discussion was moving an appropriate solution forward as well as moments that were clouding the solution.

Peer feedback shifting the problem-solving ahead

As previously outlined, peer feedback may be beneficial for helping students in the process of solving the problem as compared with students who were solving the problem on their own. An advantage of peer feedback over teacher feedback is that students have a better understanding of each other. As students use “similar language”, they are able to mutually clarify the meanings (Kaur et al., 2013). For example, communication transcribed in Excerpt 2 leads to a learning moment when Pedram realised that square metres do not mean that the rooms in the apartment need to be squares (turn 208). Although the

Tab. 2: Transcript Excerpt 2: Example of a dialogue when the peer feedback contributed positively to the problem-solving process (Group 3)

Time	Dialogue	Code
201 [41:08.10]	Pedram: Are we going to have square rooms though? Oh yeah, it's square rooms.	Idea (eliciting)
202 [41:11.02]	Katie: Do you say one room is 12 metres, one off side of the wall would have to be three, the other would have to be three and four and four. He said square metres.	Feedback (ask for clarification)
203 [41:19.00]	Pedram: No, they all have to be three. You can't get 12. That'll be a rectangle. Like you can't have one side of the length and ones on that side of the width. They've...	Feedback (clarification)
204 [41:33.16]	Pedram: If you're having a square room all of the sides are going to be the same.	Eliciting – idea
205 [41:37.13]	Katie: Does it have to be a square?	Feedback (doubt)
206 [41:39.11]	Pedram: Yeah. Square metres.	Eliciting – idea
207 [41:42.17]	Katie: Square metres mean like a metre inside the square. It doesn't mean that the room has to be a square.	Feedback (explanation)
208 [41:47.08]	Pedram: Yeah. Can be a rectangle.	Reaction
209 [41:48.16]	Katie: Yeah.	Reaction (doubt)
210 [41:49.05]	Pedram: But should it – well isn't it easier if it's a square?	Feedback (doubt – question addressed to schoolmates)

feedback that Katie gradually provides (turns 202, 205, 207) was not easy to understand, it eventually makes Pedram aware of an important idea that will remove the obstacle that complicated his process of problem-solving.

In our data we observed that it was difficult for the students to follow the problem-solving process, while assessing the contribution of peers. In the example of communication (Transcript Excerpt 3), two subgroups were created. The girls (Pandit and Anna) tried to solve the dimensions of the whole apartment (turns 301–307 and 310, 311, 316–318). Turn 318 was clearly inspired by a classmate's suggestion. The idea of the girls was then used by the whole group. Meanwhile, the boys (John and Arman) did not follow the girls' exchange and discussion of the area of each room.

Tab. 3: Transcript Excerpt 3: Peer feedback (Group 5)

Time	Dialogue	Code
301 [40:36.19]	Pandit: No. Wait, isn't that – have to times?	Eliciting – question
302 [40:38.18]	Anna: Yeah.	
303 [40:39.16]	Pandit: Twenty times thirty is like 600.	Eliciting (development of the previous idea)
304 [40:41.14]	Anna: Six hundred.	Feedback (confirmation)
305 [40:42.17]	Pandit: It has to be 60.	Reaction (confirmation that the solution is wrong)
306 [40:43.11]	Anna: Yeah. (Sigh)	Feedback (confirmation)
307 [40:45.12]	Pandit: You did it wrongly, that's why.	Reaction.
308 [40:45.12]	John: The area of the... of each room is the same or not same?	Eliciting (question)
309 [40:46.20]	Arman: John. Fifteen times four is 60 and 15 times five is 75.	Eliciting (idea)
310 [40:51.09]	Anna: I don't know.	Feedback (doubt)
311 [40:51.18]	Pandit: I don't care about that... No, wait. Wait.	Reaction
312 [40:53.17]	John: We need care about that – because maybe some rooms are bigger and some rooms are smaller.	Feedback (confirmation)
313 [40:56.20]	Arman: Fifteen times five is 75 and 15 times four is 60.	Reaction
314 [41:00.08]	John: Huh? What?	Feedback (doubt)
315 [41:01.20]	Arman: Fifteen times four is 60, but we need five rooms. Five, five.	Reaction
316 [40:57.19]	Anna: Twenty, three. Twenty- no, no, no.	Eliciting (idea)
317 [40:58.24]	Pandit: My god, you do like 10 times six or something like that.	Eliciting (idea)
318 [41:01.18]	Anna: Okay. Let's just do 10 times six then.	Use

Sometimes although the peer's eliciting was relevant for the solution and the schoolmates reacted to it for a certain period, they did not finish the discussion, or make use of the idea, but shifted to dealing with something else. For example, during the discussion in Transcript Excerpt 4 for Group 3, several suggestions were made about the corridor in the apartment. In the end, the students concluded, that there would be no corridor in the apartment. Pedram was not satisfied with the result and the students' discussion was interrupted.

Tab. 4: Transcript Excerpt 4: Peer feedback (Group 3)

Time	Dialogue	Code
401 [50:39.23]	Pedram: ... Like how do you get out of the house?	Eliciting (question)
402 [50:42.05]	Poya: You need a back door.	Eliciting (idea)
403 [50:43.11]	Katie: There's doors everywhere.	Reaction
404 [50:43.23]	Audrey: Elevator.	Eliciting (idea)
405 [50:44.24]	Katie: The circles are the doors. Like you could label more doors if you want. (Laughter)	Reaction
406 [50:50.19]	Pedram: Okay	Feedback (agreement)
407 [50:51.13]	Katie: Just make sure there's doors –	Eliciting (idea)
408 [50:53.08]	Pedram: But there's no corridor in the house.	Feedback (objection)
409 [50:55.15]	Poya: No. (Over speaking)	Reaction
410 [50:56.04]	Audrey: There doesn't need to be a corridor.	Reaction (use)
411 [50:57.08]	Pedram: How to get out? Why don't you get lost?	Feedback (doubt)

4.4 Solved problem and educational objective of the activity

As already mentioned, the task did not have a specifically stated mathematical goal. The teaching situation was directed to the competence layer, where it supported the instrumentalization of the problem solver's experience (or mathematization) and the competence to solve problems.

In our view, uncertainties in the thematic layer of the problem caused difficulties for Group 3 in the solution process. The communication of the students repeatedly returned to the problem of whether there should be a corridor in the apartment or not. Even in real life, the question of whether the area of the corridor is a part of the whole apartment area differs according to the purpose of the calculation. The group did not manage to solve the question of the corridor in a satisfactory way. An objection to count the corridor as a room was raised multiple times during the student discussion. Nevertheless, they did not reach an agreement. The communication stagnated and did not progress the problem's solution. Transcript Excerpt 5 shows the student's helplessness with the solution of a non-mathematical content

Tab. 5: Transcript Excerpt 5: Non-mathematical problems in communication (Group 3)

Time	Dialogue	Code
501 [46:36.22]	Katie: Okay, fine. If you – if you do this, you can have a back garden. You can just walk out of your back garden to the front. (Laughter)	Eliciting (idea)
502 [46:43.12]	Pedram: Why do we even need a garden?	Feedback (request for clarification)
503 [46:46.13]	Audrey: It's an apartment though. That's okay.	Feedback (objection)
504 [46:48.05]	Pedram: That can – that can even be...	
505 [46:49.17]	Katie: Oh wait, it's an apartment.	Feedback (objection)
506 [46:50.17]	Pedram: Right.	Reaction (positive)
507 [46:51.17]	Katie: Oh my God, I forgot it was an apartment.	Reaction (awareness)
508 [46:54.13]	Pedram: Wait, you know what we can have? We can – we can bike it downstairs.	Eliciting (idea)
509 [46:59.15]	Katie: Mmm hmm.	Feedback (positive)
510 [46:59.18]	Pedram: Oh it's an apartment.	Feedback (confirmation)
511 [47:01.03]	Katie: Yeah. So you don't – you can't get out of the house. Technically, you have to take an elevator down.	Idea (eliciting)
512 [47:04.24]	Pedram: You – you do have to get out of the house.	Feedback (objection)
515 [47:10.18]	Pedram: ... if you can't get out of the house. Just go to prison. They give you free food and free beds.	Idea (eliciting)
516 [47:18.02]	Poya: You go – you go – you – you think you go to jail if you don't get out of the house.	Feedback (confirmation)

of the task. The ideas formulated in turns 501, 508, 511, 515 are unrealistic. The students formulated the feedback in the form of objections. Ambiguities in the thematic layer of the problem seemed to have caused some negotiative events to become a barrier to progressing the solving process.

Nevertheless, the group has produced a sketch in which there was no corridor (Fig. 2).

5 Conclusions and discussion

This section outlines the contributions and limitations of this study to research into student's peer feedback as formative assessment and for problem solving. We first discuss the appropriateness of the EFRU model for describing patterns in students' feedback during problem solving, and evaluate its strengths and weaknesses. We also discuss the benefits of group cooperation in solving problems.

5.1 Strengths and weaknesses of the EFRU model

We propose an EFRU model (eliciting, feedback, reaction, use) as a general pattern of students' communication in a group and the feedback they provide to each other when solving a problem. This model allowed us to describe the course of communication in a group of students, but also to get an idea of how the feedback from classmates helps students to think during the process of solving the task. This is because the model itself also requires determining the intention of the speaker. This means not only describing what the students are saying, but also determining how they are thinking. When coding communication transcripts, we followed how individual students contributed to the communication, and we watched the thinking manifested in their comments. We are aware of at least two weaknesses in this approach. First, the students certainly did not communicate all of their thoughts to their classmates. But we can assume that they communicated what they considered as important. Second, our interpretation of what the student thinks is a postulation that we made when re-reading the communication, viewing the video, and based on the student's written products. It is a postulation that was supported by various types of data. We believe that in future research, these weaknesses could be addressed through post-lesson interviews with students. This approach has been used in the Learner's Perspective Study methodology (in more details in Kaur et al., 2013). But even in that project the success of this approach has not been unequivocally confirmed, because in some cases it was not possible to establish good contact with students.

5.2 Forms of feedback and its influence on the problem-solving process

Feedback appeared very often in the discussion of a group of students and takes the form of positive and negative reactions. In our analyses, we found the following forms of peer feedback:

- negative reaction: rejection, explained or unexplained objection, doubt, question involving doubt;
- positive reaction: confirmation; question asking for clarification, request for clarification, eliciting.

It is not possible to conclude unambiguously from our data that peer feedback helps or does not help to progress the solving process in a productive way. As it was shown in part 4 of this paper, the approach to peer feedback is not consistent across groups or within groups. In some cases, peer feedback resulted in moving the solving process ahead, in other cases this effect did not occur. In our data, the latter was mostly caused by constrained interpersonal relationships among group members. For example, by attempting to work individually without listening to the ideas of others. Willingness and ability to cooperate is not a characteristic of everybody; in a case that the willingness to listen to others' ideas is missing, the peer feedback is unable to alter the student's thinking or actions.

Using a similar language is often considered a great advantage of collaborative work in the groups but there could be inaccuracies in the way in which ideas are expressed. In Excerpt 1 and our comments to it, we have already shown that despite these ambiguities, students seem to be able to understand each other. However, sometimes inaccuracies in language influence the solution process negatively. (For example: In Transcript Excerpt 2 in turn 202, Katie mentions 12 metres, but she means squared metres. She also makes no clear distinction between room dimensions and area. Pedram repeatedly returns to the idea of square rooms. In his comments, the shape of the room and the square unit are confused. But as it was already mentioned the peer communication contributes to Pedram's better understanding of the situation and the final solution of the task.)

5.3 Factors affecting the quality of peer feedback

If we look at feedback as a possible form of assessment, it may be provided (either by the teacher or a classmate) in accordance with the goal of the activity. However, this is difficult when solving an open problem, because its solution usually allows different approaches, that is, the solver can pursue different directions in his problem-solving activity. As we have already stated from the point of view of the long-term goals of education (the competence layer of the problem), the goal of the problem “Fred’s Apartment” should be aimed at strengthening the competence of problem-solving. It is difficult for students to provide peer feedback to support the development of this competence.

The thematic layer of the task is underpinned by the task formulation in the form of a word problem, a mathematical problem set within the framework of a realistic situation.¹ Proposals for solutions that had their basis in this layer appeared. However, the quality of the feedback was diverse because the students did not have enough real life experience of different living spaces. We could observe this especially in Transcript 2 and 4. But the most important, from the point of view of mathematics teaching, is the conceptual layer. Here it is necessary to recall that this was an open problem that can be viewed and solved from various points of view. The quality of feedback may increase if the task were focused on a specific mathematical concept. For example, the problem could be directed to the following aims (expressed in the form of teacher’s questions):

- What would be a suitable way of sketching the plan (of the apartment)?
- What scale was used for the plan? – ratio in a real situation
- What might be the dimensions of an apartment that is suitable for people to live in?
- What are the relationships between the area of a geometrical figure, its dimensions and shape?

This study contributes to the larger *Social Unit of Learning Project* by examining student group work and their social interactions in terms of peer feedback. The project provides a rich source of classroom data in which researchers could apply different perspectives to interrogate the data. Through this interrogation, researchers could also broaden and deepen their understanding of their own perspectives. By examining the student-student conversation in terms of what feedback the students offer each other in the problem-solving process, this study identifies some of the task and interpersonal conditions that necessitate productive problem-solving processes. Future investigations may involve juxtapositioning the different theoretical lenses (for example Brousseau, 1997, Brousseau & Sarrazy, 2002) to identify points of connections and distinctions to enhance the theorisation and analysis of student group work.

Acknowledgment

This research was supported under the Australian Research Council’s Discovery Projects funding scheme (Project number DP170102541). We would like to thank the students, parents, teachers, and school staff for their invaluable support of this project.

References

- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the Black Box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8–21. <https://doi.org/10.1177/003172170408600105>
- Boud, D., & Associates. (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Australian Learning and Teaching Council.
- Boud, D., & Falchikov, N. (2007). *Rethinking assessment in higher education*. Kogan Page.
- Brousseau, G., & Sarrazy, B. (2002). *Glossaire de quelques concepts de la théorie des situations didactiques en mathématiques* [English translation V. Warfield]. DAEST, Université Bordeaux 2.

¹Word problems constitute one of the few school mathematics domains which require mathematization of situations described in words and transformation of the mathematical solution back to the semantic context of the problem. It often happens that the student does not recognize the mathematical model of the problem immediately; one of the tasks given to the student is exactly to discover (or construct) this mathematical model. Word problems also allow development of students’ capacity to use their mathematical knowledge in extra-mathematical domains and in this way they contribute to recognition, understanding and memorising of notions, methods and results from mathematics. (Novotná, 2010)

- Brousseau, G. (1997). *Theory of Didactical situations in mathematics 1970–1990* [Translation from French M. Cooper, N. Balacheff, R. Sutherland, & V. Warfield]. Kluwer Academic Publishers.
- Chan, M. C. E., & Clarke, D. J. (2017). Structured affordances in the use of open-ended tasks to facilitate collaborative problem-solving. *ZDM-The International Journal on Mathematics Education*, 49(6), 951–963. <https://doi.org/10.1007/s11858-017-0876-2>
- Chan, M. C. E., & Clarke, D. J. (2019). Video-based research in a laboratory classroom. In L. Xu, G. Aranda, W. Widjaja & D. Clarke (Eds.), *Video-based research in education: Cross-disciplinary perspectives* (pp. 107–123). Routledge.
- Chan, M. C. E., & Sfard, A. (2020). On learning that could have happened: The same tale in two cities. *Journal of Mathematical Behavior*, 60, Art. 100815. <https://doi.org/10.1016/j.jmathb.2020.100815>
- Chan, M. C. E., Clarke, D. J., & Cao, Y. (2018). The social essentials of learning: An experimental investigation of collaborative problem-solving and knowledge construction in mathematics classrooms in Australia and China. *Mathematics Education Research Journal*, 30(1), 39–50. <https://doi.org/10.1007/s13394-017-0209-3>
- Chan, M. C. E., Ochoa, X., & Clarke, D. J. (2020). Multimodal learning analytics in a laboratory classroom. In M. Virvou, E. Alepis, G. A. Tsihrintzis, & L. C. Jain (Eds.), *Machine learning paradigms: Advances in learning analytics* (pp. 131–156). Springer. https://doi.org/10.1007/978-3-030-13743-4_8
- Chiu, M. M. (2008). Flowing toward correct contributions during groups' mathematics problem-solving: A statistical discourse analysis. *Journal of the Learning Sciences*, 17(3), 415–463. <https://doi.org/10.1080/10508400802224830>
- Clarke, D. J. (2001). Untangling uncertainty, negotiation and intersubjectivity. In D. Clarke (Ed.), *Perspectives on practice and meaning in mathematics and science classrooms* (pp. 33–52). Kluwer Academic.
- Clarke, D. J., & Chan, M. C. E. (2020). Dialogue and shared cognition: An examination of student–student talk in the negotiation of mathematical meaning during collaborative problem-solving. In N. Mercer, R. Wegerif, & L. Major (Eds.), *International handbook of research on dialogic education* (pp. 581–592). Routledge.
- Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research*, 64(1), 1–35. <https://doi.org/10.2307/1170744>
- Cohen, E. G., & Lotan, R. A. (2014). *Designing groupwork: Strategies for the heterogeneous classroom* (3rd ed.). Teachers College Press.
- Damon, W., & Phelps, E. (1989). Critical distinctions among three approaches to peer education. *International Journal of Educational Research*, 13(1), 9–19. [https://doi.org/10.1016/0883-0355\(89\)90013-X](https://doi.org/10.1016/0883-0355(89)90013-X)
- Díez-Palomar, J., Chan, M. C. E., Clarke, D., & Padrós, M. (2021). How does dialogical talk promote student learning during small group work? An exploratory study. *Learning, Culture and Social Interaction*, 30 (Part A), 100540. <https://doi.org/10.1016/j.lcsi.2021.100540>
- Francisco, J. (2013). Learning in collaborative settings: Students building on each other's ideas to promote their mathematical understanding. *Educational Studies in Mathematics*, 82(3), 417–438. <https://doi.org/10.1007/s10649-012-9437-3>
- Gillies, R. (2016). Cooperative learning: Review of research and practice. *Australian Journal of Teacher Education*, 41(3), 39–51. <https://doi.org/10.14221/ajte.2016v41n3.3> – via eric.ed.gov
- Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy and Practice*, 4(3), 365–379. <https://doi.org/10.1080/0969594970040304>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hendl, J. (2005). *Kvalitativní výzkum* [Qualitative research]. Portál.
- Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem-solving skills. In P. Griffin, & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills, Educational Assessment in an Information Age* (pp. 37–56). Springer Science + Business Media. https://doi.org/10.1007/978-94-017-9395-7_2
- Isöhätälä, J., Näykki, P., Järvelä, S., & Baker, M. J. (2018). Striking a balance: Socio-emotional processes during argumentation in collaborative learning interaction. *Learning, Culture and Social Interaction*, 16, 1–19. <https://doi.org/10.1016/j.lcsi.2017.09.003>
- Johnson, D. W., Johnson, R. T., & Holubec, E. J. (1994). *The nuts and bolts of collaborative learning*. Interaction Book Company.
- Johnson, D. W., & Johnson, R. T. (1984). Structuring groups for collaborative learning. *Organizational Behavior Teaching Review*, 9(4), 8–17. <https://doi.org/10.1177/105256298400900404>

- Kaur, B., Anthony, G., Ohtani, M., & Clarke, D. (Eds.). (2013). *Student voice in mathematics classrooms around the world*. Sense Publishers. <https://doi.org/10.1007/978-94-6209-350-8>
- Leikin, R., & Zaslavsky, O. (1999). Cooperative learning in Mathematics. *The Mathematics Teacher*, 92(3), 240–246.
- Lohman, M. C., & Finkelstein, M. (2000). Designing groups in problem-based learning to promote problem-solving skill and self-directedness. *Instructional Science*, 28(4), 291–307.
- Moate, J., Kuntze, S., & Chan, M. C. E. (2021). Students' participation in peer interaction – Use of material resources as a key consideration. *LUMAT: International Journal on Math, Science and Technology Education*, 9(1), 29–55.
- Novotná, J. (2010). *Study of solving word problems in teaching of mathematics. From atomic analysis to the analysis of situations*. Germany: LAP LAMBERT Academic Publishing.
- OECD. (2005). *Formative assessment: Improving learning in secondary classrooms*. Paris: OECD.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. National Academy Press.
- Ross, J. A., & Smyth, E. (1995). Differentiating cooperative learning to meet the needs of gifted learners: A case for transformational leadership. *Journal for the Education of the Gifted*, 19(1), 63–82. <https://doi.org/10.1177/016235329501900105>
- Ruiz-Primo, M. A., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment*, 11(3–4), 237–263.
- Silver, E. A. (1995). The nature and use of open problems in mathematics education: Mathematical and pedagogical perspectives, *Zentralblatt für didaktik der mathematik. International Reviews on Mathematical Education*, 27(2), 67–72.
- Slavík, J. (2003). *Autonomní a heteronomní pojetí školního hodnocení – aktuální problem pedagogické teorie a praxe* [Autonomous and heteronomous assessment – current issue of educational theory and practice]. *Pedagogika*, 40(1), 5–25.
- Slavík, J., Janík, T., Najvar, P., & Knecht, P. (2017). *Transdisciplinární didaktika: o učitelském sdílení znalostí a zvyšování kvality výuky napříč obory* [Transdisciplinary didactics: about teacher sharing of knowledge and increasing the quality of teaching across disciplines]. Masarykova univerzita, Pedagogická fakulta.
- Sluijsmans, D. M. A. (2002). *Student involvement in assessment. The training of peer assessment skills* [Unpublished doctoral dissertation]. Open University of the Netherlands, The Netherlands.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Sage.
- Nieminen, J. H., Chan, M. C. E., & Clarke, D. J. (2021). What affordances do open-ended real-life tasks offer for sharing student agency in collaborative problem-solving? *Educational Studies in Mathematics*. <https://doi.org/10.1007/s10649-021-10074-9>
- Tuohilampi, L. (2018). Analyzing engagement in mathematical collaboration: What can we say with confidence? In E. Bergqvist, M. Österholm, C. Granberg & L. Sumpter (Eds.), *Proceedings of the 42nd Annual Meeting of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 235–242). Sweden: PME.
- Žantovská, I. (2015). *Rétorika a komunikace* [Rhetoric and communication]. Prague: Dokořán.
- Žlábková, I., & Rokos, L. (2013). Pohledy na formativní a sumativní hodnocení žáka v českých publikacích [Formative and summative assessment in Czech publications]. *Pedagogika*, 58(3), 328–354.

Komparace kvality tzv. teacher made testů s didaktickými testy a jejich vliv na úspěšnost žáků: případová studie

Comparison of the quality of the teacher made tests with achievement tests and their influence on the success of pupils: case study

Jaroslav Říčan^{1,*}, Jiří Škoda¹, Viktorie Hermanová¹, Barbora Lanková¹

¹ Pedagogická fakulta, Univerzita J. E. Purkyně, Hoření 13, 400 96 Ústí nad Labem; jaroslav.rican@ujep.cz

Příspěvek se zabývá komparací didaktických testů a tzv. „teacher made“ testů. Hlavním cílem této případové studie je explorace diskrepancí ve skórování žáků na základě toho, je-li test stejného učiva tvořen pedagožkou nebo badatelem. Z toho důvodu autoři zjišťují: (1) kvalitu teacher made testů a didaktických testů vytvořených badateli ve vztahu k zákonitostem tvorby didaktických testů a (2) existenci spojitosti mezi výsledky žáků v rámci jejich procentuálního skórování v teacher made testech a didaktických testech. Za účelem naplnění cílů bylo sestaveno badateli 5 didaktických testů z témat, která žáci probírali ve čtvrtém ročníku 1. stupně ZŠ v předmětech vlastivěda a přírodověda, a byly komparovány s 5 teacher made testy vyhotovených pedagožkou. Studie poukazuje na nedostatečné kvalitativní aspekty teacher made testů (reliabilita, poměr subjektivně a objektivně skórovatelných úloh, poměr úloh vyžadujících nižší a vyšší kognitivní operace, způsob skórování). Přestože korelační analýza odhalila středně vysokou spojitost mezi procentuálním skóre žáků v teacher made testech a didaktických testech, očekávali bychom spíše spojitost blížíící se absolutní hodnotě. Výsledky jsou diskutovány ve vztahu k současnému paradigmatu tvorby didaktických testů, dále je poukázáno na limity studie a zároveň jsou naznačeny potenciální cesty dalšího empirického směřování.

Klíčová slova:

první stupeň ZŠ, přírodověda, vlastivěda, teacher made test, didaktický test.

Zasláno 7/2020

Revidováno 10/2020

Přijato 2/2021

This article compares achievement tests and so-called “teacher-made” tests. The main objective of this case study is to explore and discrepate the results of pupils’ scorings based on whether is test of the same subject has been matter created by the educator or by the researcher. For this reason, the authors ascertain: (1) the quality of teacher-made tests and achievement tests created by the researchers in relation to the regularities of creating achievement tests and (2) the existence of a connection between pupils’ results within their percentage scoring in teacher-made tests and in achievement tests. In order to meet the objectives, the researchers compiled 5 achievement tests on topics that the pupils had been subjected to in the fourth grade of elementary school, during primary science classes. These were then compared to 5 teacher-made tests made by the educator. The study indicates insufficient qualitative aspects of the teacher-made tests (reliability; the ratio of subjectively and objectively scoreable tasks; the ratio of tasks requiring lower and higher cognitive operations; scoring method). Although the correlation analysis revealed a moderately high correlation between the percentage score of pupils in teacher-made tests and achievement tests, we would rather expect a continuity approaching absolute continuity. The results are discussed in relation to the current paradigm of creating achievement tests. The article also points out the limitations of the study and indicates potential ways of further empirical direction.

Key words:

primary education, primary science subject, teacher made test, achievement test.

Received 7/2020

Revised 10/2020

Accepted 2/2021

1 Úvod

Jednou z nejdůležitějších činností učitele je objektivní hodnocení (Junková, 2006), které funguje jako zpětná vazba především pro učitele, ale i pro žáky a jejich rodiče. Učitel z výsledků může zjistit, v jaké míře si žáci osvojili učivo a zda si ho zvládli zařadit mezi dříve získané vědomosti. Je to důležité zvláště v oblasti kognitivní, kdy učitel zjišťuje, jestli žák probíranou látku pochopil (Jeřábek & Bílek, 2010). Získané výsledky učiteli nadále naznačují, zdali si žáci učivo adekvátně osvojili, tedy dosáhli vytyčeného vzdělávacího cíle, a zároveň také, zda jim byl daný obsah vhodně předán, což je právě úkolem didaktického testu (Byčkovský, 1982; Chráska, 2007; Komenda & Mazuchová, 1995; Švamberská Šauerová, 2016). Didaktický test není náhradou zkoušky ústní či praktické, ale je jejím vhodným doplňkem (Schindler, 2006). Měl by být sestavován osobou, která nejenže bude dobrým pedagogem a odborníkem na danou oblast, ale zároveň bude znát základy statistiky (Chráska, 2016). Při jeho konstrukci a vyhodnocování by měl tvůrce vycházet z poznatků didaktické teorie a specifického vzdělávacího kontextu (Cizek, 2004).

Kvalita didaktického testu klade na učitele vysoké nároky. Ve školní praxi se setkáváme s názvem nestandardizovaný didaktický test, tedy takový test, který je sestaven dostatečně odborně, ale nejsou zde splněny veškeré náležitosti, a nelze proto ověřit všechny vlastnosti testu (Kalhous & Obst, 2002; Skutil, 2011). Dále k nestandardizovanému didaktickému testu budeme referovat v terminologii Škody a Doulíka (2007) jako k tzv. teacher made testu. Každý učitel by měl být schopný vytvořit kvalitní didaktický test, dokázat ho vyhodnotit a analyzovat jeho kvalitu, přesto v České republice není situace ohledně tvorby testů uspokojivá. Chráska (1999) popisuje, že v názorech mnoha pedagogů přetrvává nedůvěra a skepse ohledně testů a měření v pedagogice. Spousta učitelů nemá dostatečné zkušenosti s tvorbou didaktických testů a nevědomují si, jaké parametry musí kvalitní didaktický test splňovat. Pokud chceme, aby zjišťování výsledků bylo dostatečně spolehlivé a platné, musí být při konstrukci testu použity osvědčené postupy. V praxi se často setkáváme s případy, kdy sestavené testy vyžadují pouze namemorované poznatky a jejich následné hodnocení probíhá konzervativním způsobem známkování (Čapek, 2015). Tento přístup je minimálně problematický.

Cílem této studie je popsat parametry pěti vytvořených teacher made testů jedné pedagožky a pěti didaktických testů vytvořených badateli ke stejnému učivu. Zároveň chceme poukázat na sílu spojitosti v žákovském skórování v rámci procentuálního výsledku z teacher made testů a didaktických testů.

2 Historická kontinuita a současný stav testování

Historie testování ve školní praxi sahá podle Vrány (1948) do 16. století (jezuitské školy). Na našem území můžeme sledovat kořeny didaktického testování ve druhé polovině 19. století (Jeřábek & Bílek, 2010). O rozvoj metod objektivního hodnocení pro srovnání vzdělávacích výsledků se zasloužil Joseph Mayer Rice z USA, který roku 1894 navrhl a vyzkoušel metodu používání objektivních měřítek při nácviku pravopisu, po níž mohly být výsledky srovnány s dalšími školami (Fayol et al., 2012). Škoda, Doulík a Hajerová-Müllerová (2006) jako další důležitý mezník považují rok 1908, kdy C. W. Stone sestavil v USA první standardizovaný test, což způsobilo velký rozvoj tvorby didaktických testů. Příznivá situace byla pro rozvoj české (československé) pedagogické diagnostiky po první světové válce, a to hlavně díky Václavu Příhodovi, který didaktické testy považoval jako nástroj hodnocení a do školního zkoušení zahrnoval objektivitu, srovnatelnost zkoušek a výkonnost. Další významný pedagog té doby, Otokar Chlup, vystupoval proti didaktickým testům, které mu připadaly neúčelné. Byl názoru, že školství nepodporuje tvořivé myšlení žáků.

Mechanismus této metody [myšleno bylo didaktické testování] nemůže správně hodnotit ani soustavu vědění a myšlení žáka, ani jednotlivých poznatků, jejichž důležitost odhadnouti jest koneckonců vždy věcí subjektivního posouzení. Měřením odvrací se žactvo od vážného myšlení a tvořivá činnost žáka, rozvíjející se pod vedením učitelovým, ubíjí se v bičovaném jezuitském závodění. (Chlup, 1931, s. 56)

Spory O. Chlupa a V. Příhody o didaktické testy se projevovaly ve dvou rovinách. V té první vedl spor ke zdokonalování didaktických testů a hledání nových způsobů měření a hodnocení žákových výstupů, zatímco ve druhé vedl k nepoužívání didaktických testů. S rozvojem empirického bádání v kontextu kvality psychometrických vlastností testů, nástrojů a přístupů, akcentem na komparaci výsledků z mezinárodních srovnávacích testů (PIRLS, TIMSS) či problematikou přijímacího řízení na střední a vysoké školy je nepochybné, že didaktický test „hraje“ jednu z ústředních rolí v pedagogickém dění, a to v rovině teorie i praxe.

3 Kvalita didaktického testu

Didaktický test má řadu funkcí. Popham (2017) popisuje tři hlavní účely pedagogického testování, kterými je srovnávání mezi testovanými, zlepšování výuky a hodnocení výuky. Hališka (1999) hovoří o funkci diagnostické, motivační a stimulační, klasifikační, kontrolní a prognostické. Nesmíme zapomenout na funkci formativní, která je v současné době zřejmě nejvíce akcentována (Laufková & Starý, 2016). Jak však může jakýkoliv test naplňovat tyto funkce, když panují pochybnosti o jeho kvalitě (psychometrické vlastnosti testu)? Termín „didaktický test“ je definován s nuancemi. Podle Chráska (1999, s. 12) didaktický test „měří hlavně vědomosti a dovednosti žáků.“ Byčkovský (1982) nazval didaktický test nástrojem systematického zjišťování, při kterém se zjišťují výsledky výuky. Za nejbližší termín k pojmu didaktický test užívaný v zahraniční literatuře považujeme označení *achievement test*, který je definován jako nástroj určený k měření relativního úspěchu v dané oblasti prostřednictvím jednoho ze dvou typů, mezi které patří testy zjišťující úspěch a testy diagnostické (Hawkes et al., 1936; Ward et al., 1996).

Škoda a Doulík (2007, s. 11) popisují test jako „zkoušku, jejíž podmínky jsou pro všechny testované jedince shodné a jejíž výsledky mají číselný charakter.“ Z této závěrečné definice excerptujeme první ze základních vlastností didaktického testu: test by měl být maximálně (1) objektivní. Objektivitu testu lze dle Smékala, Švece a Zajace (1973) zajistit tím, že bude daný test sestaven z otázek s možností jednoznačné odpovědi, nadále je klíčové, aby hodnotitel mohl u jednotlivých odpovědí rozhodnout bez pochyb o správnosti odpovědi, a v posledním případě je nezbytné, aby výkon žáka byl posuzován a interpretován na základě systému norem. Klíčová je nadále jednoznačná formulace testových úloh a při jejich vyhodnocování by výsledky neměly být ovlivněny postoji a názory hodnotitele (Linn, 2008; Schindlera, 2006). Výše uvedeným kritériím odpovídají z větší části úlohy objektivně skórovatelné, k jejichž převaze při konstruování testu se přiklání i někteří odborníci (Pulpán, 1991; Škoda & Doulík, 2007). Při tvorbě didaktických testů se badatelé drželi co možná největší míry objektivity rozložení testových položek (92 % položek bylo objektivně skórovatelných). Někteří autoři opírají míru objektivity také o Bloomovu taxonomii a rozdělení jednotlivých úloh dle jejich cílení na využívání vyšších a nižších kognitivních operací (Anderson et al., 2000).

Na základě aktuálního paradigmatu tvorby didaktického testu uvádíme další vlastnosti, kterými jsou (Chráška, 1999): (2) reliabilita, přesnost a spolehlivost, kdy se výsledky testů musí co nejméně lišit od skutečných hodnot. Výsledky didaktického testu jsou tvořeny dvěma složkami, a to pevnou složkou, ve které se jedná o skutečné vědomosti, a náhodnou složkou, kdy na žáka působí vnější činitelé, jako je stav žáka, prostředí ve třídě, hluk z ulice apod. U správně vypracovaného didaktického testu by neměl být vliv náhodné složky vysoký, a pokud test poskytuje výsledky, na které mají vnější činitelé minimální vliv, mluvíme o testu s vysokou reliabilitou. Pokud chceme posuzovat míru reliability, slouží nám k tomu tzv. koeficient reliability, který v praxi nabývá hodnot od 0 (naprostá nespolehlivost) až po hodnoty, které se blíží k číslu 1 (naprostá spolehlivost). Existuje řada přístupů pro zjišťování reliability (split half, opakované měření, ...), mnohé z nich souvisí s charakterem testové položky. V případě binárně skórovatelných položek didaktického testu používáme Kuder-Richardsonův test, v případě, že testové položky jsou skórovatelné v intervalu (0; 1) a zároveň je prokázána jednofaktorová dimenzionalita, můžeme použít Cronbachovo alfa (Urbánek, 2002). Používat binární skórování (každá položka je hodnocena jedním bodem, nezávisle na obtížnosti dané položky) je doporučováno autory Škodou, Doulíkem a Hajerovou-Müllerovou (2006). Opakem je tzv. vážené skórování, kdy se jednotlivým položkám v testu přiřazuje různé bodové ohodnocení. Vážené skórování se ve školní praxi užívá častěji, své opodstatnění však má pouze tehdy, vyžadují-li některé úlohy na své řešení výrazně více času než ostatní úlohy (rovněž Pulpán, 1991). Podobně se vyjadřuje Chráška (2002), že „široké testové úlohy se navrhuji poměrně snadno, ale jejich hlavní nevýhodou je nemožnost objektivního skórování“ (s. 183). V teacher made testech vyhotovených pedagožkou ani v didaktických testech vytvořených badateli se nevyskytovaly položky, jejichž vyplnění by zabralo žákům významně více času než vyplňování položek jiných (jako v případě položek otevřených se strukturou nevymezenou – Škoda et al., 2006; podobně tzv. široké testové úlohy – Chráška, 2002). V tomto případě proto akcentujeme binární skórování.

Na základě faktorové analýzy se v této studii nepotvrdila jednodimenzionalita umožňující užití Cronbachovo alfy, a tak bylo využito Split-half reliability, kdy se testové položky rozdělí na dvě přibližně stejné části. Tento způsob výpočtu reliability je u didaktických testů velmi rozšířený (Johnson & Penny, 2005; Marinova et al., 2005). Nevýhoda podstatného snížení hodnoty koeficientu korelace u Split-half reliability se koriguje použitím Spearman-Brownova vzorce (Johnson & Penny, 2005), který stanovuje reliabilitu pro celý nezkrácený test. (3) Obsahová validita (testová doména) je shoda mezi obsahem testu a obsahem výuky, tj. poměr počtu položek v didaktickém testu reprezentující subtémata okruhu, pro který je test určen. Prakticky to znamená, že pokud učitel věnoval např. subtématu č. 1 30 % výuky, mělo by být i 30 % testových položek konstruováno k tomuto obsahu. Tento aspekt byl při tvorbě didaktických testů problematizován dvěma faktory: (i) při konstrukci jsme výhradně vycházeli ze zaznamenaných informací žáků (výpisky v sešitu, odkazy na domácí úkoly v pracovním sešitu) a nebylo tak reflektováno realizované kurikulum (reálná výuka) a (ii) některé testy žáci absolvovali po několika hodinách. Tuto záležitost blíže řešíme v empirické části tohoto textu. Disman (2005) uvádí, že (4) konstruktová validita referuje k tomu, jaké pedagogické nebo psychologické konstrukty jsou vlastně diagnostikovány (vyjádření míry vztahu mezi nástrojem – testem a teoretickým konstruktem – v našem případě myšlenkové úrovně dle klasifikace Blooma). V kontextu didaktického testování se jedná o to, zdali se diagnostikují vyšší nebo nižší kognitivní operace. Jestliže učitel věnoval subtématu č. 1 30 % výuky, mělo by být 20 % konstruovaných položek na vyšší a 10 % na nižší úrovni kognitivních operací (poměr 2 : 1). Na tento způsob konstrukce testových položek v didaktickém testu poukazují například autoři Škoda, Doulík a Hajerová-Müllerová (2006) nebo Junková (2006), která dále uvádí možné využívání specifikační tabulky či techniky seznamu výukových cílů.

Za další vlastnost didaktického testu můžeme považovat senzibilitu (citlivost). Test musí být pro žáka rovněž přiměřený, jelikož příliš obtížné, či příliš snadné úlohy, by ho objektivně nehodnotily (hodnota

obtížnosti úloh). Závěrem můžeme konstatovat, že didaktický test se od jiných diagnostických způsobů (hlavně v komparaci s teacher made testy) liší hlavně daným metodologickým postupem s cílem zajištění co možná nejvíce optimálních vlastností s akcentem na reliabilitu, validitu a objektivitu (Marinova et al., 2005).

4 Cíle, výzkumné problémy, předpoklady a hypotéza

Cílem této studie je popsat parametry pěti vytvořených teacher made testů jedné pedagožky a pěti didaktických testů vytvořených badateli ke stejnému učivu. Zároveň chceme poukázat na sílu spojitosti v žákovském skórování v rámci procentuálního výsledku z teacher made testů a didaktických testů. Jelikož se studie zabývá jevem, který je běžný v pedagogické praxi, tedy konstrukcí testů, a zaměřuje se na práci jednoho učitele, můžeme hovořit o deskriptivní (popisující) případové studii (Mareš, 2015). Johanson (2003) uvádí, že jedním z možných způsobů usuzování v rámci případové studie je prostřednictvím verifikace hypotéz testovat teorii. Z toho důvodu jsme na základě vytyčených cílů projektovali jak deskriptivní, tak i relační výzkumný problém (VP).

VP₁ (deskriptivní): Jaké jsou vlastnosti teacher made testů v komparaci s aktuálním paradigmatem tvorby didaktických testů?

V souvislosti s výše uvedeným deskriptivním výzkumným problémem jsme vytvořili následující předpoklady (P):

P₁: Teacher made testy neodpovídají současnému paradigmatu tvorby didaktických testů ve vztahu k doporučenému počtu testových položek.

P₂: Teacher made testy neodpovídají současnému paradigmatu tvorby didaktických testů ve vztahu k doporučenému poměru objektivně a subjektivně skórovatelných testových položek.

P₃: Teacher made testy neodpovídají současnému paradigmatu tvorby didaktických testů ve vztahu k doporučenému poměru testových položek na vyšší a nižší kognitivní operace.

P₄: Teacher made testy neodpovídají současnému paradigmatu tvorby didaktických testů ve vztahu k doporučenému způsobu skórování (úlohy jsou hodnoceny jiným počtem bodů).

VP₂ (relační): Jaká je spojitost mezi výsledky žáků v rámci jejich procentuálního skórování v teacher made testech a didaktických testech?

V souvislosti s výše uvedeným relačním výzkumným problémem jsme vytvořili následující hypotézu (H):

H (relační): Spojitost mezi výsledky žáků v rámci jejich procentuálního skórování v teacher made testech a didaktických testech je pozitivní.

5 Nástroje, procedura a výzkumný vzorek

Za účelem naplnění vytyčených cílů sestavili badatelé (autoři této studie) celkem pět didaktických testů z témat, která žáci probírali ve čtvrtém ročníku 1. stupně základní školy (Sedláčková, 1993). Autoři při tom výhradně čerpali ze zaznamenané práce v hodině (tj. výpisků žáků v sešitech, včetně odkazů na domácí cvičení). Uvědomujeme si, že vztahové kritérium „zaznamenaná práce v hodině“ pro konstrukci didaktických testů je nekompletním přístupem, jelikož nezahrnuje kritérium „skutečná práce v hodině“. Jedná se o nedostatek, který si autoři tohoto příspěvku uvědomují, avšak dlouhodobé záměrné pozorování bylo pedagožkou odmítnuto. Testy byly konstruovány pro dva vyučovací předměty, a to vlastivědu a přírodovědu. Vzdělávací obsah těchto předmětů lze snadno strukturovat, vykazuje vysokou míru integrace¹ poznatků různých vědních disciplín, lze ho posuzovat s ohledem na mezinárodně platné vymezení přírodovědné gramotnosti. K danému vzdělávacímu obsahu se snadno vytvářejí úlohy vyžadující různé úrovně osvojení poznatků a při konstrukci a hodnocení těchto úloh se lze snadno inspirovat pravidelnými mezinárodními srovnávacími studii (např. TIMSS).

V rámci prezentované výzkumné studie se jednalo o následující témata: a) Má vlast (vlastivěda); b) Živá a neživá příroda a společné znaky rostlin (přírodověda); c) Orientace v krajině (vlastivěda); d) Rozmnožování rostlin (přírodověda); e) Stavba rostliny, kořen, stonk (přírodověda).

Každý z didaktických testů obsahoval vždy deset položek, což je podle Chrásky (1999) minimální počet pro zajištění akceptovatelné reliability testu, pokud bereme v potaz chybu měření. V testu se vždy nacházely rozmanité typy testových položek (blíže k typologii Kalhous & Obst, 2002). Žáci v dané škole nebyli informováni o způsobu vyhodnocení testů, aby nevznikaly nejasnosti při započítání do běžné klasifikace. Uvědomujeme si, že způsob skórování by měl být žákům zpřístupněn, jelikož

¹Posuzování testů společně pro vlastivědu a přírodovědu vnímáme za legitimní (viz vzdělávací oblast Člověk a jeho svět – RVP, 2017, s. 42).

... pokud máme informace o formě testu (zkoušky), přizpůsobíme tomu svou strategii učení, vyčleníme si čas na učení, nebo zpracujeme zpětnou vazbu efektivněji, než když tyto informace o podobě testu (zkoušky) nemáme. (Dutke et al., 2010, s. 195)

Ve výzkumu (zadávání didaktických testů autory tohoto příspěvku žákům) jsme však zachovali konzistenci s přístupem pedagožky (žákům nezpřístupňuje bodové vyhodnocování testových položek). Žákům bylo řečeno, že se jedná o pracovní list, po jehož vyplnění obdrží zpětnou vazbu. Žáci vždy nejprve vypracovávali teacher made test se svou pedagožkou, a poté vyhotovili didaktický test. Didaktické testy byly vyplněny v těchto datech: 11. 10. 2019 (a) dva dny po testu pedagožky; 8. 11. 2019 (b) den po testu pedagožky; 15. 11. 2019 (c) dva dny po testu pedagožky; 18. 11. 2019 (d) čtyři dny po testu pedagožky; 2. 12. 2019 (e) čtyři dny po testu pedagožky. Časový odstup mezi vyplňováním obou typů testů je samozřejmě intervenující proměnnou. Odstup byl způsoben dvěma faktory: (1) bylo zapotřebí přijmout disponibilní možnosti pedagožky pro zadávání didaktických testů (kdy a v jaké hodině proběhne zadávání tak, aby nenarušilo „chod“ výuky) a (2) vypracování didaktického testu bezprostředně po vyhotovení testu zadaného pedagožkou by nemuselo být k žákům přiměřené (únava, kognitivní vyčerpání). Zadávání teacher made testů ani didaktických testů nebylo realizováno první ani poslední vyučovací hodinu daného dne. Výzkum byl realizován po dobu 5 měsíců (9/2019–1/2020) v jedné 4. třídě středočeského kraje. Přestože se jednalo pouze o dostupný výběr, tedy jednoho pedagoga a jednu 4. třídu ZŠ, jak uvádí Flybvjerg (2006), i na základě jednoho individuálního případu (případové studii) lze dojít k určitému rozvoji vědeckého poznání.

6 Analýza a zpracování dat

6.1 Didaktické testy

Abychom zajistili co nejvyšší platnost výzkumných zjištění, bylo zapotřebí nejprve učinit analýzu vlastností vyhotovených didaktických testů, jakožto vztahových kritérií při komparaci s teacher made testy. Reliabilita byla nejprve určena na základě relativní velikosti chyby měření, která se zvyšuje snižujícím se počtem úloh. Z toho důvodu všech pět didaktických testů obsahovalo deset testových položek. Dále za účelem stanovení spolehlivosti didaktických testů byla použita metoda Split-half reliability a korigována použitím Spearman-Browna vzorce. Hodnota reliability pro jednotlivé didaktické testy činila: 0,75; 0,705; 0,667; 0,928; 0,904. Obecně přijatelné hodnoty reliability jsou stanoveny v intervalu 0,70–0,60 (Sekaran, 1992) a 0,95 (Cronbach & Meehl, 1955; Tavakol & Dennick, 2011). Z hlediska spolehlivosti se tak všechny didaktické testy jeví jako dostatečně spolehlivé. Poměr objektivně vs. subjektivně skórovatelných úloh činil 10 vs. 0 (did. test č. 1) a 9 vs. 1 (did. testy 2–5). Pro účely určení charakteru testových položek na základě jejich kognitivní náročnosti jsme podnikli doplňující šetření, ve kterém jsme využili tři expertní² posudky (stejný přístup jsme podnikli u teacher made testů – viz navazující oddíl).

Tab. 1: Expertní posouzení poměru testových položek. Poměr položek vyšší : nižší kognitivní náročnosti pro jednotlivé didaktické testy

	Did. test č. 1	Did. test č. 2	Did. test č. 3	Did. test č. 4	Did. test č. 5
Expert 1	1 : 9	4 : 6	3 : 7	5 : 5	4 : 6
Expert 2	2 : 8	5 : 5	3 : 7	5 : 5	5 : 5
Expert 3	2 : 8	5 : 5	2 : 8	5 : 5	2 : 8
Expert 4	1 : 9	5 : 5	3 : 7	4 : 6	2 : 8
Expert 5	2 : 8	4 : 6	2 : 8	5 : 5	3 : 7
Me	2 : 8	5 : 5	3 : 7	5 : 5	3 : 7

Zdroj: autoři

Výše uvedené expertní posouzení nekoresponduje s dostatečným poměrem úloh na vyšší a nižší kognitivní operace (2 : 1 ve prospěch úloh na vyšší kognitivní operace). Autoři si uvědomují, že tento fakt bude mít dopad na celkovou platnost zjištěných nálezů, avšak: a) teacher made testy byly zadávány po výuce (podle našeho názoru) nízkého počtu realizovaných hodin (test č. 1: 3 vyučovací hodiny; testy č. 2 a 5: 4 vyučovací hodiny; testy č. 3 a 4: 2 vyučovací hodiny), což problematizovalo konstrukci položek v didaktickém testu (a to z hlediska konstruktové i obsahové validity) a b) při konstrukci didaktických testů jsme vycházeli výhradně ze zaznamenané práce žáky v hodině a z tohoto důvodu jsme čelili dilematu

²Experta jsme definovali jako osobu s vzdělaností úrovní ISCED 8 připravující budoucí učitele prvního stupně v oblasti obecné didaktiky nebo didaktiky přírodovědných předmětů (příp. kombinací těchto předmětů).

vytvoření teoreticky žádoucí úlohy vyšší kognitivní náročnosti, avšak bez písemného dokladu o tom, jestli je vůbec možné od žáků očekávat její vypracování.

Dále jsme posuzovali testové položky z hlediska jejich obtížnosti (p) a citlivosti (d). Vhodné úlohy nabývají indexu obtížnosti $p = \langle 20; 80 \rangle$, za podezřelé považujeme ty s hodnotami $p < 20$ (velmi obtížné) a $p > 80$ (velmi snadné) a zakázané jsou ty úlohy, ve kterých se p blíží k 0 (Škoda et al., 2006). Chráska (1999) rovněž označuje úlohy s indexem obtížnosti s $p > 80$ za podezřelé, avšak ty s $p < 20$ za zakázané. Pro vyhodnocení indexu obtížnosti jsme adoptovali přístup Chrásky (1999). Index obtížnosti určuje tu procentuální část celkového počtu žáků, kteří mají úlohu vyřešenou správně. Počítáme ho podle vzorce

$$p = 100 \cdot \frac{n_s}{n},$$

kde p je index obtížnosti, n_s jsou žáci, kteří odpovídali správně a n představuje celkový počet testovaných žáků.

Hodnota citlivosti pro vhodné úlohy má být $d > 0,25$ pro $p = \langle 30; 70 \rangle$ a $d > 0,15$ pro $p = \langle 20; 30 \rangle$ a $\langle 70; 80 \rangle$. Podezřelé hodnoty proto nabývají hodnot $d = \langle 0; 0,15 \rangle$ až $\langle 0; 0,25 \rangle$ a zakázané úlohy jsou ty se zápornou hodnotou ($d < 0$).

Koeficient citlivosti ULI počítáme podle vzorce

$$d = \frac{n_L - n_H}{0,5n},$$

kde d je koeficient citlivosti ULI, n_L je počet žáků s vyšším skórem (lepší polovina z celkového počtu testovaných žáků), kteří mají úlohu řešenou správně, n_H je počet žáků s horším skórem (horší polovina z celkového počtu testovaných žáků), již mají úlohu řešenou správně a $0,5n$ představuje polovinu všech testovaných žáků.

Tab. 2: Přehled vhodných, podezřelých a zakázaných úloh na základě určení indexu obtížnosti a citlivosti

	Test č. 1	Test č. 2	Test č. 3	Test č. 4	Test č. 5	Celkem
Vhodné úlohy (p)	2	5	4	7	2	20
Podezřelé úlohy (p)	6	5	6	3	8	28
Zakázané úlohy (p)	2	0	0	0	0	2
Vhodné úlohy (d)	4	5	4	9	4	26
Podezřelé úlohy (d)	5	5	6	1	6	23
Zakázané úlohy (d)	1	0	0	0	0	1

Zdroj: autoři

Autoři si uvědomují relativně vysoký počet podezřelých úloh (velmi snadných; $p > 80$) objevujících se ve vytvořených didaktických testech. Autorům však není známa publikace, která by definovala jejich přijatelný počet (poměr).

6.2 Teacher made tests

Abychom odpověděli na vyřčené předpoklady, v tabulce 3 uvádíme analýzu teacher made testů ve vztahu k (1) počtu testových položek (P_1), (2) poměru objektivně a subjektivně skórovatelných položek (P_2), (3) poměru testových položek na vyšší a nižší kognitivní operace (P_3) a (4) způsobu skórování (P_4).

6.3 Syntéza: komparace testů

Abychom odpověděli na druhý výzkumný problém korelačního charakteru: „Jaká je spojitost mezi výsledky žáků v rámci jejich procentuálního skórování v teacher made testech a didaktických testech?“, využili jsme korelační analýzu.

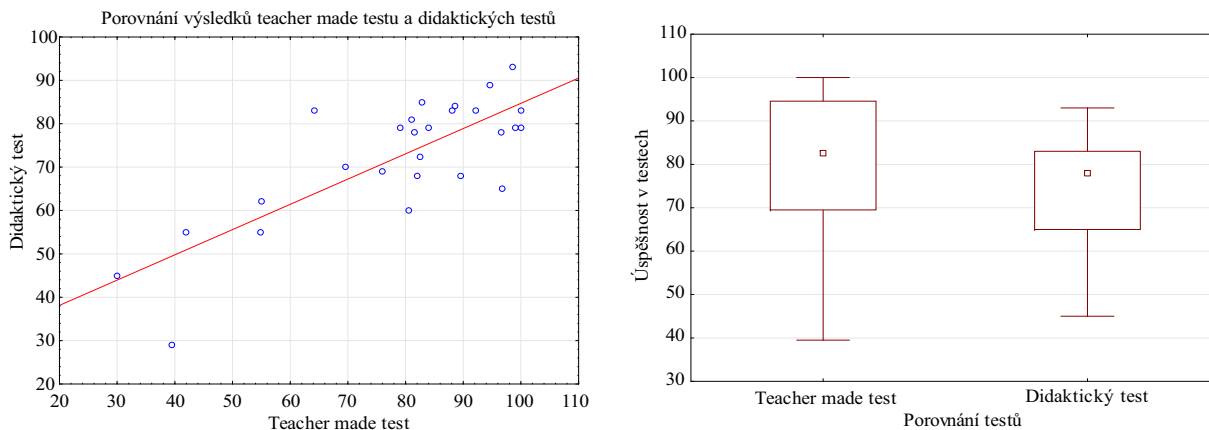
Z toho důvodu, že každý z teacher made testů byl tvořen jiným počtem položek (test č. 1: 8 úloh, test č. 2: 6 úloh, test č. 3: 7 úloh, test č. 4: 4 úlohy, test č. 5: 6 úloh), nemůžeme uvažovat o rovnoměrném rozestupu (typické pro metrická data). Rovněž si uvědomujeme, že taktó vyvozovat z výsledků žáků průměrné procentuální skóre je a priori z matematického hlediska nesmyslné (průměrujeme výsledky testů, přičemž v každém z nich existovaly jiné rozestupy v důsledku rozmanitého počtu úloh a jejich skórování). Na druhou stranu tato studie odráží problematiku praxe a věříme, že výsledky přináší důležitý vhled a podnět do otázky konstrukce testů pedagogy. Také si uvědomujeme, že počty respondentů pro statistické zpracování dat induktivním přístupem jsou hraniční (1 pedagog, 27 žáků). Normalitu dat jsme tedy z důvodu rozdílných rozestupů v kontextu skórování a počtu úloh v případě teacher made testů neověřovali a rovnou jsme přistoupili k užití neparametrického Spearmanova koeficientu pořadové korelace.

Tab. 3: Základní deskripce parametrů teacher made testů

	Test č. 1	Test č. 2	Test č. 3	Test č. 4	Test č. 5
Počet úloh	8	6	7	4	6
Poměr objektivně vs. subjektivně skórovatelných úloh	1 vs. 7	2 vs. 4	0 vs. 7	0 vs. 4	0 vs. 6
Poměr úloh na vyšší: nižší	0 : 8 × 0 : 8×	0 : 6 × 1 : 5×	1 : 6 × 0 : 7×	1 : 3 × 0 : 4×	1 : 5 × 1 : 5×
kognitivní operace	0 : 8	1 : 5	1 : 6	0 : 4	1 : 5
	Me : 0 : 8	Me : 1 : 5	Me : 1 : 6	Me : 0 : 4	Me : 1 : 5
Způsob skórování	1 bod: 7 úloh 7 bodů: 1 úloha	1 bod: 5 úloh 2,5 bodů: 1 úloha	1 bod: 3 úlohy 2 body: 4 úlohy	1 bod: 3 úlohy 2 body: 1 úloha	1 bod: 6 úloh
Procentuální převod (%) a klasifikace	100 – 90 = 1 89 – 79 = 1 – 2 78 – 68 = 2 67 – 57 = 2 – 3 56 – 46 = 3 45 – 54 = 3 – 4 34 – 24 = 4 23 – 13 = 4 – 5 12 – 0 = 5	100 – 93 = 1 92 – 85 = 1 – 2 84 – 70 = 2 69 – 62 = 2 – 3 61 – 54 = 3 53 – 45 = 3 – 4 44 – 37 = 4 36 – 29 = 4 – 5 28 – 0 = 5	100 – 96 = 1 95 – 85 = 1 – 2 84 – 65 = 2 64 – 54 = 2 – 3 53 – 45 = 3 44 – 33 = 3 – 4 32 – 24 = 4 23 – 15 = 4 – 5 14 – 0 = 5	100 – 90 = 1 89 – 71 = 1 – 2 70 – 61 = 2 60 – 51 = 2 – 3 50 – 41 = 3 40 – 31 = 3 – 4 30 – 21 = 4 20 – 11 = 4 – 5 10 – 0 = 5	100 – 93 = 1 92 – 68 = 1 – 2 67 – 60 = 2 59 – 51 = 2 – 3 50 – 40 = 3 39 – 28 = 3 – 4 27 – 16 = 4 15 – 8 = 4 – 5 7 – 0 = 5

Zdroj: autoři

Hodnota Spearmanova korelačního koeficientu byla $r = 0,662$ a koeficient determinace $r^2 = 43,8 \%$. Ačkoliv je možné nulovou hypotézu o nulovém korelačním koeficientu zamítnout na jednoprocentní hladině významnosti ($p = 0,0002$), jedná se stále pouze o středně silnou asociaci (Hendl, 2012; Chráska, 2016). Vzhledem k charakteru testů bychom očekávali spíše vysokou až absolutní korelaci (přímou úměrnost). Ke stejnému závěru bychom došli také na základě Kendalova tau ($\tau = 0,501$). Srovnání jednotlivých testů je dobře patrné z korelačního grafu (zároveň data také vizualizujeme prostřednictvím kvartilového grafu, přestože pro tyto účely nebývá primárně využíván).



Obr. 1: Porovnání teacher made testu a didaktického testu na základě bodového a kvartilového grafu

Z bodového grafu je patrné, že závislost nebude silná, jelikož vykreslené hodnoty se značně odchyľují od přímky (lineární interpolace), a to bez ohledu na to, jakou analytickou křivku využijeme. Z kvartilového grafu je dobře patrné rozložení obou měření. V případě didaktického testu dochází k posunu na vertikále směrem dolů a „ořezání“ krajních hodnot (velmi dobrého a velmi špatného výsledku žáka).

Tab. 4: Srovnání typů testů. V tabulce uvádíme počet získaných procent každého z testů. V závorkách je u teacher made testů uvedena klasifikace pedagožky. První sloupec: teacher made test, druhý sloupec: didaktický test. Symbol „×“ vyjadřuje neúčast žáka při vyplňování testu. Závěrečný sloupec prezentuje průměrné procentuální skóre žáků z obou typů testů (TMT = teacher made test; DT = didaktický test)

Žák	Test č. 1		Test č. 2		Test č. 3		Test č. 4		Test č. 5		Ø skór	
	TMT	DT	TMT	DT	TMT	DT	TMT	DT	TMT	DT	TMT	DT
Dívka 1	82,86 (1)	76	86,67 (1-2)	76,7	55 (2-3)	80	90 (1)	95	100 (1)	96,7	82,91	85
Dívka 2	64,29 (2-3)	60,7	86,67 (1-2)	80	82 (2)	60	80 (1-2)	65	100 (1)	96,7	82,59	72,4
Chlapec 1	92,86 (1)	70	100 (1)	100	100 (1)	87,5	90 (1)	×	100 (1)	96,7	94,57	89
Dívka 3	96,43 (1)	64	86,67 (1-2)	80	100 (1)	100	100 (1)	65	100 (1)	80	96,62	78
Chlapec 2	64,23 (3)	64	66,67 (2-3)	80	45 (3)	45	80 (1-2)	68,6	91,67 (1-2)	90	69,51	70
Chlapec 3	71,43 (2)	62	46,67 (3-4)	60	45 (3)	30	30 (4)	61,4	16,66 (4)	63,3	41,95	55
Dívka 4	85,71 (1-2)	76	86,67 (1-2)	70	×	85	80 (1-2)	88,6	100 (1)	93,3	88,10	83
Chlapec 4	64,29 (2-3)	76	86,67 (1-2)	90	100 (1)	80	100 (1)	85	91,67 (1-2)	90	88,53	84
Dívka 5	×	32	40 (4)	50	41 (3-4)	70	40 (3-4)	45,7	0 (5)	25	30	45
Chlapec 5	64,29 (2-3)	64	66,67 (2-3)	86,7	95 (1-2)	70	100 (1)	71,4	×	100	81,49	78
Dívka 6	85,71 (1-2)	66	93,33 (1)	93,3	100 (1)	65	90 (1)	92,1	91,67 (1-2)	100	92,14	83
Dívka 7	100 (1)	72	×	70	100 (1)	80	70 (2)	22,9	58,33 (2-3)	96,7	82	68
Chlapec 6	71,43 (2)	60	80 (2)	76,7	91 (1-2)	45	80 (1-2)	30	×	86,7	80,61	60
Dívka 8	71,43 (2)	64	86,67 (1-2)	96,7	100 (1)	50	90 (1)	40	100 (1)	90	89,62	68
Chlapec 7	×	×	×	80	100 (1)	90	100 (1)	80	×	66,7	79	79
Dívka 9	78,57 (1-2)	60	73,33 (2)	60	64 (2-3)	80	80 (1-2)	70	66,66 (2)	70	100	79
Dívka 10	50 (3)	60	86,67 (1-2)	90	100 % (1)	70	×	28,6	66,66 (2)	96,7	76	69
Chlapec 8	92,86 (1)	68	100 (1)	100	100 (1)	100	100 (1)	98,6	100 (1)	100	98,57	93
Chlapec 9	100 (1)	80	93,33 (1)	83,3	100 (1)	75	100 (1)	67,1	100 (1)	90	99	79
Chlapec 10	100 (1)	86	100 (1)	100	100 (1)	85	100 (1)	48,6	100 (1)	96,7	100	83
Dívka 11	92,86 (1)	62	93,33 (1)	90	73 (2)	72,5	70 (2)	87,1	83,33 (2)	93,3	81	81
Dívka 12	64,29 (2-3)	46	73,33 (2)	53,3	27 (4)	47,5	60 (2)	×	50 (3)	71,7	54,92	55
Chlapec 11	85,71 (1-2)	90	80 (2)	76,7	55 (2-3)	77,5	50 (3)	98,6	50 (3)	73,3	64,14	83
Chlapec 12	50 (3)	54	66,67 (2-3)	76,7	73 (2)	35	70 (2)	64,3	16,66 (4)	78,3	55	62
Dívka 13	×	62	86,67 (1-2)	60	100 (1)	65	100 (1)	50	100 (1)	86,7	96,67	65
Chlapec 13	×	44	26,67 (5)	3,3	73 (2)	70	0 (5)	×	58,33 (2-3)	0	39,50	29
Chlapec 14	78,57 (2)	76	73,33 (2)	76,7	×	80	×	60	100 (1)	100	83,97	79

Zdroj: autoři

7 Interpretace dat a diskuze

7.1 Počet testových položek

První předpoklad P_1 : Teacher made testy neodpovídají současnému paradigmatu tvorby didaktických testů ve vztahu k doporučenému počtu testových položek referuje k reliabilitě užitých nástrojů. Námí vytvořených pět didaktických testů vždy obsahovalo deset testových položek, což koresponduje s doporučením aktuální literatury (Škoda et al., 2006). Na druhou stranu pět teacher made testů bylo tvořeno 8, 6, 7, 4 a 6 testovými položkami s průměrnou reliabilitou. Podle odborné literatury se nejedná o dostatečnou hranici vzhledem k tomu, že relativní velikost chyby měření se zvyšuje se snižujícím se počtem úloh. Ve výsledku to znamená, že zatímco v případě didaktických testů ovlivňují nahodilé jevy (únava, vnější biologické podmínky atp.) výsledné hodnocení z 10 %, v případě teacher made testů se jedná průměrně o 17 %.

7.2 Poměr objektivně a subjektivně skórovatelných položek

Analýzu druhého předpokladu P_2 : Teacher made testy neodpovídají současnému paradigmatu tvorby didaktických testů ve vztahu k doporučenému poměru objektivně a subjektivně skórovatelných testových položek prezentujeme prostřednictvím tabulky.

Tab. 5: Procentuální výsledek počtu objektivně skórovatelných úloh v jednotlivých testech

Testy	Test č. 1	Test č. 2	Test č. 3	Test č. 4	Test č. 5	Průměr
Teacher made test	87,5 %	66,67 %	100 %	100 %	0 %	70,8 %
Didaktický test	100 %	90 %	90 %	90 %	90 %	92 %

Zdroj: autoři

Zjištěné průměrné hodnoty je možné interpretovat tak, že v případě teacher made testů vstupuje subjektivita do výsledného hodnocení z 29,2 %, zatímco v případě didaktických testů se jedná pouze o 8 %. V žádném případě nechceme polemizovat nad tím, zda subjektivita do vyučovacího procesu patří (z našeho úhlu pohledu jednoznačně patří). Jsme však přesvědčeni, že subjektivita má své místo v případě hodnocení úsilí, schopnosti kooperace, verbálního projevu apod. V případě testování (testujeme-li znalosti) by měl být její vliv minimalizován. Vyhodnocování subjektivně skórovatelných úloh může podléhat percepčním chybám učitele (kvalita písma žáka, vnímání žáka učitelem, gender aj.). Percepční chyby jsou mentální zkratky (heuristiky – zrychlený úsudek o druhém člověku) z důvodu nedostatku času či zájmu. V praxi se jednotlivé chyby doplňují a navazují na sebe, čímž je umocněn jejich efekt. Mohou si však také odporovat. Holeček (2014) zdůrazňuje významnost nezájatosti učitele při posuzování žáků: „Vnímání učitele by mělo být objektivní a nezkreslené. . .“ (s. 36), přestože je zřejmé, že každý člověk, tedy i učitel, je neustále ovlivňován mnoha vjemy, které následně zpracovává a utváří si z nich názor a postoj k ostatním. Pokud si však vyučující dostatečně neuvědomuje své vlastní percepční omezení, a nesnaží se je napravit, jeho chování může přinášet dalekosáhlé nežádoucí dopady na školní výkony žáka, na jeho školní úspěšnost, budoucí studium i celý jeho život. „Každý člověk, tedy i každý učitel, má předsudky. Podmínkou pro odstranění předsudků je jejich rozpoznání. Snaha uvědomit si vlastní předsudky je akt zodpovědnosti.“ (Lazarová & Pol, 2002, s. 5) Rozborem percepčních chyb bychom se již dostali mimo ústřední linii tohoto příspěvku, nicméně považujeme za nutné výběrově uvést příklady, ve kterých se subjektivita posuzovatele promítla do hodnocení (blíže Man et al., 2000). Například v kontextu neměnnosti očekávání učitele vůči žákům (tzv. perseverační tendence) bylo již v roce 1928 prokázáno, že u zdatnějších žáků měli učitelé tendenci v náhodně vybraných diktátech přehlížet chyby častěji než u méně zdatných žáků. Další příklad uvádí, že když byly dvěma skupinám učitelů předloženy dvě stejné písemné práce s informací, že práce byla napsána jazykově nadaným žákem s výborným rodinným zázemím nebo průměrným žákem z průměrného rodinného zázemí, učitelé hodnotili stejnou práci fiktivně lepšího žáka s dobrým zázemím značně lépe než průměrného žáka. Za určitých okolností proto může nepřiměřené očekávání nabýt podoby sebenaplňující předpovědi (pygmalion efekt, golem efekt). Jiné studie prokazují vliv vzhledu žáka na výkon (Fitzpatrick et al., 2016), poruchy učení, pozornosti, chování (učitelé předpovídali vyšší úspěch studentům, kteří četli pod úrovní požadované čtenářské gramotnosti a zároveň nebyli označeni jako žáci s poruchou učení, pozornosti či chování (Tournaki, 2003), nebo genderu (Hofer, 2015). Subjektivita by měla být co nejvíce eliminována právě konstrukcí objektivně skórovatelných položek.

7.3 Poměr testových položek na nižší a vyšší kognitivní operace

Třetí předpoklad P_3 : Teacher made testy neodpovídají současnému paradigmatu tvorby didaktických testů ve vztahu k doporučenému poměru testových položek na vyšší a nižší kognitivní operace souvisí s tzv. konstruktovou validitou, tj. zda (a jak dobře) určený nástroj zjišťuje specifické pedagogické nebo psychologické konstrukty (v našem případě úrovně kognitivních operací dle Bloomovy taxonomie). Pro určení typu položek jsme využili expertní validizace (též mínění soudců – Disman, 2005). Každý z pěti expertů pro každou z testových položek v případě didaktických i teacher made testů dichotomicky určil, zda se jedná o položku vyžadující vyšší nebo nižší kognitivní operace. Jednotlivé testy byly tvořeny následujícími položkami (poměr nižší : vyšší kognitivní náročnost):

- Didaktické testy (*Me* expertního posouzení): 2 : 8, 5 : 5, 3 : 7, 5 : 5, 3 : 7. Celkem: 18 : 32 (36 % : 64 %).
- Teacher made testy (*Me* expertního posouzení): 0 : 8, 1 : 5, 1 : 6, 0 : 4, 1 : 5. Celkem: 3 : 24 (11 % : 89 %).

Na základě expertního posouzení můžeme konstatovat, že v teacher made testech nebyl reflektován poměr otázek na vyšší a nižší kognitivní operace, jenž je doporučován odbornou literaturou. Na druhou stranu nebyl tento poměr splněn ani v kontextu didaktických testů. Jak však již bylo jednou uvedeno, konstrukce didaktických testů byla limitována nízkým počtem realizovaných vyučovacích hodin (2–4), ze kterých pedagožka vytvářela teacher made testy, a při jejich tvorbě jsme vycházeli výhradně ze zaznamenané práce žáky v hodině (nebylo nám umožněno přímé pozorování). Z tohoto důvodu jsme čelili dilematu vytvoření teoreticky žádoucí úlohy vyšší kognitivní náročnosti, avšak bez písemného dokladu o tom, jestli je vůbec možné od žáků očekávat její vypracování.

7.4 Doporučovaný způsob skórování

Závěrečný předpoklad (P_4 : Teacher made testy neodpovídají současnému paradigmatu tvorby didaktických testů ve vztahu k doporučenému způsobu skórování – úlohy jsou hodnoceny jiným počtem bodů) odkazuje k bodování jednotlivých testových položek. Šatánek a Hubalovská téměř před půl stoletím (1972) tvrdili, že by každá úloha měla být ohodnocena podle náročnosti. Tím pádem by testové položky měly být hodnoceny jiným počtem bodů jako ve většině námi analyzovaných teacher made testech. Jejich výzkum se věnuje třem žákům, kdy každý z žáků zvládl stejný počet úloh, ale každá úloha byla jiné náročnosti, proto nevidí důvod, proč by žáci měli být ohodnoceni stejně. Jak však bylo výše uvedeno, v případě, kdy test neobsahuje široké úlohy vyžadující delší čas na jejich vyplnění, doporučuje se binární skórování (Chráška, 1999; Škoda et al., 2006). V teacher made testech nebyl tento přístup respektován. Tento fakt může vést k paradoxní situaci:

- Chlapec 4, který v teacher made testu č. 1 vyřešil 3 úlohy z 8 (vyřešil 37,5 % testu), získal 9 bodů (jelikož první úloha byla hodnocena maximem 7 bodů) ze 14 možných a (64,3 %) byl hodnocen známkou 2–3. Chlapec 13 ve stejném testu vyřešil 4 úlohy za 1 bod a z první úlohy za 7 bodů získal 3 body (tuto úlohu řešil na 42,9 %). Z hlediska testové domény tento žák vyřešil téměř 4,5 úlohy (56,3 % testu), získal však 7 bodů ze 14 (50 % bodového maxima) a byl ve výsledku hodnocen hůře (známka 3) než chlapec 4.
- Nejednotnost v bodovém ohodnocení jednotlivých testových položek a jejich různý počet v jednotlivých teacher made testech způsobuje druhou paradoxní situaci – získání 60 % bodů z celkového množství znamenalo v případě teacher made testu č. 1 a č. 3 hodnocení 2–3, testu č. 2 hodnocení 3, testu č. 4 a 5 hodnocení 2.

7.5 Souvislost výsledků žáků v rámci užití dvou nástrojů měření

V závěru předkládané případové studie jsme testovali nulovou hypotézu tvrdící, že: „Neexistuje spojitost mezi výsledky žáků v rámci jejich procentuálního skórování v teacher made testech a didaktických testech.“ Na jednocentní hladině významnosti jsme odmítli nulovou hypotézu a přijali hypotézu alternativní: spojitost mezi výsledky žáků v rámci jejich procentuálního skórování v teacher made testech a didaktických testech je pozitivní. Potvrzení této hypotézy se samozřejmě dalo předpokládat, naše pozornost však směřovala k hodnotě Spearmanova korelačního koeficientu ($r = 0,662$) vyjadřující pouze středně silnou spojitost (Hendl, 2012; Chráška, 2016).

Vzhledem k tomu, že žáci pětkrát po sobě vyplňovali vždy dva testy na stejný obsah vzdělávání, očekávali bychom spojitost blízkou až k absolutní korelaci. Tento fakt je rovněž patrný z celkového průměrného procentuálního skóre, které prezentujeme v tabulce 6.

Tab. 6: Syntéza – celkový průměrný procentuální skór žáků

Typ testu	Teacher made test	Didaktický test
Celkový průměrný výsledek žáků	78,91	72

Zdroj: autoři

Podle našeho mínění tato skutečnost vyvolává otázky ohledně činitelů, které tuto situaci zapříčinily (spolehlivost testů? orientace teacher made testů primárně na položky na nižší kognitivní operace a s tím související nižší schopnost žáků řešit položky na vyšší kognitivní operace v didaktických testech), a zároveň otevírá pole pro budoucí výzkumy v této věci.

Z didaktického testu měli žáci většinou horší výsledky než z testu pedagožky (viz tab. 4). Sedláčková (1993) uvádí možnost klasifikace na základě procenta správných odpovědí. Podle autorky v případě běžné klasifikace představuje každá ztráta 10 % jeden klasifikační stupeň, avšak ve velmi přísně nastaveném měřítku je to již 5 %. Uvažujeme-li v intencích výsledků této studie a při nastavení běžného klasifikačního měřítko Sedláčkové (1993), 12 z 27 žáků by obdrželo jiné hodnocení v závislosti na tom, zda by byl jako evaluační nástroj použit teacher made test, nebo didaktický test. V osmi případech se průměrné výsledky z obou typů testů lišily o 5–9 % a pouze v sedmi případech se výsledky lišily méně než o 4 % (což by podle autorky nemělo vliv na výsledné hodnocení). To je podle našeho názoru nejpodstatnější zjištění této studie. I zde se otevírá možnost dalšího šetření, jelikož potenciálně přínosné by mohlo být interview se sedmi žáky, kteří v této studii (poněkud paradoxně) získali vyšší průměrné skóre z didaktických testů než z teacher made testů pedagožky.

8 Limity studie

Přes prokázané diference (v intencích míry reliability a validity této případové studie) mezi teacher made testy a didaktickými testy musíme mít na paměti intervenující proměnné (interní validita šetření) bránící jednoznačné interpretaci výsledků (tak, jak je to v neexaktních vědách pravidlem). Zde uvádíme činitele, které mohly zkreslit výsledky: efekt měření odkazuje k situaci, kdy si žáci z prvního testu osvojí určité informace nebo dovednosti, což následně zkreslí výsledky druhého testu. Prodleva mezi testem zadaným pedagožkou a námi zadaným testem činila 1–4 dny. Žáci mohli po testu zadaném pedagožkou diskutovat jeho obsah a správné odpovědi na otázky nebo si své odpovědi ověřit ze zápisů v sešitu (na druhou stranu by pak zřejmě neobdrželi průměrně horší hodnocení, jak se také stalo). Ne všechny testy byly vyplňovány stejným počtem žáků (všech 10 testů vyplnilo 15 žáků, 9 žáků vyplnilo 9 testů, 2 žáci vyplnili 8 testů a 1 žák vyplnil 6 testů; ve 13 případech se jednalo o absenci při vyplňování teacher made testů, ve 4 případech při vyplňování didaktických testů). Za značný limit této studie shledáváme konstruktovou validitu didaktických testů (nebyl splněn doporučený poměr položek na vyšší a nižší kognitivní operace). Příčiny této situace byly diskutovány výše. Dalším limitem této studie je malý výzkumný vzorek (1 pedagog, 5 teacher made testů a didaktických testů, 27 dětí jedné 4. třídy ZŠ) a jeho dostupný výběr znemožňující generalizaci výsledků. Rovněž způsob konstrukce teacher made testů (nejednotné skórování úloh) znemožňuje ověření normality a případné použití rigoróznějších parametrických testů. Zároveň si uvědomujeme, že došlo k porovnávání známek napříč dvěma předměty (vlastivěda, přírodověda), což může vyvolávat otázku ve vztahu k parametrům testů pedagožky v jiných vyučovacích předmětech.

9 Závěr

Trčková ve své práci (2013) uvádí, že 97 % učitelů používá didaktické testy. Můžeme si pokládat otázku, co si přesně učitel představuje pod pojmem *didaktický test*. Autorka dále zmiňuje, že pouhých 7 % učitelů si dle jejího výzkumu tvoří didaktický test sami, ostatní čerpají z učebnic nebo jiné literatury. Nabízí se otázka o kvalitě „testů“ (spíše úloh nebo otázek?) uvedených v učebnicích (ve smyslu jejich reliability, validity a objektivitu). Zároveň je důležité podotknout, že je stále nedostatečné množství zdrojů, které by poukazyvaly na jednoznačná a sjednocující pravidla pro psaní jednotlivých položek v textu, stále se tak jedná o určitý proces, ve kterém se pedagogové řídí spíše svým osobním pocitem (Millman & Greene, 1993; Haladyna et al., 2002). Přestože využívání teacher made testů lze považovat za jeden z efektivních nástrojů pro zajišťování zpětné vazby od žáků, je stále nutné myslet na skutečnost, že ne vždy disponují dostatečnou mírou reliability a validity a zároveň že ne každý pedagog má dostatečné dovednosti k jejich správné evaluaci (srov. Mpofo, 2011; Hartell & Strimel, 2019). Učitelé dávají nejčastěji do popředí jednoduché úlohy na pamětní operace (Germ & Harms, 2008). Janík a Stuchlíková (2010) poukazují na fakt, že takto koncipované úlohy snižují zájem žáků například o přírodovědné obory. Suchoradský (2008)

považuje testy jako jednoduchou a zejména pak spravedlivou formu zkoušky. Oceňuje objektivní a rychlý způsob, jak prověřit schopnosti žáka, ale zdůrazňuje, že by se nemělo jednat o jediný způsob hodnocení. K tomuto stanovisku se připojujeme i my. Je potřeba hledat rozmanité formy a způsoby hodnocení, didaktické testy představují pouze jednu z využitelných možností.

Nebylo cílem zde didaktický test představovat jako „ideální či optimální“ způsob evaluace žákovské činnosti. Jsme si vědomi, že vše má svá pozitiva a negativa, což samozřejmě platí i pro didaktický test. Za silné stránky se považují stejné podmínky pro všechny žáky při testu, reliabilita, validita, objektivita nebo způsob vyhodnocení (praktičnost). Za slabé stránky se naopak shledává časová náročnost při tvorbě, absence mluveného projevu žáka (nemožnost si tak ověřit určité zvládnuté kompetence mluveného projevu) nebo ztráta možnosti interakce s učitelem, která může navést žáka správným směrem při zodpovězení otázek (Hališka, 1999). Záměrem této studie v žádném případě nebylo ani poukazovat na potenciální nedostatky v oblasti kompetencí pedagogů při tvorbě didaktických testů, avšak z našeho úhlu pohledu si tato oblast zaslouží zvláštní pozornost. Předkládaná deskriptivní případová studie měla za cíl poukázat na kvalitu vyhotovených testů jedné pedagožky ve dvou vyučovacích předmětech. Design celého výzkumného směřování (dostupný výběr, jedna třída, nízký počet žáků, pět teacher made testů) neumožňuje generalizaci výsledků (externí validita) a existuje řada intervenujících proměnných potenciálně zkreslujících jednoznačnou interpretaci výsledků (interní validita), včetně uvažování nad vyhotovenými pěti didaktickými testy badatelů jako potenciálního vztahového kritéria kvality vůči teacher made testům pedagožky. Záměrem této práce je vzbudit pozornost odborné (akademiků, praktikujících učitelů) i laické (municipalita, rodiče) veřejnosti k této problematice, jelikož se podle našeho mínění jedná o zcela zásadní téma, a zároveň nabídnout badatelům na tomto poli potenciální způsob při realizaci podobné studie). Závěrem si dovolueme položit dvě otázky: Vyplývají známky na vysvědčeních žáků ze spolehlivých, platných a objektivních nástrojů měření a je vůbec možné je poměřovat? Je možné uvažovat o budoucí hypotetické možnosti vyvážení autonomie učitelů při tvorbě testových materiálů s organizovaným (centrálním) modelem testování vycházejícího z více rigorózního přístupu při tvorbě testových materiálů?

Poděkování

Tento příspěvek vznikl za podpory grantové agentury UJEP-SGS-2020-43-008-2.

Literatura


- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Rath, J., & Wittrock, M. C. (2000). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives, abridged edition*. Pearson PLC.
- Byčkovský, P. (1982). *Základy měření výsledku výuky*. Tvorba didaktického testu, ČVUT VÚIS.
- Cizek, G. J. (2004). Achievement tests. *Encyclopedia of Applied Psychology, 1*, 41–42.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302. <https://doi.org/10.1037/h0040957>
- Čapek, R. (2015). *Moderní didaktika: lexikon výukových a hodnotících metod*. Grada.
- Disman, M. (2005). *Jak se vyrábí sociologická znalost*. Karolinum.
- Dutke, S., Barenberg, J., & Leopold, C. (2010). Learning from text: Knowing the test format enhanced metacognitive monitoring. *Metacognition and Learning, 5*(2), 195–206. <https://doi.org/10.1007/s11409-010-9057-1>
- Fayol, M., Alamargot, D., & Berninger, V. W. (2012). *Translation of thought to written text while composing: advancing theory, knowledge, research methods, tools, and applications*. Psychology Press/Taylor & Francis Group.
- Fitzpatrick, C., Côté-Lussier, C., & Blair, C. (2016). Dressed and groomed for success in elementary school: Student appearance and academic adjustment. *Elementary school journal, 117*(1), 30–45. <https://doi.org/10.1086/687753>
- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative inquiry, 12*(2), 219–245. <https://doi.org/10.1177/1077800405284363>
- Germ, M., & Harms, U. (2008). What do biology tests look like in German grammar schools? A descriptive study about task formats and teachers' intentions for surveying different cognitive dimensions. In M. Hamman, M. Reiss, C. Boulter, & S. D. Tunnicliffe (Eds.), *Biology in Context: Learning and Teaching for the twenty-first century* (pp. 248–258). Institute of Education.

- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied measurement in education*, 15(3), 309–333. <https://doi.org/10.1207/S15324818AME1503-5>
- Hališka, J. (1999). *Jak testy sestavit a pracovat s nimi*. Středisko služeb školám.
- Hartell, E., & Strimel, G. J. (2019). What is it called and how does it work: examining content validity and item design of teacher-made tests. *International Journal of Technology and Design Education*, 29(4), 781–802. <https://doi.org/10.1007/s10798-018-9463-2>
- Hawkes, H. E., Lindquist, E. F., & Mann, C. R. (1936). *Construction and use of achievement examinations*. Houghton Millin Company.
- Hendl, J. (2012). *Přehled statistických metod*. Portál.
- Hofer, S. I. (2015). Studying gender bias in physics grading: The role of teaching experience and country. *International Journal of Science Education*, 37, 2879–2905. <https://doi.org/10.1080/09500693.2015.1114190>
- Holeček, V. (2014). *Psychologie v učitelství*. Grada.
- Chlup, O. (1931). *O školu měšťanskou*. Nové školy. Knihovny nových škol.
- Chráška, M. (1999). *Didaktické testy: příručka pro učitele a studenty učitelství*. Paido.
- Chráška, M. (2002). *Didaktické testy ve školní praxi*. Brno.
- Chráška, M. (2007). *Metody pedagogického výzkumu*. Grada.
- Chráška, M. (2016). *Metody pedagogického výzkumu*. Grada.
- Janík, T., & Stuchlíková, I. (2010). Oborové didaktiky na vzestupu: přehled aktuálních vývojových tendencí. *Scientia in educatione*, 1, 5–32.
- Jeřábek, O., & Bílek, M. (2010). *Teorie a praxe tvorby didaktických testů* [online] [cit. 10. 5. 2020]. UPOL. Dostupné z <http://zvyp.upol.cz/publikace/bilek-jerabek.pdf>.
- Johanson, R. (2003). Case study methodology. *Acta Linguistica Hungarica – ACTA LINGUIST HUNG*, 32, 22–24.
- Johnson, R., & Penny, J. (2005). Split-Half Reliability. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (pp. 649–654). Elsevier.
- Junková, J. (2006). *Didaktické testování* [online] [cit. 17. 5. 2020]. Dostupné z https://is.muni.cz/el/1441/podzim2009/ZS1BK_PDD/didakticke_testovani.pdf.
- Kalhous, Z., & Obst, O. (2002). *Školní didaktika*. Portál.
- Komenda, S., & Mazuchová, J. (1995). *Tvorba a testování testu*. UP.
- Laufková, V., & Starý, K. (2016). *Formativní hodnocení ve výuce*. Portál.
- Lazarová, B., & Pol, M. (2002). *Multikulturalita a rovné příležitosti v české škole*. Institut pedagogicko-psychologického poradenství ČR.
- Linn, R. L. (2008). *Measurement and assessment in teaching*. Pearson Education India.
- Man, F., Mareš, J., & Stuchlíková, I. (2000). Paradoxní účinky učitelových motivačních postupů. *Pedagogika*, 50(3), 224–235.
- Mareš, J. (2015). Tvorba případových studií pro výzkumné účely. *Pedagogika*, 65(2), 113–142.
- Marinova, V., Tsvetkov, D., & Hristov, L. (2005). On the reliability of didactic tests. *Pedagogical Almanac*, 13(1), 242–248.
- Millman, J., & Greene, J. (1993). The specification and development of tests of achievement and ability. In R. L. Lin (Ed.), *The American Council on Education/Macmillan series on higher education. Educational measurement* (pp. 335–366). Macmillan Publishing Co, Inc; American Council on Education.
- Mpofu, B. (2011). *Formative evaluation versus summative evaluation*. Longman.
- Popham, W. J. (2017). *The ABCs of educational testing: demystifying the tools that shape our schools*. Corwin.
- Pulpán, Z. (1991). *Základy sestavování a klasického vyhodnocování didaktických testů*. Kotva.
- Rámcový vzdělávací program pro základní vzdělávání. (2017). MŠMT. [on-line] [cit 5. 3. 2020]. Dostupné z: <https://www.msmt.cz/file/41216/>
- Sedláčková, J. (1993). *Diagnostické metody ve vyučování matematice*. PřF UPOL.
- Sekaran, U. (1992). *Research methods for business: A skill building approach*. 2nd ed. Wiley.

- Schindler, R. (2006). *Rukověť autora testových úloh*. Centrum pro zjišťování výsledků vzdělávání.
- Skutil, M. (2011). *Základy pedagogicko-psychologického výzkumu pro studenty učitelství*. Portál.
- Smékal, V., Švec, V., & Zajac, J. (1973). *Didaktické testy a jejich vyhodnocování*. Středisko pro výzkum učebních metod a prostředků.
- Suchoradský, O. (2008). *Testy a jejich užití při hodnocení žáků*. Metodický portál [on-line] [cit. 11. 3. 2020]. Dostupné z <https://clanky.rvp.cz/clanek/c/Z/2666/TESTY-A-JEJICH-UZITI-PRI-HODNOCENI-ZAKU.html>.
- Šatánek, A., & Hubálovská, H. (1972). Příspěvek k hodnocení výkonů v testových zkouškách. *Pedagogika*, 2, 185–189.
- Škoda, J., & Doulík, P. (2007). *Tvorba a hodnocení didaktických testů: cvičebnice pro studenty učitelství a účastníky kurzu DPS*. PF UJEP.
- Škoda, J., Doulík, P., & Hajerová-Müllerová, L. (2006). *Zásady správné tvorby, použití a hodnocení didaktických testů v přípravě budoucích učitelů* [on-line] [cit. 2. 2. 2020]. Dostupné z <http://cvicebnice.ujep.cz/cvicebnice/FRVS1973F5d/>
- Švamberg Šauerová, M. (2016). *Hyperaktivita nebo hypoaktivita – výchovný problém*. Wolters Kluwer.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Tournaki, N. (2003). Effect of student characteristics on teachers' predictions of student success. *Journal of Educational Research*, 96(5), 310–319. <https://doi.org/10.1080/00220670309597643>
- Trčková, V. (2013). *Didaktické testování jako profesní kompetence učitele matematiky na základní škole* (Diplomová práce). [online] [cit. 9. 3. 2020]. Dostupné z <https://theses.cz/id/03mwq9/>.
- Urbánek, T. (2002). *Základy psychometriky*. Masarykova univerzita.
- Vrána, S. (1948). *Zkoušení a známkování*. Komenium, učitelské nakladatelství.
- Ward, A., Stoker, H. W., & Murray-Ward, M. (1996). *Educational measurement: Theories and applications* (Vol. 2). University Press of America.

Vlastnosti úloh z obecné chemie vyplývající z analýzy přijímacího řízení na Přírodovědeckou fakultu Univerzity Karlovy

The Qualities of Tasks in General Chemistry Based on the Analysis of Entrance Procedure at the Faculty of Science, Charles University

Martin Šrámek^{1,*},  Milada Teplá¹

¹ Přírodovědecká fakulta, Univerzita Karlova, Albertov 6, Praha 2, Česká republika; sramekm123@gmail.com

Článek představuje výsledky položkové analýzy oborových testů z chemie zadávaných pro účely přijímacího řízení na Přírodovědecké fakultě Univerzity Karlovy v letech 2016 až 2019. Cílem položkové analýzy, která byla zaměřena na úlohy z obecné chemie, bylo odhalit úlohy, které byly pro uchazeče o studium nejvíce obtížné a úlohy necitlivé tedy takové úlohy, které nedostatečně rozlišovaly úspěšné uchazeče od uchazečů neúspěšných. Obtížné úlohy byly identifikovány na základě výpočtu indexu obtížnosti. Necitlivé úlohy byly stanoveny na základě výpočtu koeficientů ULI(1/2) a RIR. Výsledky položkové analýzy odhalily, že pro uchazeče obtížným tématem je stavová rovnice ideálního plynu. Dále byla odhalena čtyři témata, která obsahovala alespoň jednu obtížnou úlohu (výpočet pH, elektronová konfigurace, iontový součin vody a elektrolyza). Necitlivé úlohy byly detekovány u třech témat (tepelný rozklad látky, rychlostní rovnice a chemické rovnováhy). Necitlivost úlohy často souvisela s vysokým počtem kroků nutných k jejímu úspěšnému vyřešení.

Klíčová slova:
přijímací test, obecná chemie, položková analýza, obtížnost položky, citlivost položky.

Zasláno 10/2021
Revidováno 10/2021
Přijato 12/2021

This article introduces results of item analysis of chemistry tests used as a part of the admission procedure at the Faculty of Science, Charles University, between 2016 and 2019. The aim of the analysis focusing on general chemistry was to determine which tasks were the most troublesome for applicants and those with a low discrimination index. The troublesome tasks were found based on the calculation of difficult indexes whereas the item discrimination index was represented by ULI(1/2) and RIR coefficients. The results of the item analysis showed that ideal gas law is the hardest and only troublesome topic. Moreover, there were four topics including at least one troublesome item (pH calculation, electron configuration, ion product calculation, and chemical equilibrium). Additionally, the analysis revealed three topics containing an item with a low discrimination index (thermal decomposition, rate law, chemical equilibrium) and, in many cases, its value decreased with a number of steps necessary for reaching the solution in the task.

Key words:
admission tests, general chemistry, item analysis, item difficulty, item discrimination.

Received 10/2021
Revised 10/2021
Accepted 12/2021

1 Úvod

Přijímací zkoušky na vysokou školu jsou milníkem pro studenty, kteří se na školu hlásí. Jsou však také důležitým podkladem pro vysoké školy, které studenty vzdělávají. Přijímací zkoušky totiž mohou napomoci odhalit úseky učiva, které pro žáky středních škol, budoucí studenty škol vysokých, mohou být obtížné. Vyučující se na tyto úseky studia mohou ve své výuce specificky zaměřit a tím pomoci studentům překonat překážky, které je ve studiu čekají. Pro přírodovědné obory, mezi nimi i chemii, může být takový přístup zásadní, protože se potýkají s vysokým procentem neúspěšných studentů, kteří studium nedokončí a přestupují na jiné obory, nebo studia ukončí (viz např. Pikálková et al., 2014).

Prezentovaný výzkum byl zaměřen na vzdělávací obor Chemie. Ten lze dále studovat na řadě vysokých škol, na kterých mohou úspěšní uchazeči o studium studovat odborné studijní programy (např. Biochemie, Chemie životního prostředí ad.) či programy zaměřené na přípravu budoucích učitelů předmětu chemie. Obě zmíněné skupiny studijních programů jsou nabízeny např. na Přírodovědecké fakultě Univerzity Karlovy (dále jen PŘF UK), kde proběhl tento výzkum.

2 Teoretická východiska

2.1 Oborový test jako jedna z forem přijímacího řízení

Přijímací řízení by mělo mít co nejvyšší možnou predikční validitu, tedy schopnost předpovědět akademický úspěch daného studenta. Akademická úspěšnost udává, zda-li student dokončí zdárně první ročník studia, eventuálně dané studium jako celek (Brown, 2003). Výzkumem prediktorů akademické úspěšnosti se zabývá řada českých i zahraničních studií (např. Berger, 2012; Rubešová, 2009; Škaloudová, 2003).

Uvedení autoři dochází shodně k závěru, že mezi validní prediktory akademického úspěchu, tj. úspěšného dokončení 1. ročníku, eventuálně studia, patří středoškolský prospěch. Rubešová (2009, str. 97) ve svém článku uvádí, že „kolísání konečného průměru studentů můžeme vysvětlit ze 47 % závislosti na předchozí úspěšnosti studia na střední škole.“ Při současném zohlednění úspěšnosti v přijímacím řízení (oborový test z matematiky + oborový test z geografie) však dojde k nárůstu koeficientu determinace modelu na 62 %. Využití oborových testů v přijímacím řízení tedy přispívá k výběru vhodných uchazečů (Rubešová, 2009) a oborové testy patří taktéž mezi validní prediktory akademického úspěchu. Oba prediktory (středoškolský prospěch i výsledek v přijímacím testu na vysokou školu) se navzájem doplňují a společně lépe predikují úspěšnost studia sledovaného jedince (Šrámek et. al., 2020).

Přijímací řízení formou pouze jednoho oborového testu je od akademického roku 2017/2018 taktéž využíváno pro výběr vhodných uchazečů o studijní program Chemie na PřF UK. Oborový test si jednotliví uchazeči mohou zvolit v závislosti na profilu jednotlivých studijních oborů (matematika, chemie, nebo biologie). Informace o středoškolském prospěchu uchazeče se v aktuálních podmínkách přijímacího řízení na PřF UK již nevyžaduje (PřF UK, 2021).

Kromě predikce akademické úspěšnosti může využití oborových testů jako jedné z formy realizace přijímacího řízení sloužit ke zjištění úrovně výsledků vzdělávání v rámci předchozích stupňů studia.

2.2 Analýza oborového testu z chemie se zaměřením na obecnou chemii – výsledky dosavadních studií

Za účelem zvýšení kvality přijímacího řízení byly provedeny analýzy oborových testů na vysoké škole. V oblasti chemie provedly výzkum např. Belháčová (2002), Jedličková (2007), Martincová (2001) a Štefanová (2005). K dosažení uvedeného cíle využily všechny výše uvedené autorky položkovou analýzu, na jejímž základě interpretovaly zjištěné poznatky. V rámci obecné chemie se došlo k následujícím závěrům:

Mezi úlohy s **nízkým indexem obtížnosti** (pro uchazeče obtížné úlohy) patří úlohy zaměřené na: (i) chemické rovnováhy (Martincová, 2001; Štefanová, 2005); (ii) termochemii (Martincová, 2005); (iii) teorii kyselin a zásad, včetně výpočtů pH (Belháčová, 2002; Jedličková, 2007; Štefanová, 2005) a (iv) poznatky plynoucí z trendů v periodické soustavě prvků (Jedličková, 2007; Štefanová, 2005).

Na druhou stranu mezi úlohy s **vysokým indexem obtížnosti** (pro uchazeče nejméně obtížné úlohy) patří (i) obecně úlohy z oblasti složení atomu (Martincová, 2001; (ii) výpočet reakčního tepla látky (Štefanová, 2005); (iii) tepelný rozklad látky (Štefanová, 2005) a (iv) vyčíslení chemické rovnice (Belháčová, 2002).

Martincová (2001) však poukazuje na závislost indexu obtížnosti s konkrétním zadáním úlohy a vytýčila témata, do kterých spadaly úlohy s **různorodou náročností** závisující právě na konkrétním zadání. Jednalo se o úlohy zaměřené na (i) práci s oxidačními čísly; (ii) teorii kyselin a zásad, včetně výpočtů pH a (iii) poznatky plynoucí z trendů v periodické soustavě prvků (Martincová, 2001). Na souvislost indexu obtížnosti s konkrétním zadáním a taktéž počtem kroků nutných k vyřešení úlohy poukazuje též Cowan (2001), Miller (1985) či Čipera et al. (1956).

Problematika úloh a obsahu přijímacího testu z chemie byla řešena i v zahraničí – např. v rámci studie Herridge (2016). Autorka studie identifikovala ve své práci ty oblasti obecné chemie, které vnímají studenti (shodně s míněním vyučujících) v základním kurzu chemie na vysoké škole jako obtížné. V souladu se zjištěními ve shora uvedených studiích uvádí autorka, že jako nejvíce obtížné vnímají studenti i učitelé odvozování vlastností prvků a sloučenin na základě obecných znalostí periodické soustavy prvků (konkrétně je uvedena elektronová konfigurace, elektronová afinita, velikost atomů, reaktivita a ionizační energie). Další výzkum autorky v průběhu semestru odhalil s tím související problémy, jakými je vnímaná náročnost při určování produktů reakcí. Na konci semestru uváděli studenti jako komplikovaná témata chemickou rovnováhu, práci s oxidačními čísly a některé typy výpočtů například molární koncentraci. Výpočet pH vnímali studenti jako průměrně náročný. S podobnými výsledky přichází studie Tilahun a Tirfu (2016), která v souladu se shora uvedenými autorkami uvádí mj. jako obtížné téma chemickou rovnováhu a téma termochemie.

3 Cíle a výzkumné otázky

Hlavním cílem prezentované studie bylo zmapovat nejvíce obtížné úseky z učiva středoškolské chemie se zaměřením na obecnou chemii, které vyplývají z analýzy oborového testu chemie zadávaného na PřF UK v letech 2016 až 2019 a navázat tak na výsledky dříve realizovaných studií. Druhým neméně důležitým cílem bylo zjistit, které úlohy z oblasti obecné chemie nejméně přispívaly k výběru úspěšných uchazečů na PřF UK (tedy byly nejméně citlivé). Výsledky analýzy budou reflektovány při tvorbě oborových testů z chemie použitých v rámci přijímacího řízení.

Vzhledem k výše vymezeným cílům studie byly stanoveny následující výzkumné otázky:

1. Které úlohy z obecné chemie zařazené do oborového testu byly obtížné pro uchazeče o studium na PřF UK?
2. Které úlohy z obecné chemie zařazené do oborového testu nedostatečně rozlišovaly úspěšné a neúspěšné uchazeče o studium na PřF UK?
3. Které úlohy z obecné chemie zařazené do oborového testu vykazovaly dostatečnou citlivost a zároveň adekvátní obtížnost?

4 Metodologie

Výzkumný vzorek byl tvořen 1 780 uchazeči o studijní obor Chemie. Výzkumnými nástroji bylo 8 oborových testů z chemie zadaných v řádných termínech přijímacího řízení v letech 2016 až 2019 (viz tab. 1). Každý test obsahoval 30 uzavřených úloh s výběrem jedné správné odpovědi ze čtyř nabízených alternativ. Přibližně 25 až 30 % z celkového počtu úloh bylo zaměřených na obecnou chemii, zbývající úlohy se zaměřovaly na anorganickou chemii, organickou chemii a biochemii. Aby bylo možné zodpovědět výše stanovené výzkumné otázky, byla u každého testu provedena položková analýza úloh z obecné chemie. Dále byly určeny následující parametry – úspěšnost, koeficient ULI(1/2), koeficient RIR (Chvál et al., 1999; Chráska, 2015; Varma, 2020).

Tab. 1: Počty uchazečů absolvujících danou variantu testu

Test	2016 –	2016 –	2017 –	2017 –	2018 –	2018 –	2019 –	2019 –	Celkem
	A	B	A	B	A	B	A	B	
Počet uchazečů	219	220	230	221	238	213	223	216	1 780
Počet úloh z obecné chemie	9	9	10	10	8	8	9	9	72

Vzhledem k tomu, že podkladem pro tvorbu oborových testů zadávaných na PřF UK slouží běžně dostupné středoškolské učebnice chemie a též Rámcové vzdělávací programy pro gymnázia (PřF UK, 2021), bylo pro stanovení obsahové validity zkoumáno zařazení konkrétních témat obecné chemie v Rámcovém vzdělávacím programu pro gymnázia (RVP G) (MŠMT, 2007) a též ve středoškolských učebnicích chemie. Z toho důvodu byla provedena obsahová komparace úloh s obsahem středoškolských učebnic chemie, které Huvarová (2010) ve svém výzkumu označila za nejčastěji používané. Konkrétně se jednalo o učebnice autorů: Mareček a Honza (2002), Mareček a Honza (2005a), Vacík (2005b) a Benešová (1999). Obsahová komparace byla provedena též z důvodu, že lze předpokládat, že žáci středních škol právě ze středoškolských učebnic čerpají své poznatky při přípravě na přijímací řízení.

Konstruktová validita úloh je u přijímacího testu na PřF UK stanovována expertním panelem, který se skládá ze tří tvůrců testů, a dále jedním až dvěma kontrolory. Experti mj. posuzují proveditelnost, čtivost, konzistenci stylu, či jasnost vyjadřování (srov. Taherdoost, 2016).

4.1 Stanovení obtížných úloh

Aby mohla být nalezena odpověď na první výzkumnou otázku, byla pomocí položkové analýzy úloh stanovena jejich obtížnost. Ta byla vymezena prostřednictvím úspěšnosti uchazečů v dané úloze, která je definována jako podíl uchazečů, kteří úspěšně vyřešili danou úlohu, a celkového počtu uchazečů (Chvál et al., 2015). Za **obtížnou úlohu** byla následně považována taková položka, v níž byla úspěšnost uchazečů méně než 30 %. Hranice úspěšnosti 30 % byla zvolena s ohledem na skutečnost, že všechny úlohy byly uzavřenými úlohami, kdy uchazeč volil správnou odpověď z právě čtyř alternativ. Tato forma úlohy má za následek, že pouhým hádáním lze dosáhnout statistické úspěšnosti rovné 25 %.

Aby bylo možné vymezit, která témata z obecné chemie jsou pro uchazeče nejvíce náročná, byla každá úloha, která byla shledána jako obtížná, následně zařazena do tematického celku z oblasti obecné chemie. V případě, že úspěšnost úloh spadajících do daného tematického celku nepřesahuje 30 %, je tento okruh úloh označen jako **obtížné téma**. V případě, že úspěšnost alespoň jedné úlohy v daném tematickém celku dosahuje nižší hodnoty než 30 % a zároveň daný tematický celek obsahuje úlohy s průměrnou úspěšností převyšující 30 %, je tento okruh úloh označen jako **potenciálně-obtížné téma**. Zbývající témata lze nahlížet jako témata s adekvátní či nižší obtížností (resp. adekvátní či vyšší úspěšností), jelikož úspěšnost ve všech úlohách převyšovala 30 % dle Šrámka et. al. (2021).

4.2 Stanovení úloh nedostatečně rozlišujících úspěšné uchazeče od uchazečů neúspěšných

Aby mohla být nalezena odpověď na druhou výzkumnou otázku, byla u každé úlohy vypočtena její citlivost, která indikuje, zda úloha dostatečně rozlišuje úspěšné a neúspěšné uchazeče. K tomu byly použity koeficienty ULI (1/2) a RIR. Koeficient ULI(1/2) byl vypočten následujícím způsobem. Uchazeči, kteří řešili danou úlohu byli rozděleni na dvě stejně velké skupiny dle celkového skóre dosaženého v daném testu. Pro každou skupinu byl vypočten index obtížnosti dané úlohy. Rozdíl těchto hodnot je následně roven právě koeficientu ULI(1/2) (Chvál et al., 2015). Koeficient RIR byl u každé úlohy určen tak, že byl vypočítán korelační koeficient mezi bodovým ziskem každého uchazeče v dané položce a jeho celkovým počtem v testu bez započtení bodů za danou položku (Chvál et al., 2015).

Pro účely prezentované studie byla úloha považována za **citlivou**, jestliže v případě výpočtu koeficientu ULI(1/2) došlo k naplnění jedné z následujících podmínek: (i) obtížnost úlohy je mezi 0,2 až 0,3 a zároveň hodnota vypočteného koeficientu ULI(1/2) je vyšší nebo rovna 0,15; (ii) obtížnost úlohy je mezi 0,3 až 0,7 a zároveň hodnota vypočteného koeficientu ULI(1/2) je vyšší nebo rovna 0,25 a (iii) obtížnost úlohy je mezi 0,7 až 0,8 a zároveň hodnota vypočteného koeficientu ULI(1/2) je vyšší nebo rovna 0,15 (Chráška, 1999). Při využití koeficientu RIR je za citlivou úlohu považována taková, jejíž hodnota tohoto koeficientu byla alespoň 0,15 s tím, že velmi dobré úlohy vykazují koeficient RIR vyšší než 0,25 (Varma, 2020). Citlivost byla ve spojitosti s jejich úspěšností odstupňovaná, neboť u úloh s úspěšností do 30 % (či naopak u úloh s úspěšností velmi vysokou) bývá v literatuře snížen požadavek na koeficienty citlivosti (ULI(1/2) a RIR) s ohledem na jejich zvýšenou náročnost (Chráška, 1999; Varma, 2020).

Úloha byla považována za **necitlivou**, jestliže hodnota obou koeficientů citlivosti (ULI(1/2) a zároveň RIR) byla nedostatečná dle výše uvedených kritérií. Každá úloha, která byla shledána jako necitlivá, byla následně zařazena do tematického celku z oblasti obecné chemie, aby mohly být predikovány důvody její necitlivosti. V tomto případě byly rovněž posuzovány všechny položky, které svým zaměřením spadaly do vymezeného tematického celku v případě, že daná úloha nebyla v daném tématu jediná.

4.3 Stanovení dostatečně citlivých úloh s adekvátní obtížností

Aby mohla být nalezena odpověď na třetí výzkumnou otázku, byla opět pomocí položkové analýzy úloh stanovena jejich obtížnost a citlivost. Za dostatečně citlivé úlohy s adekvátní obtížností byly označeny úlohy s úspěšností mezi 30 % a 80 % (tedy úlohy ne příliš obtížné, a ne příliš snadné) a zároveň se jedná o úlohy, které na základě definice uvedené v kapitole 3 lze označit jako citlivé.

5 Výsledky a diskuze

5.1 Analýza obtížných úloh

Analýza úloh z obecné chemie odhalila jedno obtížné téma a čtyři potenciálně-obtížná témata vyžadované v rámci přijímacího řízení. Mezi **obtížná témata obecné chemie**, tedy taková, ve kterých úspěšnost jednotlivých položek nepřesáhla 30 %, patřila *stavová rovnice ideálního plynu včetně výpočtu na základě zápisu chemické rovnice*. Toto téma bylo hlavním předmětem tří úloh. Dvě úlohy byly zaměřeny na výpočet změny tlaku, která nastala v důsledku hoření práškové glukosy v uzavřené nádobě. Obě úlohy vykazovaly úspěšnost přibližně 20 %. Navzdory skutečnosti, že obě úlohy měly téměř shodné zadání (pouze různé pořadí nabízených alternativ), dosahovala citlivost různých hodnot – ULI(1/2) hodnoty 0,16, resp. 0,25; RIR hodnoty 0,18, resp. RIR 0,29. S ohledem na náročnost úlohy se tedy jedná o položky s přijatelnou citlivostí. Pro vyřešení úlohy bylo zapotřebí provést větší počet kroků nutných k jejímu vyřešení (zápis a následné vyčíslení chemické rovnice, výpočet změny látkového množství plynných složek, dosazení do stavové rovnice pro ideální plyn a porovnání změny tlaku v souvislosti se změnou teploty a též změnou látkového množství). Třetí úlohou byla úloha, která byla zaměřena na reakci železa s kyselinou sírovou za laboratorních podmínek. Úspěšnost této úlohy činila 25 % a lze předpokládat, že nízká úspěšnost úlohy souvisela se zvolením laboratorních podmínek nikoliv podmínek normálních. Navzdory poměrně nízké úspěšnosti vykazovala tato položka velmi vysokou citlivost – ULI(1/2) = 0,30, RIR = 0,36. Na základě analýzy výše uvedených učebnic bylo zjištěno, že úlohy obdobného typu nebyly v těchto učebnicích středoškolské chemie uvedeny. V Přehledu středoškolské chemie autorů Vacíka a kol. (1999) byla sice uvedena stavová rovnice, avšak ani jeden z řešených příkladů neobsahoval aplikaci daného vztahu.

Na druhou stranu mezi analyzovanými položkami bylo dalších osm úloh, ve kterých byl při výpočtu použit výpočet objemu plynu za normálních podmínek (teplota 0 °C, tlak 101,325 kPa) s tím, že tyto položky vykazovaly vyšší úspěšnost (62 % až 90 %) i citlivost – koeficienty ULI(1/2) a RIR nabývaly hodnot na intervalu 0,15 až 0,46, tedy jednalo se o úlohy s akceptovatelnou až velmi dobrou citlivostí.

V rámci analyzovaných úloh lze shrnout, že pokud k vyřešení úlohy je nutné použít výpočet objemu plynu za normálních podmínek, jedná se o úlohy vhodné pro přijímací řízení. V RVP G není explicitně učivo stavová rovnice ideálního plynu ani učivo molární objem plynu zmíněno, avšak témata lze zařadit k učivu výpočty v chemii.

Mezi *potenciálně-obtížná témata* chemie patřila tato témata: (i) výpočet pH; (ii) elektronová konfigurace (určování počtu volných ne vazebných elektronových párů částic); (iii) iontový součin vody a (iv) elektrolýza.

Úlohy zaměřené na *výpočet pH* se velmi lišily svou obtížností. Okruhem úloh s nízkou úspěšností byl výpočet pH směsi, která vznikla smícháním roztoku hydroxidu a roztoku kyseliny. Tento typ úlohy byl zadán celkem v šesti úlohách – u všech těchto úloh se pohybovala úspěšnost mezi 18 % a 27 %, koeficienty citlivosti se pohybovaly mezi hodnotami 0,13 a 0,42. S ohledem na velmi vysokou náročnost úlohy se tedy obecně jednalo o položky s akceptovatelnou až velmi dobrou citlivostí. Zvýšená náročnost tohoto typu úloh byla zaznamenána již v přechozích studiích – např. Belháčové (20092002) a v některých úlohách též u Martinové (20012001). Nelze však tvrdit, že téma výpočty pH je jako celek téma náročné, neboť úlohy zaměřující se na výpočet pH roztoku kyseliny vykazovaly poměrně vysokou úspěšnost (76 %) i přijatelnou citlivost – $ULI(1/2) = 0,18$; $RIR = 0,20$, což je v souladu se závěry Ruska et. al. (2021). Úlohy spadající do tohoto tématu ve vztahu k RVP G lze téma zařadit opět k učivu výpočty v chemii.

Druhým potenciálně-obtížným tématem bylo *určování počtu volných ne vazebných elektronových párů částic*. Do tohoto tématu bylo zařazeno celkem 5 položek, jejichž průměrná úspěšnost se velmi lišila a pohybovala od 23 % do 79 % s přijatelnou citlivostí jednotlivých položek s ohledem na úspěšnost každé z nich. Úlohy v tomto tématu bylo možné rozdělit do tří kategorií:

- určení počtu volných elektronových párů částice;
- porovnání počtu volných elektronových párů částice;
- zařazení prvku do skupiny na základě elektronové konfigurace.

Úspěšnost úloh, které spadaly do první zmíněné kategorie (např. Určení počtu volných elektronových párů sulfidového aniontu), dosahovala nízkých hodnot (přibližně 27 %), koeficienty citlivosti se pohybovaly mezi 0,14 a 0,33, což lze s ohledem na náročnost úloh považovat za přijatelnou až velmi vysokou citlivost. Do druhé zmíněné kategorie spadala pouze jedna úloha, ve které bylo zadáno 5 částic (atomů či iontů) a úkolem uchazečů bylo vybrat částice se shodným počtem volných elektronových párů. Tato úloha vykazovala velmi dobré parametry – úspěšnost 44 %, koeficienty citlivosti – $ULI(1/2) = 0,32$; $RIR = 0,37$, což svědčí o velmi vysoké citlivosti. Uvedené výsledky korespondují se závěry Martinové (2001), která upozorňuje na nedostatečné porozumění některých pojmů – např. volný elektronový pár. Ve třetí uvedené kategorii byly dvě úlohy, jejichž úspěšnost činila 78 %, resp. 79 % a koeficienty citlivosti se pohybovaly mezi 0,13 a 0,31, což lze s ohledem na poměrně nízkou náročnost úloh považovat za přijatelnou citlivost. Z uvedených dat plyne, že způsob zadání úlohy z tohoto tématu chemie silně ovlivňuje úspěšnost, nikoliv však citlivost položky. V RVP G učivo ne vazebné elektronové páry není explicitně zmíněno, nicméně téma lze zařadit k učivu stavba atomu a chemická vazba.

Třetím potenciálně-obtížným tématem byl **iontový součin vody**. Toto téma obsahovalo čtyři úlohy. Dvě úlohy byly zaměřeny na výpočet iontového součinu vody, další dvě na souvislost mezi iontovým součinem vody a koncentrací oxoniových, resp. hydroxidových iontů. Obtížnější z úloh zaměřených na výpočet iontového součinu vody (výpočet iontového součinu vody na základě zadané koncentrace hydroxidových aniontů při dané teplotě) dosahovala hodnoty úspěšnosti 25 %, dále pro tuto úlohu byly vypočteny koeficienty $ULI(1/2) = 0,18$ a $RIR = 0,14$, které poukazují na citlivost úlohy blížící se hranici přijatelnosti a svědčící spíše o tom, že jak úspěšní tak neúspěšní řešitelé správnou odpověď pouze tipovali. Druhá úloha spadající do téže kategorie dosahovala úspěšnosti 38 % a velmi výborné citlivosti $ULI(1/2) = 0,34$, $RIR = 0,43$. Do druhé kategorie spadaly úlohy (vztah mezi koncentrací hydroxidových a oxoniových iontů) s relativně vyšší úspěšností (úspěšnost dosahovala 54 %, resp. 56 %) a s koeficienty citlivosti mezi 0,23 až 0,34, což svědčí o vysoké citlivosti daných úloh. Co se týče analýzy tématu v učebnicích chemie, bylo téma iontový součin vody zmíněno spíše okrajově, např. v učebnici autorů Vacík et al. (1999). V RVP G učivo iontový součin není přímo zmíněno, avšak lze jej zařadit k učivu výpočty v chemii a chemická rovnováha.

Posledním potenciálně-obtížným tématem bylo téma *elektrolýza*, které bylo obsaženo ve dvou typech úloh. Prvním typem je elektrolýza směsi látek (s využitím zadaných elektrodových potenciálů), druhým typem je elektrolýza chemicky čisté látky. Dvě úlohy zaměřující se na elektrolýzu směsi látek vykazovaly úspěšnost 21 %, resp. 39 %. Koeficienty citlivosti těchto úloh se pohybovaly na intervalu 0,02 až 0,25, což svědčí o nedostatečné, popř. omezené citlivosti daných položek. Výrazně odlišnou úspěšnost (přesahující 70 %) vykazovaly dvě úlohy zaměřující se na elektrolýzu čisté látky. Tyto položky vykazovaly též velmi vysokou citlivost – koeficienty citlivosti ležely na intervalu 0,30 až 0,42. Z uvedených dat je možné se

domnívat, že uchazeči o studium ovládají základní úlohy na elektrolýzu tavenin čistých látek, úlohy zaměřené na elektrolýzy tavenin směsí, ve kterých je nutné využít zadané elektrodové potenciály jednotlivých kovů, jim činí obtíž. Na základě analýzy učebnic středoškolské chemie bylo zjištěno, že tato problematika je podrobněji zmíněna pouze v Chemii pro čtyřletá gymnázia (Mareček et al., 2005b). V RVP G učivo elektrolýza není explicitně zmíněno, nicméně téma lze zařadit k učivu výpočty v chemii.

5.2 Analýza úloh s nedostatečnou citlivostí

Analýza úloh z obecné chemie odhalila tři témata, která obsahovala necitlivé úlohy. Mezi tato témata patřily (i) tepelný rozklad látky; (ii) rychlostní rovnice a (iii) chemické rovnováhy.

První vymezené téma *tepelný rozklad látek* bylo kromě jedné necitlivé úlohy zařazeno v dalších jedenácti úlohách, které byly zadány v rámci přijímacího řízení. Necitlivá úloha se zaměřovala na tepelný rozklad dusičnanu olovnatého, správné vyčíslení redoxní reakce a následný výpočet látkového množství jednoho z produktů. Přestože úloha vykázala poměrně vysokou úspěšnost 69 %, nelze ji považovat za vhodnou vzhledem k vypočteným koeficientům citlivosti ($ULI(1/2) = 0,08$; $RIR = 0,13$). Úloha kombinovala výpočet molární hmotnosti produktu spolu s vyčíslením slovně-zadané redoxní rovnice. Na rozdíl od ostatních úloh, které se týkaly stejného tématu (tepelný rozklad látek), byla pro úspěšné vyřešení této úlohy potřeba vysokého počet kroků, což mohlo být příčinou nižší citlivosti. Jedním z možných důvodů neúspěšného řešení mohlo být chybné vyčíslení chemické rovnice, což vedlo k nevhodnosti jednoho z nabízených distraktorů (tj. tuto chybnou odpověď volili častěji úspěšnější řešitelé než řešitelé méně úspěšní) a úloha tak nedostatečně rozlišovala mezi řešiteli úspěšnými a neúspěšnými. Obdobně nízký koeficient citlivosti vykazovala i velmi obdobná položka, která byla zaměřena na tepelný rozklad oxidu rtuťnatého. Úloha vykazovala velmi vysokou úspěšnost 91 % avšak s tím spojenou nízkou citlivostí ($ULI(1/2) = 0,10$ a $RIR = 0,23$). Nízká hodnota koeficientu $ULI(1/2)$ je způsobena velmi vysokou úspěšností, což je limitujícím faktorem pro tento koeficient citlivosti a při zohlednění koeficientu RIR lze tuto položku považovat za přijatelně citlivou. Je pravděpodobné, že vysoká úspěšnost v úloze byla způsobena nízkým počtem kroků k jejímu úspěšnému vyřešení. Na úlohu lze tedy pohlížet jako na úlohu motivační bez dostatečné funkce rozlišit úspěšné uchazeče od uchazečů neúspěšných. Ostatní úlohy týkající se tématu tepelný rozklad látek byly zaměřeny buď na výpočet molární hmotnosti rozkládané látky či na výpočet poměru objemů/hmotností dvou látek vystupujících v dané reakci a vykazovaly vhodnou úspěšnost (50 % až 73 %) i citlivost. V předchozí studii Štefanová (2005) poukázovala právě na úlohy zaměřené na tepelný rozklad látek jako na úlohy s vysokou citlivostí. Výše uvedená zjištění lze shrnout následovně: úlohy zaměřené na výpočty spojené s tepelným rozkladem látky jsou vhodné, pokud nezahrnují další obtížný krok, např. vyčíslení slovně zadané redoxní reakce. Vzhledem k RVP G lze výše analyzovaný tematický celek zařadit k učivu výpočty v chemii.

Druhé vymezené téma, které se zaměřovalo na *rychlostní rovnici*, bylo obsaženo ve dvou úlohách. Obě tyto úlohy vykázaly úspěšnost přibližně 35 %, avšak též poměrně nízkou citlivostí – hodnoty $ULI(1/2)$ a RIR se pohybovaly na intervalu 0,12 až 0,20. Předpokladem nízké citlivosti mohlo být, že žáci správnou odpověď pouze tipovali či se na správnou odpověď snažili přijít doplněním zadaných hodnot do výsledných vztahů (úloha byla zaměřena na určení závislosti rychlosti na koncentraci výchozích látek tzv. metodou počátečních rychlostí). Úloha tedy netestovala znalosti z oboru chemie a nerozlišovala úspěšné uchazeče od uchazečů neúspěšných. Přestože v RVP G téma přímo souvisí s učivem rychlost chemických reakcí, na základě analýzy středoškolských učebnic chemie se zjistilo, že problematika byla stručně vysvětlena pouze v učebnici autorů Vacík et al. (1999).

Posledním tématem, ve kterém byly nalezeny necitlivé úlohy, byly *chemické rovnováhy*. Toto téma bylo obsaženo celkem v deseti úlohách, které byly zaměřeny na ovlivňování rovnovážné konstanty přidáním či odebráním reaktantu, resp. produktu, ustanovení rovnováhy v závislosti na pH a na výpočet rovnovážné konstanty. Dvě úlohy z deseti byly označeny jako necitlivé. Jedna byla zaměřena na dvojici chemických rovnováh (a posunu těchto rovnováh v závislosti na změně pH; rovnováhy byly v zadání zapsány), druhá byla zaměřena na ovlivnění chemické rovnováhy mezi chromanem a dichromanem v závislosti na změně pH (celá reakce byla opět v zadání zapsána). Koeficienty citlivosti těchto dvou položek se pohybovaly na intervalu 0,07 až 0,21, což poukazuje na velmi malou rozlišovací schopnost úlohy. Je možné předpokládat, že možnou příčinou snížené citlivosti bylo, stejně jako v již zmíněné úloze o tepelném rozkladu dusičnanu olovnatého, vysoký počet kroků nutný k úspěšnému vyřešení úlohy. Tuto skutečnost podporuje i fakt, že citlivost zbylých osmi úloh byla vyhovující, v některých případech dokonce velmi dobrá, neboť koeficienty citlivosti nabývaly hodnot 0,21 až 0,47. Vzhledem k tomu, že Martincová (2001) i Štefanová (2005) ve svém výzkumu poznamenaly, že úlohy zaměřené na problematiku chemických rovnováh obvykle vykazují zvýšenou náročnost, předpokládalo se, že tyto úlohy budou celkově vykazovat nižší úspěšnost. Nicméně úspěšnost obou necitlivých úloh činila 43 %, resp. 47 %. Úspěšnost všech (citlivých i necitlivých) námi analyzovaných úloh zaměřených na chemické rovnováhy kolísala a pohybovala se mezi 38 % a 64 %. Výše

uvedené skutečnosti lze shrnout tak, že úlohy zaměřené na chemické rovnováhy jsou vhodné za předpokladu, že k vyřešení úlohy není potřeba uvědomění si dalších souvislostí, např. vztahu mezi koncentrací oxoniových kationtů a pH. Téma chemická rovnováha je učivem uvedeným v RVP G.

5.3 Úlohy dostatečně citlivé s adekvátní obtížností

Vzhledem k tomu, že předchozí podkapitoly se zaměřily především na analýzu velmi obtížných či necitlivých úloh, v této podkapitole jsou představeny úlohy, které naopak vykazovaly adekvátní úspěšnost (30 až 80 %) a zároveň přijatelnou citlivost, tedy položky vhodné pro účely přijímacího řízení. Celkem bylo vytyčeno devět témat, které právě takovéto úlohy obsahovaly, z nichž pět bylo představeno v kapitolách 5.1 a 5.2: (i) výpočet pH kyseliny/zásady o dané koncentraci (nikoliv však směsi látek); (ii) elektronová konfigurace; (iii) elektrolýza chemicky čisté látky; (iv) tepelný rozklad látek; (v) chemické rovnováhy; (vi) teorie kyselin a zásad; (vii) výběr pravdivého tvrzení o kationtech; (viii) termochemie a (ix) redoxní reakce.

V následujícím textu jsou uvedena zbývající čtyři, dosud nepředstavená, témata z obecné chemie.

Úlohy zaměřující se na *teorii kyselin a zásad* byly v rámci přijímacího řízení zadány celkově čtyři. Ve všech případech byla dána dvojice částic a úkolem uchazeče bylo vybrat z nabízených možností tu, která obsahuje částice, které by ze zadané dvojice částic vznikly v případě, že by tyto částice chovaly jako kyseliny. Všechny úlohy vykazovaly velice vysokou citlivost. Koeficienty citlivosti nabývaly hodnot od 0,32 do 0,54. Což poukazuje na to, že všechny položky dostatečně rozlišují úspěšné uchazeče od uchazečů neúspěšných a jsou velmi vhodnými pro účely přijímacího řízení. Úspěšnost jednotlivých úloh však byla různorodá a pohybovala se mezi 39 % a 67 %. Příčina tohoto rozptylu mohla souviset s počtem nutných kroků vedoucích k vyřešení jednotlivých úloh. Uvedené zjištění je v rozporu se závěry Štefanové (2005), která toto téma obecně označila za obtížné, avšak v souladu s Martincovou (2001). Vzhledem k tomu, že v RVP G není téma kyselin a zásad explicitně zmíněno, lze jej zařadit k učivu vlastnosti látek.

Analýzované testy dále obsahovaly dvě úlohy zaměřující se na *výběr pravdivého tvrzení o kationtech* (oxoniovém, resp. amonném). Tyto úlohy se zaměřovaly na strukturu iontu, typ vazby mezi jednotlivými částicemi tvořící ion a vlastnosti daného iontu. Obě úlohy vykazovaly poměrně vysokou úspěšnost – 64 %, resp. 66 % a velmi vysokou citlivost – koeficienty citlivosti nabývaly hodnot mezi 0,30 a 0,41. Vzhledem k vypočteným charakteristikám úloh lze tyto úlohy označit jako vhodné pro účely přijímacího řízení. Ve vztahu k RVP G lze téma zařadit k učivu stavba atomu či chemická vazba.

Tématermochemie bylo obsaženo ve čtyřech úlohách s různou úspěšností, která se pohybovala mezi 45 % a 69 % s tím, že všechny úlohy vykazovaly adekvátní až vysoké koeficienty citlivosti (0,25 až 0,42). Rovněž se jedná o položky vhodné do přijímacího řízení. Zjištění je v rozporu se závěry Martincové (2001) a též se závěry autorů Tilahun a Tirfu (2016), kteří téma termochemie obecně shledávají jako téma velmi obtížné. Co se týče RVP G téma termochemie je obsaženo v učivu tepelné přeměny při chemických reakcích.

Analýzované testy obsahovaly celkem deset úloh zaměřující se na *redoxní reakce*. Tyto úlohy mohou být dále děleny na úlohy zaměřené na vyčíslení redoxní rovnice a na úlohy zaměřené na výběr redoxní rovnice. Úspěšnost úloh zaměřených na vyčíslování redoxní reakce se pohybovala mezi 54 % a 87 %, koeficienty citlivosti nabývají hodnot na intervalu 0,18 až 0,41. Rovněž úlohy zaměřené na výběr redoxní rovnice z nabídky nabývají obdobných hodnot – úspěšnost mezi 71 % a 86 % a koeficienty citlivosti na intervalu 0,20 až 0,33. Úlohy spadající do tohoto tématu tedy vykazovaly obdobné vlastnosti, a to poměrně vysokou úspěšnost i citlivost nezávisle na způsobu zadání a zvoleném úkolu. Uvedené závěry jsou ve shodě s výzkumem Martincové (2001), která elementární redoxní reakce (s nízkým počtem myšlenkových kroků k úspěšnému vyřešení úlohy) rovněž uvádí jako příklad úloh s nadprůměrnou úspěšností. Belháčová (2002) uvádí vyčíslování chemických reakcí jako oblast s velmi vysokou úspěšností, avšak z její práce není patrné, zda v rámci přijímacího řízení byly zadány redoxní reakce či reakce, během nichž nedocházelo ke změně oxidačních čísel prvků. V RVP G učivo redoxní reakce není přímo zmíněno, nicméně je možné jej zařadit k učivu stavba atomu, chemická vazba či vlastnosti látek.

6 Limity výzkumu

Cílem studie byla analýza oborových testů z chemie zadávaných v rámci přijímacího řízení na PřF UK. Uvedené závěry tedy nelze zobecnit na všechny absolventy českých či dokonce zahraničních středních škol a taktéž na realizaci přijímacího řízení na jiných českých či zahraničních vysokých školách.

Dále je důležité upozornit na skutečnost, že necitlivost je vlastnost úlohy (vycházející z jejího zadání) nikoliv tématu, kterého se úloha týká. Zařazení necitlivé úlohy do tematického celku si kladlo za cíl odhalit důvody necitlivosti dané úlohy.

Podrobnější komparaci zjištěného závěru se závěry v teoretické části uvedených studií nebylo možné provést vzhledem k malému průniku analyzovaných témat z obecné chemie. V teoretické části uvedené studie se vždy zabíraly pouze částí námi analyzovaných tematických celků.

Prezentovaná studie je dílčí studií, která byla zaměřena pouze na položkovou analýzu přijímacího testu jako jednoho možného prediktoru akademické úspěšnosti, a to vzhledem k rozsahu článku pouze na položkovou analýzu úloh spadající do oblasti obecné chemie. Na studii navazují další analýzy autorů článku – analýza úloh spadající do zbylých oblastí chemie (anorganické, organické a biochemie), které slouží jako podklad nejen pro úpravu přijímacího řízení na PřF UK, ale i úpravu obsahu povinných předmětů v prvním ročníku studijního programu Chemie.

7 Závěr

V rámci analýzy oborového testu z chemie použitých v letech 2016 až 2019 v rámci přijímacího řízení na PřF UK bylo odhaleno celkem pět tematických celků z oblasti obecné chemie, které obsahovaly úlohy obtížné pro uchazeče o studium chemie na PřF UK (stavová rovnice ideálního plynu (včetně výpočtu na základě zápisu chemické rovnice), výpočet pH, elektronová konfigurace, iontový součin vody a elektrolyza). Na obtížnost některých úloh již upozorňovaly dříve realizované studie. S ohledem na přetrvávající nízkou úspěšnost v těchto oblastech doporučujeme těmto úsekům věnovat při přípravě budoucích studentů chemicky zaměřených studijních programů zvýšenou pozornost v rámci studia na vysoké škole.

Položková analýza následně odhalila tři témata, která obsahovala necitlivé úlohy (tepelný rozklad látky, rychlostní rovnice a chemické rovnováhy). Necitlivost úlohy často souvisela s vysokým počtem kroků nutných k jejímu úspěšnému vyřešení. Na všechny v článku analyzované necitlivé úlohy vzhledem k tomu, že nejsou schopné odlišit úspěšné uchazeče od uchazečů neúspěšných, je možné nahlížet jako na úlohy nevhodné pro přijímací řízení, vyjma úloh s motivační funkcí (tedy necitlivých úloh s vysokou úspěšností).

Podrobnou analýzou úloh bylo vytýčeno pět témat, které lze považovat za nevhodná pro účely přijímacího řízení (aplikace stavové rovnice ideálního plynu, výpočet pH směsi (kyselina + zásada), iontový součin vody, elektrolyza směsi látek s využitím elektrodových potenciálů a závislost rychlosti reakce na počátečních koncentracích výchozích látek). Nevhodnost témat plyne ze skutečnosti, že tato témata jsou v běžných středoškolských učebnicích zmíněna pouze okrajově a úlohy, jejichž předmětem byla výše uvedená témata, byly buď velmi obtížné či (a zároveň) necitlivé.

V neposlední řadě položková analýza odhalila devět témat z oblasti obecné chemie, které obsahovaly pouze úlohy s adekvátní náročností a taktéž citlivostí (výpočet pH kyseliny/zásady o dané koncentraci (nikoliv však směsi látek), elektronová konfigurace, elektrolyza chemicky čisté látky, tepelný rozklad látek, chemické rovnováhy, teorie kyselin a zásad, struktura iontů, termochemie a redoxní reakce), přestože tři z témat (chemické rovnováhy, teorie kyselin a zásad a pH) byla považována v předchozích výzkumech za náročné partie středoškolské chemie. Všechna uvedená položky věnované uvedeným tématům lze považovat s ohledem na parametry jednotlivých položek za vhodná pro účely přijímacího řízení.

Výše uvedené závěry mohou být reflektovány nejen v úpravě přijímacího řízení na vysoké školy, ale i v obsahu přípravných kurzů k přijímacím zkouškám, případně i obsahu předmětu Obecná chemie na vysoké škole.

Poděkování

Tvorba příspěvku byla podpořena grantovým programem Univerzitní výzkumná centra UK č. UNCE/HUM/024 a projektem Progres Q17.

Literatura

Benešová, M., et al. (2002). *Odmaturuj z chemie* (1. vydání). Didaktis, spol. s r. o.

Belháčová, Z. (2009). *Přijímací zkoušky na vysoké školy a nová maturita z chemie* [Diplomová práce, Univerzita Karlova]. Repozitář Univerzity Karlovy. <https://dspace.cuni.cz/handle/20.500.11956/22792>

Berger, S. J. (2012). The rise and demise of the SAT: The University of California. *American Educational History*, 39(1), 165–180.

Brown, H. D. (2003). *Language assessment: principles and classroom practices* (1st ed.). Pearson/Longman.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–185. <https://doi.org/10.1017/S0140525X01003922>

- Čípera, J., Klímová, H., & Číperová, M. (1985). Ovlivňování složitosti úloh. *Pedagogika*, (26), 563–574.
- Herridge, M. (2016). *Student Identification of Problem Topics in General Chemistry*. [Diplomová práce, Missouri State University]. Archiv Missouri State University. <https://bearworks.missouristate.edu/theses/2964/>
- Huvarová, M. (2010). *Nejpoužívanější středoškolské učebnice chemie na gymnáziích*. [Bakalářská práce, Univerzita Palackého v Olomouci]. Archiv Univerzity Palackého. <https://theses.cz/id/bmn3n5/>
- Chráska, M. (1999). *Didaktické testy*. Paido.
- Chvál, et al. (2015). *Hodnocení výsledků vzdělávání didaktickými testy*. TISKÁRNA BÍLÝ SLON, s. r. o.
- Jedličková, A. (2007). *Přijímací zkoušky z chemie – analýza a tvorba úloh*. [Diplomová práce, Univerzita Karlova]. Repozitář Univerzity Karlovy. <https://dspace.cuni.cz/handle/20.500.11956/93513>
- Martincová, J. (2001). *Souhrnná analýza úloh přijímacích testů z chemie na Přírodovědeckou fakultu UK z let 1995, 1996, 1998 a 2000*. [Disertační práce, Univerzita Karlova]. Přírodovědecká Fakulta UK.
- Mareček, A., & Honza, J. (2005a). *Chemie pro čtyřletá gymnázia 1. díl* (3. vydání). Nakladatelství Olomouc, s. r. o.
- Mareček, A., & Honza, J. (2005b). *Chemie pro čtyřletá gymnázia 2. díl* (3. vydání). Nakladatelství Olomouc, s. r. o.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97. <https://doi.org/10.1037/h0043158>
- MŠMT ČR. (2007). *Rámcový vzdělávací program pro gymnázia: RVP G*. Výzkumný ústav pedagogický v Praze.
- Pikálková, S., Vojtěch, J., & Kleňha, D. (2014). *Úspěšnost absolventů středních škol ve vysokoškolském studiu, předčasné odchody ze vzdělávání*. Národní ústav pro vzdělávání.
- PřF UK (Přírodovědecká fakulta Univerzity Karlovy). (2021). *Fakulta / Uchazeči o studium / Podmínky přijímacího řízení k bakalářskému studiu na Přírodovědecké fakultě pro akademický rok 2021/2022*. <https://www.natur.cuni.cz/fakulta/uchazeci/podminky-prijimacihorizeni-k-bakalarskemustuui-na-prirodovedecke-fakulte-pro-akademicky-rok-2021-2022>
- Rubešová, J. (2009). Souvisí úspěšnost studia na vysoké škole se středoškolským prospěchem? *Pedagogická orientace*, 19(3), 89–103.
- Rusek, M., Vojtíš, K., & Chroustová, K. (2021). *An investigation into freshman chemistry teacher students' difficulty in performing chemistry calculations*. In M. Nodzynska (Ed.), *Scientific Thinking in Chemical Education* (pp. 61–68). Pedagogical University of Kraków.
- Škaloudová, A. (2003). *Predikce úspěšnosti ve studiu učitelství*. [Disertační práce, Univerzita Karlova]. Pedagogická fakulta UK.
- Šrámek, M., & Teplá, M. (2020). Prediktory akademického úspěchu. In A. Vargová & K. Szarka. *15. Medzinárodná konferencija študentov doktorandského štúdia v oblasti teórie prírodovedného vzdelovania (Zborník príspevkov)*. 1. vyd. Komárno: J. Selye University, Komárno, Slovak Republic, (s. 96–100).
- Šrámek, M., & Teplá, M. (2021). „Problematické“ partie středoškolské chemie vyplývající z analýzy přijímacího řízení na Přírodovědeckou fakultu Univerzity Karlovy. In V. Machková, *16th International seminar for PhD students of chemistry didactics and related doctoral study programs. PROCEEDINGS*. 1. vyd. Univerzita Hradec Králové (s. 101–108).
- Štefanová, L. (2005). *Přijímací zkoušky z chemie – analýza a tvorba úloh*. [Diplomová práce, Univerzita Karlova]. Repozitář Univerzity Karlovy. <https://dspace.cuni.cz/handle/20.500.11956/98334>
- Taherdoost, H. (2016). *Validity and reliability of the research instrument; How to test the validation of a questionnaire/survey in a research* (August 10, 2016). Available at SSRN: <https://ssrn.com/abstract=3205040> or <https://doi.org/10.2139/ssrn.3205040>
- Tilahun, K., & Tirfu, M. (2016). Common Difficulties Experiences by 12 Grade Students in Learning Chemistry in Ebinat Preparatory School. *Semantic Scholar*, 6(2), 16–32.
- Vacík, J., et al. (1999). *Přehled středoškolské chemie* (4. vydání). Státní pedagogické nakladatelství, a. s.
- Varma, S. (2020). *Preliminary item statistics using point-biserial correlation and P-Values*. Eddata. https://eddata.com/wp-content/uploads/2015/11/EDS_Point_Biserial.pdf

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)

prof. RNDr. Hana Čtrnáctová, CSc.

RNDr. Martina Kekule, Ph.D.

prof. RNDr. Jarmila Novotná, CSc.

PhDr. Martin Rusek, Ph.D.

RNDr. Lenka Pavlasová, Ph.D.

Mezinárodní redakční rada

Dr. John Carroll (Nottingham Trent University, Great Britain)

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)

Dr. Paola Iannone (University of East Anglia, Norwich, Great Britain)

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)

PaedDr. Svatava Kubicová, CSc. (Ostravská univerzita v Ostravě)

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)

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. 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 Mgr. Zdeňka Janušová