Informatics in Education, 2025, Vol. 24, No. 1, 175–198

DOI: 10.15388/infedu.2025.06

Collaborative Solving of Computer Science Tasks: How Pairs Differ from Individuals?

Václav ŠIMANDL, Jiří VANÍČEK, Václav DOBIÁŠ
Department of Computing, Faculty of Education, University of South Bohemia in České Budějovice, České Budějovice, Czech Republic
e-mail: simandl@pf.jcu.cz, vanicek@pf.jcu.cz, dobias@pf.jcu.cz

Received: September 2024

Abstract. Research on collaborative learning of computer science has been conducted primarily in programming. This paper extends this area by including short tasks (such as those used in contests like the Bebras Challenge) that cover many other computer science topics. The aim of this research is to explore how problem-solving in pairs differs from individual approaches when tackling contest tasks.

An observational study was conducted on tens of thousands of contestants aged 8–12 years. Statistical tests showed that, compared to individuals, pairs have a higher ratio of correct answers and solve slightly more tasks. They seem to be more successful in some components of computational thinking and are more confident in their answers. In tasks with instant feedback, pairs find the correct solution faster than individuals. As the age of the pupils increases, a trend of decreasing advantages of working in pairs can be observed. These results could be useful for curriculum makers who create computer science textbooks.

Keywords: collaborative learning, computational thinking, pair problem-solving, Bebras Challenge, computer science tasks, primary school.

1. Introduction

The recent reform of computer science education in the Czech Republic introduced a new compulsory subject with computer science content in primary and lower-secondary schools. It emphasises group learning among the supported forms of work with pupils. There are many reasons for this approach. Some of them are of practical nature, such as providing enough computers or robotic aids for learning, or organising lessons more efficiently (a teacher who monitors the independent work of pupils checks half of the pupils' work compared to individual learning). However, the more important reasons are of pedagogical nature, aiming to improve the quality of learning. As discussed below, research studies have demonstrated the benefits of group learning in programming activities that involve hands-on and project-based activities. This study investigates whether there is such a benefit of pair learning when solving short computer science tasks. In these tasks, pupils work in pairs to decide on the correct answer to the question posed in a "real life" situation given by the task. The Czech version of Bebras Challenge which contains tasks of this nature was selected as the subject of the research. The aim of this research is to ascertain how solving these tasks in pairs differs from solving them individually.

2. Background

2.1. *Developing Computational Thinking through Collaborative Learning*

Learning is a sociocultural phenomenon. Almost a century ago, Vygotsky focused on the factor of cooperation of individuals in their learning. He emphasized the importance of communication and social interaction in learning in his theory of social development. He argued that children's cognitive development and their ability to learn can be guided and mediated by their social interactions (Vygotsky, 1980). The collaborative approach to learning actively engages pupils in processing and synthesizing information and concepts (Power, 2020). The social, psychological, and academic benefits of collaborative learning (Laal and Ghodsi, 2012) have been demonstrated many times and the approach is widely used in school practice. The potential of collaborative learning in computer science education has been highlighted by Schubert (2001). Currently, online tools such as Padlet, G Suite, and Slack support collaborative learning (Power, 2020).

Results applicable to the fields of collaborative learning and computational thinking (abbr. CT) can also be found in current pedagogical research. Laal and Ghodsi (2012) mention various benefits of collaborative learning:

- Promoting critical thinking skills
- Involving students actively in the learning process
- Modelling appropriate student problem-solving techniques

Lai and Wong (2021) created the Collaborative Computational Problem-Solving Competency Model. In it, the authors divided the competencies into three categories (social, affective, and cognitive learning outcomes). According to their findings, collaborative learning primarily affects cognitive learning outcomes (Lai and Wong, 2021).

Li *et al.* (2023) point out that developing CT through collaborative learning requires a certain level of metacognition of the participants. Students often fail to mobilize metacognition to regulate and control their cognitive activities in a collaboration. This results in poor learning effects (Li *et al.*, 2023).

Some research highlights the way the group is formed. According to Israel *et al.* (2016), in collaborative computing, the quality of the educational process is affected by the following influences:

- Group composition.
- Role of the adult/teacher.
- Defined problems that result in students experiencing uncertainty.

Lai and Wong (2021) identified these factors:

- Educational level.
- Programming environment.
- Study duration.
- Grouping method.
- Group size.

When forming groups, it is important to take into account the roles in the group. According to Keith *et al.* (2019), when teaching robotics, students have the opportunity to adopt roles within the group that are aligned to multiple dimensions of robotics (e.g., programmer, builder, and analyst).

Collaborative programming is largely an effective pedagogical tool for student outcomes in regard to programming practice (Scherer *et al.*, 2020). It helps to improve student CT (Li *et al.*, 2023; Wei *et al.*, 2021). Pair programming gives better results for beginners compared to programming as an individual (Iskrenovic-Momcilovic, 2019; Šimandl and Dobiáš, 2024). Such learning can also increase learners' self-efficacy (Wei *et al.*, 2021). Pair learning of programming can be a very promising way to educate girls, who often prefer to work in pairs (Bodaker and Rosenberg-Kima, 2022). However, in another study, lower levels of collaborative behaviours such as giving and receiving feedback and helping other partners were observed among girls (Hopcan *et al.*, 2022). The collaboration is more effective when both partners make substantive dialogue contributions, express uncertainty, and resolve it (Rodríguez *et al.*, 2017).

There are different opinions about the speed of problem solving. Bodaker and Rosenberg-Kima (2022) state that the disadvantage of learning programming in pairs can be seen in the longer time required to complete tasks. In contrast, according to Salleh *et al.* (2011), students in pairs usually complete tasks in less time than individual students.

Many studies whose stated theme is the development of CT have been conducted only on pair programming (Huang and Parker, 2023; Li *et al.*, 2023; de Jesus and SIlveira, 2022; Wu *et al.*, 2019). Also, meta-studies of research focused on the development of CT predominantly use resources from the area of pair programming (Lai and Wong, 2021; Li *et al.*, 2023).

Only a few sources (Bellettini *et al.*, 2022; Šimandl and Dobiáš, 2024) address the development of CT through collaborative learning using educational tools other than programming and robotics. Such suitable tools are, for example, short informatics tasks (Dagiené *et al.*, 2019) used in Bebras-like contests that develop CT.

2.2. *Collaborative Learning and the Revision of the Curriculum Framework in Czechia*

Since 2021, there has been a significant shift in the approach to teaching computer science and working with computers in Czechia (Ministry of Education, Youth and Sports of the Czech Republic, 2021).The new curriculum framework began to emphasise an author-centred approach instead of the previous user-centred approach. The aim of the new approach is to use the creation of both software (e.g., programming) and hardware (e.g., robotics) to understand computers and information systems and to look at the problems and challenges of today's world from a computer science perspective.

There have been changes in the approach to the goals of education, to the relationship of computer science to other subjects, and to the methods used. In addition, we can also observe a change in the organisation of computer science education in terms of the number of pupils in one learning unit. Before the reform, i.e., five or more years ago, learning with a computer was perceived as the work of an individual. One pupil per computer was the standard, which was the basis for the contemporary arguments on the need to equip schools with technology (Ministry of Education, Youth and Sports of the Czech Republic, 2008). From the point of view of that time, it made sense, because it is harder to apply a collaborative approach when developing digital literacy, training writing, or drawing on the computer. The current approach sees the benefits of two or more students working together to solve computing problems. For example, when developing a program, one pupil from the group can take on the role of "thinker, the head that thinks", and one can take on the role of "coder" and / or "checker". Computing emphasizes discussion as a learning method and a goal of education (Ministry of Education, Youth and Sports of the Czech Republic, 2021). Previously this was neither required nor seen as important in the education focused on the use of computers.

This change in approach can be seen in the content of computer science textbooks. By analysing the tasks, we can see that old textbooks focused on digital literacy and user access (Vaníček and Řezníček, 2004; Vaníček, 2012) contain mostly tasks for individuals. While solving them, it would be difficult to work in pairs in a way that pupils do not interfere with each other and that such learning brings benefits for developing the required skills. The new computer science textbooks that have been produced since 2018 (Berki and Drábková, 2020; Filipi *et al.*, 2020; Vaníček, 2019; Procházka *et al.*, 2020; Kalaš *et al.*, 2018) often mention collaborative work and discussion as a supported form of learning. Most of the activities from these textbooks are applicable when working collaboratively.

A similar trend from individual to team-based can be observed in computer science contests. Older contests, such as the Olympiad in Informatics, are organised in Czechia (under the name Mathematical Olympiad category P) as in many other countries as a contest for individuals. Younger contests such as the VEX IQ Robotics Competition, the FIRST Lego League Challenge, or the World Robot Olympiad are organised for teams of contestants.

In some countries (Switzerland, France, Germany), it is possible to compete in pairs in Bebras-like contests in the younger age categories (Datzko-Thut, 2024). In Italy, contestants are only allowed to compete in teams of three or four pupils. In Czechia, there has been an evolution from contests for individuals only, through allowing pairs at primary schools about three years ago, to currently also enabling pairs for younger pupils of lower-secondary schools.

3. Motivation and Aim of the Research

Allowing pairs to compete in the Bebras Challenge alongside individuals creates a non-homogeneous field of contestants. The question is whether pairs are at an advantage over individuals, and if so, which indicators point to this.

According to Datzko-Thut and Pohl (2024), it is unclear whether the content of the task, the opportunity to interact while solving it, its difficulty, or other characteristics affect the success rate of pairs. It is also unclear how pairs interact during the solution, especially under time pressure. Findings that answer these questions may be useful for curriculum designers and organizers of computer science contests.

The aim of this research is to map the performance of pairs in a contest compared to individuals and to find out in what ways they differ from each other. The research is focused on pupils at primary school and early lower secondary school. Based on the above research aim, we set the following research questions:

- **RQ1:** What is the success rate of pairs compared to individuals?
  - We are interested in whether pairs perform better than individuals on the test. We are also interested in what components of CT and types of answers this difference is manifested.
- **RQ2:** What is the efficiency of pairs compared to individuals?
  - We are interested in whether pairs solve more tasks than individuals and whether they solve the tasks in a shorter time. In the case of programming tasks, we are also concerned with whether pairs solve them in fewer iterations.
- **RQ3:** What is the stability of the choice of answers for pairs compared to individuals?
  - We are interested in whether individuals or pairs are more confident in their answers. We are interested in the extent that they revisit tasks they have already answered, reconsider their opinion, and whether changing their opinion is beneficial to the solution.

These three research questions were divided into sub-questions when conducting the research. Their specification is described in the chapter *Data processing and analysis*. Another important aim of the research is to uncover age trends in the indicators observed in RQ1 to RQ3.

4. Methodology

To meet the research aims, we conducted an observational study on data obtained from the Czech version of the Bebras Challenge. In this contest, not only individuals but also pairs can compete in certain age categories. In the years 2021 and 2022, this was only possible in the category for Grades 4 and 5. In 2023, pairs were allowed to compete in the categories for Grades 3 to 6.

4.1. *Selection of Tasks, Testing*

The national round of the Bebras Challenge was used as a basis for this research. Thus, the tasks were not selected with regard to the possibility of solvers working together in a team. The international Bebras database (Bebras.org, 2024) was a source of the contest tasks. Each such task focused on one or more components of CT. The tasks were selected based on the assessment of their quality and attractiveness, age appropriateness, and compliance with the content of the Czech national curriculum of computer science. Some of the tasks, especially programming tasks that involve building a program from blocks (Vaníček *et al.*, 2023; Vaníček *et al.*, 2022), were created at the national level, as such tasks do not appear in the international database.

According to the method of answering, the tasks were divided into several types:

- Classical Bebras tasks (multiple-choice, typing).
- Interactive (click to choose, drag & drop).
- Programming (building a program from blocks).

Some tasks (especially programming ones) provided feedback on the correctness of the answer after submission, allowing contestants to correct their solution. However, most tasks did not provide such feedback.

The tasks in the Czech version of the Bebras Challenge are classified according to components of CT. The definition of CT provided by Wing (2006) which emphasises the role of the executor was used. Concepts of algorithmisation, abstraction, decomposition, evaluation, and generalisation, as proposed by Selby and Woollard (2014), was accepted as the components of CT. The organisers of the contest assign to each task the CT components which the task is related to and whose mastery is demonstrated by the successful completion of the task (Vaníček, 2024).

Each contest test comprised 12 tasks, with the youngest pupils given 30 minutes and the older ones allowed 40 minutes to complete them. It was possible to freely switch between tasks and return to those already answered. For a correct answer, contestants received points according to the predetermined difficulty of the task. For an incorrect answer, some of the points were removed.

4.2. *Respondents and their Selection*

The research was carried out among Czech contestants of the Bebras Challenge in the above-mentioned years and categories. Respondents were pupils aged 9–11 in the years 2021 and 2022, and aged 8–12 in the year 2023. To ensure the integrity of the data, we excluded “pseudo-contestants” who did not complete the test or finished it in less than 5 minutes. We excluded 5% of contestants in this way.

In order to be able to better select a representative sample of contestants, e.g. to minimize the potential bias from teachers pairing only weaker students together, we included only pairs from schools with majority of competing pairs and at least 7 pairs. Individuals were taken from schools with no pairs and more than 13 contestants. We chose these

Table 1
Number of respondents by grade and contest year

Grade	2021		2022		2023	
	Pairs	Individuals	Pairs	Individuals	Pairs	Individuals
3 (age 9)	-	-	-	-	157	2,812
4 (age 10)	181	5,444	642	13,250	875	18,923
5 (age 11)	265	11,937	608	19,929	732	21,023
6 (age 12)	-	-	-	-	351	27,034

numbers so that the results would not be influenced by schools where only a narrow selection of pupils was made instead of competing in whole classes.

In total, we included 127,974 pupils in the research, 120,352 individuals and 3811 pairs. The numbers for each grade and contest year are shown in Table 1.

4.3. *Collection of Data*

Data from testing, stored by the contest application, were used for the research. No personal data of the pupils were used; all data were anonymized before the analysis. We collected summary data about contestants: the school they attended, whether they were competing as a pair or as an individual, when they began competing and when they finished, and how many points they scored. We also collected data about the correctness of the final solution or not answering the task.

In 2023, we also collected data about each answer submitted, including a time stamp. For programming tasks, we thus had data about each execution of the pupil's program. This data provided detailed insight into how the test was solved. For example, we could identify situations where contestants were returning to previously solved tasks and changing the answers. We were also able to detect situations where contestants answered a task but later cancelled the answer.

4.4. *Data Processing and Analysis*

We divided the research questions into subquestions.

RQ1: What is the success rate of pairs compared to individuals?

We set three subquestions:

- **RQ1a:** Do pairs have a higher proportion of correct answers than individuals?
- **RQ1b:** In what components of CT is there a difference in success rates between pairs and individuals?
- **RQ1c:** In what types of answer is there a difference in success rate between individuals and pairs?

All subquestions were answered based on data from 2021 to 2023.

RQ1a: To answer this research question, we used information about the correctness of the task solutions. We understood correctness (here and in RQ1b and RQ1c) as the ratio of the number of correct answers to the sum of the number of correct and incorrect answers.

We tested the null hypothesis that the correctness of answers is the same for pairs and individuals.

RQ1b: We classified the contest tasks according to the components of CT. We dealt with the following components: abstraction (9 tasks), algorithmization (24), decomposition (7), evaluation (8), and generalisation (6). We investigated the correctness of the answers of individuals and pairs in answering the tasks containing the given part. Subsequently, we tested the null hypothesis that the correctness of answers to tasks containing a given component is the same for pairs and individuals.

RQ1c: We divided the tasks according to the type of answer. There were 8 programming tasks, 23 interactive tasks, and 17 classical Bebras tasks. We tested the null hypothesis that the correctness of a given type of answer is the same for pairs and individuals.

RQ2: What is the efficiency of pairs compared to individuals?

We set three subquestions:

- **RQ2a:** Do pairs answer the same number of tasks as individuals?
- **RQ2b:** Are pairs as fast in solving tasks as individuals?
- **RQ2c:** In what types of answer do pairs find the correct solution as fast as individuals?

All subquestions were answered based on data from the year 2023, because only these data allowed us to observe the process of task solving.

In subquestions where time is measured, we took into account that not all contestants were able to answer all the tasks within the time limit. In RQ2b, we calculated the time needed to answer the first two-thirds of the whole test. In RQ2c, where types of answers were investigated, this approach would reduce number of tasks, so we decided to examine all tasks but only from contestants who answered those tasks correctly.

RQ2a: To answer this research question, we used information about the number of tasks answered by each contestant. If a task was answered and then the answer was cancelled, we interpreted that task as unanswered. We examined the average number of tasks answered by individuals and pairs. We tested the null hypothesis that individuals and pairs answered on average the same number of tasks.

RQ2b: We needed all contestants to solve exactly the same set of tasks. The first third of the contest test contains easy tasks, the second one middle tasks and the last one difficult tasks. We calculated the time needed to answer the first two-thirds of the whole test. We used data about when contestants started competing, which tasks they answered, and when they did so.

Contestants who did not solve the test in the expected manner (e.g., they skipped some tasks or did not solve them in the recommended order) may have led to bias in our findings. We excluded such contestants from the analysis.

We investigated the average time it took for individuals and pairs to solve the tasks, testing the null hypothesis that individuals and pairs needed the same amount of time.

RQ2c: We used the classification of tasks according to the type of answer. There were 3 programming tasks, 10 interactive tasks and 11 classical Bebras tasks. We focused on successful solutions of these tasks, investigating the time needed to solve the tasks. We tested the null hypothesis that individuals and pairs needed the same amount of time to solve the task. In the case of programming tasks, we followed an analogous procedure also for the number of program builds. We interpreted the term "building a program" as not only assembling it from blocks, but also running it.

RQ3: What is the stability of the choice of answers for pairs compared to individuals?

We set three subquestions:

- **RQ3a:** Is the proportion of contestants who changed their opinion about the correctness of their answer the same for pairs and individuals?
- **RQ3b:** Do pairs return to previously solved tasks as often as individuals?
- **RQ3c:** For pairs and individuals, what impact do changes of opinion have on the final correctness of the answer?

All subquestions were answered based on data from the year 2023, because only these data allowed us to observe the process of task solving.

RQ3a: We used tasks that do not provide immediate feedback to the solver. There were 11 such tasks in the younger category and 10 in the older category. For further analysis, we used information about who answered which task and when. We were not only interested in the final answer, but in each submission of the answer. We were thus able to identify how many times contestants changed the answer in these tasks.

A special issue was that some answers were changed many times. An extreme example found in the data was a change of an answer 120 times. We believed that volatile contestants changing their answers many times were not actually changing their opinions. We thought that they were either using answer marking when discussing the answer, or mindlessly clicking on answers, i.e., they were playing. We believed that is not probable to change an opinion while solving a task more than four times so we decided to exclude them. It was 6.4% of all records.

For each task, we examined the proportion of pairs and individuals who changed their answer at least once. We tested the null hypothesis that the proportion of individuals and pairs who changed their answer was the same. Similarly, we tested the proportion of pairs and individuals who changed their answers more than once.

RQ3b: We identified the number of times contestants returned to previously completed task during the contest after dealing with another task. We examined the average number of returns for individuals and pairs. We then tested the null hypothesis that, on average, individuals and pairs returned to previously answered tasks equally often.

RQ3c: We used tasks that do not provide immediate feedback. We identified the time sequences of the answers to a given task and were interested in how often some answer scenarios were used, for example:

- A change of the answer.
- A correct answer on the first attempt.
- A changed answer that is better or worse than the previous one.

For pairs and individuals, we measured the relative frequencies of these scenarios and compared them according to the formula $\frac{\text{relative frequency of pairs}}{\text{relative frequency of individuals}} - 1$. If the result was positive, it meant that pairs used the scenario more often. Compared to a simple comparison of frequencies, this calculation helped us to see and compare trends between grades.

4.5. *Statistical Testing*

In some cases, we wanted to statistically compare data obtained in different grades. Different grades had different means, variances, and sample sizes. To reflect these differences, we used nonparametric graphical test of significance (Myllymäki and Mrkvička, 2024; Mrkvička *et al.*, 2021). This test allows to test several characteristics at the same time and still obtain formally correct results. This non-parametric test has no technical assumption about the distribution and homogeneity of the tested characteristics as other tests typically have. We used this test in research questions RQ1a; RQ1c; RQ2a; RQ2b; RQ2c; RQ3a; RQ3b.

We performed the data analysis using this test in R software, where the result is represented by a graph. In Fig. 1, we present the results of two tests for illustration. The left part of the figure shows the results of testing whether there are differences between pairs and individuals in the average time needed to solve a given set of tasks (RQ2b). The right part of the figure shows the results of testing whether the proportion of contestants who changed their opinion about the correctness of their answer is the same for pairs and individuals (RQ3a).

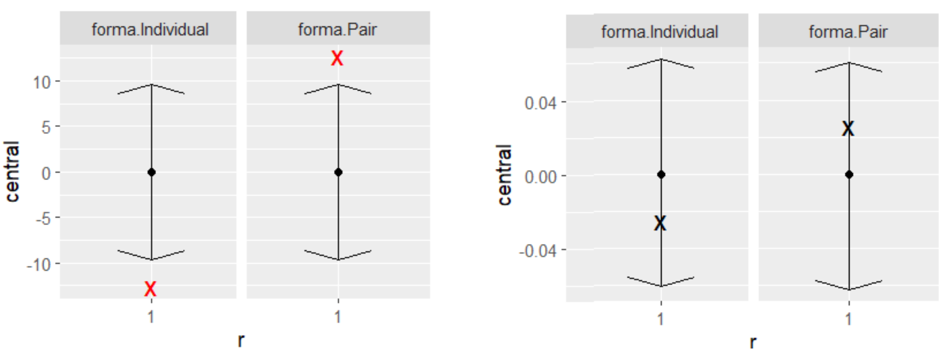

Fig. 1. Results of the nonparametric graphical test of significance.

As illustrated in the figure, the arrows denote the confidence intervals at the designated significance level, with the position of the cross indicating the magnitude of the observed difference. If the difference is statistically insignificant, the cross is situated within the confidence interval and is represented by black. Conversely, if the difference is deemed to be statistically significant, the cross is positioned outside the confidence interval and is indicated by red. The group for which the cross lies in positive values has a higher mean value of the specified characteristic. The distance of the cross from the central value indicates the effect size. The further the cross is from the confidence interval, the effect size is bigger.

Interpreting the graphs in Fig. 1, the test on the left shows a statistically significant difference where the mean of the pairs is greater. The test on the right shows statistically insignificant difference.

In the case of internally homogeneous data, we used the non-parametric two sample Mann-Whitney U test. This test was used in the research questions RQ2a; RQ2b; RQ3b. We used the Chi-square test to statistically examine the relationships of the categorical variables in the research questions RQ1b and RQ3c.

All statistical testing was based on null hypothesis significance testing. As the threshold for p-values, the alpha = 0.01 level of significance was used to avoid distortion of results caused by large research samples we had. For the same reason, the chosen non-parametric graphical test of significance provides the graphical output where the real difference together with its significance can be observed.

The analysis of trends with increasing age was done by simply comparing the values of different age groups of contestants.

5. Results

5.1. *RQ1: What is the Success Rate of Pairs Compared to Individuals?*

RQ1a: Do pairs have a higher proportion of correct answers than individuals?

We assessed the correctness of task-solving by both individuals and pairs. We found that while the correctness of pairs was 52.9%, the correctness of individuals was 48.8%. Using nonparametric graphical tests of significance (Mrkvička *et al.*, 2021), we found this difference to be statistically significant at the 0.01 level. It does mean that correctness of pairs is hard to justify under the hypothesis that pairs and individuals perform the same. We interpret this result that the correctness of pairs' answers was higher than individuals. The same interpretation of the statistically significant results we use also in the following sections.

The difference between pairs and individuals is 4.1 percentage points (%pt) in favour of pairs. Having 12 test problems, this means that pairs answered on average one half of one problem better. In other words, half of the pairs answered one task better than individuals. This difference can be illustrated by the fact that an individual in the middle of the ranking list of the contest would have moved up in the rankings by an estimated 5,6%, if he had competed in a pair.

Table 2

Differences in percentage points of correct answers between pairs and individuals by grades. Asterisk * indicates statistically significant differences at the α=0.01 level

Grade	3	4	5	6
Difference in %pt	7.5 *	6.5 *	4.4 *	3.4 *

The results of the general statistical testing mentioned above were shown in all observed grades. The difference in correctness between pairs and individuals decreased as the age of the pupils increased (see Table 2), being two times greater in the third grade compared to sixth grade. We interpret this to mean that the advantage of competing in pairs considerably decreases with increasing age, i.e., working in pairs is most beneficial for the youngest pupils.

RQ1b: In what component of CT is there a difference in success rates between pairs and individuals?

We investigated the correctness of answers to tasks according to the components of CT. We wondered whether it is advantageous to solve in pairs computing tasks focused on certain component of CT. We found the difference of correctness to be statistically significant only in tasks containing abstraction, algorithmization, and generalisation. We interpret it that the correctness of pairs was higher in these types of tasks but not in the tasks containing decomposition and evaluation (see Table 3).

RQ1c: In what types of answer is there a difference in success rate between individuals and pairs?

We examined the correctness of answers to the tasks according to the type of answer. We believe that programming tasks with their immediate feedback and their possibility to iterate the solution are more suitable for pairs as they allow discussion about the feedback. We found the difference in correctness to be statistically significant. We

Table 3

Differences between pairs and individuals in the correctness of answers to the tasks, according to the components of CT

Type of tasks	Correctness of answers to the tasks			Total number of tasks answered	
	Pairs	Individuals	Difference in %pt	Pairs	Individuals
ABS	51.2%	47.4%	3.8 *	9,472	256,278
ALG	58.7%	53.4%	5.3 *	18,524	594,183
DEK	44%	45.7%	-1.6	6,076	176,622
EVL	49.1%	47.4%	1.8	7,113	270,796
GEN	52.1%	46.6%	5.5 *	6,136	200,131

Table 4
Correctness of individuals' and pairs' answers to the tasks by answer type

Type of tasks	Correctness of answers to the tasks			Total number of tasks answered	
	Pairs	Individuals	Difference in %pt	Pairs	Individuals
Programming	63.1%	57.1%	6.1*	6,704	224,266
Interactive	49.4%	44.6%	4.8*	18,136	508,902
Classical Bebras	53.7%	50.5%	3.1*	16,422	527,490

interpret it that pairs' answers were more correct than that of individuals for all task types. The difference in programming tasks was higher than in the other answer types, two times higher than among classical multiple-choice and typing Bebras tasks (see Table 4).

5.2. *RQ2: What is the Efficiency of Pairs Compared to Individuals?*

RQ2a: Do pairs answer the same number of tasks as individuals?

We examined the number of tasks answered. We believe that pair cooperation leads to less tendency to give up tasks. Thus, pairs should solve more tasks. While pairs answered an average of 11.35 tasks, individuals answered 11.18. The difference between the two is statistically significant. Thus, we assume that pairs answered a higher number of tasks than individuals.

We examined the above differences by grades. As the fourth column of Table 5 demonstrates, these differences were statistically significant in grades 3 to 5. It does mean that in these grades the number of solved tasks is hard to justify under the hypothesis that pairs and individuals perform the same. We interpret this result that pairs answered more tasks than individuals. On the other hand, in Grade 6, pairs and individuals answered the same number of tasks.

Table 5 shows a trend towards a decreasing difference in the number of tasks answered between pairs and individuals as the age of the pupil increases. It shows that the need for collaboration in solving tasks decreases with the age of pupils.

Table 5
Number of tasks answered by pairs and individuals according to the grades

Grade	The average number of tasks answered by one contestant			Total number of contestants involved	
	Pairs	Individuals	Difference	Pairs	Individuals
3	11.30	10.80	0.50 *	157	2,812
4	11.64	11.35	0.29 *	875	18,923
5	11.46	11.26	0.20 *	731	21,024
6	11.64	11.52	0.13	351	27,034

Table 6
Time needed to solve the set of tasks solved by pairs and individuals according to the grades

Grade	The average time needed to solve the set of tasks, in seconds			Total number of sets of tasks answered	
	Pairs	Individuals	Difference	Pairs	Individuals
3	897	904	-7.3	1,622	125
4	806	795	11.3	12,649	668
5	958	943	15.5	12,900	492
6	967	871	96.9*	17,301	244

RQ2b: Are pairs as fast in solving tasks as individuals?

We investigated how solving time differs between pairs and individuals. We supposed that pairs do not want to give up solving the tasks so often and the discussion over the task needs some time. So, we assumed that pairs would spend more time.

In the analysis, we noticed a big difference in the proportion of excluded contestants who did not solve the set of tasks in the recommended order, or who did not answer some tasks. We excluded 27.7% of contestants from the pair category, compared to 36.3% from the individual category. This difference was statistically significant. This finding is supported by our other results that pairs answered more tasks than individuals and spent more time on the test. We interpret these findings to mean that pairs worked more systematically, i.e. they needed less to take advantage of the flexibility in the order of task solving provided by the contest environment. This may have been due to the fact that they understood the assignment more easily when reading together, were more confident that they would find the solution, and did not give up on the question so often.

The result was that pairs took an average of 888 seconds to solve the set of tasks, while individuals took 872 seconds. The difference between them is statistically significant. Thus, it can be assumed that pairs spent more time on the test than individuals. We do not perceive this result as important because when examined by grade (see Table 6), we only found a statistically significant difference between pairs and individuals in Grade 6 which was in favour of individuals. In the other grades, the results suggest that individuals and pairs worked at similar speed.

RQ2c: In what types of answer do pairs find the correct solution as fast as individuals?

We investigated the time needed to solve successfully one task according to the type of answer. We assumed that pairs would work slower as the discussion about the solution is time-consuming.

For programming tasks, average time of solving one task was 273 seconds for pairs and 299 seconds for individuals which is about 10% more. This difference is statistically significant and we interpret this result as pairs solved programming tasks in less time than individuals. For other types, the difference of solving time was statistically

Table 7
Time needed to solve tasks by pairs and individuals according to the type of answer

Type of tasks	Time needed to solve a task			Total number of tasks answered	
	Pairs	Individuals	Difference	Pairs	Individuals
Programming	273	299	-16*	1,927	61,116
Interactive	106	109	-3	4,409	98,002
Classical Bebras	111	110	1	5,754	209,028

insignificant (see Table 7). This finding suggests that the time taken to complete non-programming tasks is equivalent for pairs and individuals.

In programming tasks, pupils built a program and iterated their solutions based on immediate feedback. Therefore we examined one more indicator: number of program builds. We supposed that pairs would have lower number of builds as they have the opportunity to discuss the feedback rather than use the trial and error method.

Pairs needed an average of 7.77 builds to solve one task. Individuals needed 8.25 builds thus one half of one build per task more. This difference is also statistically significant. It can be said that pairs solved the tasks using fewer builds than individuals. According to both indicators, it seems that pairs solved programming tasks more effectively.

5.3. *RQ3: What is the Stability of the Choice of Answers for Pairs Compared to Individuals?*

RQ3a: Is the proportion of contestants who changed their opinion about the correctness of their answer the same for pairs and individuals?

In this research question, we examined the proportion of contestants who changed their answer. This proportion was 14.13% for pairs and 13.83% for individuals. The difference is not statistically significant. Thus, it can be said that pairs change their opinion about the correctness of their answer as often as individuals. These findings were also confirmed for multiple changes of opinion.

RQ3b: Do pairs return to previously solved tasks as often as individuals?

We examined the average number of returns to previously answered tasks. For pairs, the average number of returns during the whole contest was 0.56, for individuals it was 0.8. This difference is statistically significant. It does mean that number of returns of pairs is hard to justify under the hypothesis that pairs and individuals return equally often. In our opinion, 42% more returns of individuals can be seen as relatively big difference. Thus, it can be said that individuals are more likely to return to previously solved tasks than pairs.

We examined the above differences by grades (see Table 8). With the exception of Grade 3, we found a statistically significant difference. A trend of gradually increasing

Table 8

Differences between pairs and individuals in the average number of returns to previously solved tasks during the whole contest by grade

Grade	3	4	5	6
Difference	-0.077	-0.191 *	-0.226 *	-0.291 *

differences between pairs and individuals in the number of returns can be observed. It can therefore be assumed that pairs work more systematically than individuals as they get older.

RQ3c: For pairs and individuals, what impact do changes of opinion have on the final correctness of the answer?

In this research question, we observed the following answer scenarios (see Table 9):

- *Once, correct*: the contestant answered only once, and correctly.
- *Once, incorrect*: the contestant answered only once, but incorrectly.
- *Worse*: the contestant answered multiple times, one of the answers was correct, but the final answer was incorrect.
- *Improved*: the contestant answered multiple times, the contestant's first answer was incorrect, but his final answer was correct.
- *no Risk*: the contestant selected an answer but in the end decided not to answer.
- *Volatile*: the contestant changed the answer more than four times.

The quotient in Table 9 shows the ratio between the relative frequency of the scenario occurrence among pairs compared to individuals. In the *once, correct* scenario, the proportion of pairs is higher than the proportion of individuals. Conversely, in the *once, incorrect* scenario, the proportion of pairs is lower than the proportion of individuals. Pairs answered correctly on the first time more often than individuals.

In the *worse* scenario, the proportion of pairs appears to be lower than the proportion of individuals. Conversely, in the *improved* scenario, the proportion of pairs appears to be higher than the proportion of individuals. However, these differences are not statistically significant. Thus, it can be said that changes of opinion lead to both improved and worse answers in the case of pairs as often as in the case of individuals.

In the *no risk* scenario, the proportion of pairs is lower than the proportion of individuals. It can be said that pairs were more confident about their answers than individuals. In the *volatile* scenario, the proportion of pairs is higher than the proportion

Table 9

Quotient of pairs to individuals in each answer scenario

Scenario	once, correct	once, incorrect	worse	improved	no risk	volatile
Quotient	0.081*	-0.036*	-0.051	0.061	-0.319*	0.334*

of individuals. This may be because pairs were more likely to have a longer discussion about the correctness of their answer. During this, pupils might mark the answers they were thinking about. It is also possible that pupils are mindlessly clicking on objects on the screen. However, we do not find reasons why pairs should do it more often than individuals.

We also looked at trends according to increasing age (the data are shown in Table 10). We did not perform statistical testing due to small sample sizes and do not interpret individual results. Not all scenarios showed any trends, only the following trends were observed (see Fig. 2):

- *Once, correct*: decrease in the proportion of pairs
  For younger pupils, pairs choose the correct answer the first time noticeably more often than individuals, but this difference gradually disappears with age.
- *Worse*: increase in the proportion of pairs
  This is in line with the decreasing trend seen in the *improved* scenario. This suggests that the effectiveness of pair discussions decreases with age.
- *No risk*: decrease in the proportion of pairs
  Thus, it can be said that with increasing age, pairs' risk willingness increases, which could be due to their increasing self-efficacy. The feeling of working in a pair probably gives pupils more courage.

Table 10

Quotient of pairs to individuals in each answer scenario by grade

Grade	once, correct	once, incorrect	worse	improved	no risk	volatile
3	0.28	-0.07	-0.29	0.14	-0.29	-0.10
4	0.20	-0.09	-0.04	0.23	-0.27	0.19
5	0.10	-0.04	0.01	-0.05	-0.45	0.03
6	0.08	-0.08	0.17	-0.01	-0.45	0.83

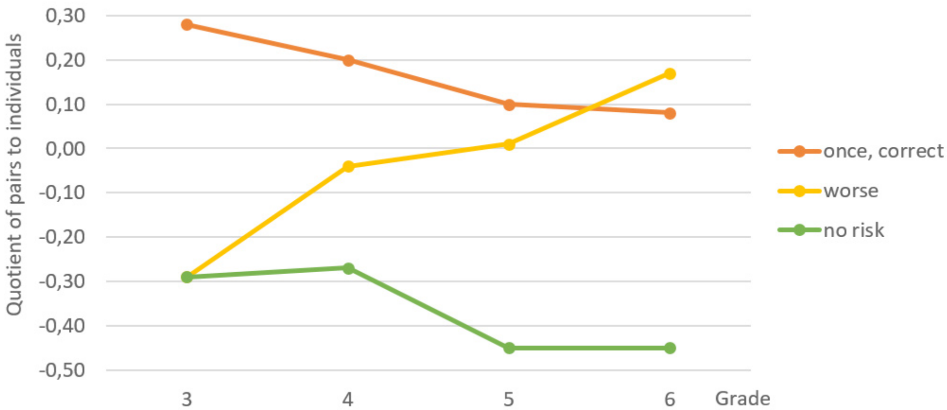

Fig. 2. Trends in scenarios with increasing age of pupils.

It can be seen in Table 10 in *volatile* scenario that there is a big increase in the proportion of pairs in Grade 6. These pairs are more likely to change the answer multiple times than the younger ones.

5.4. *Background to the Pairing of the Contestants*

To understand the context, in December 2023, we sent a questionnaire to teachers from schools where some of the pupils were competing in pairs. We asked them about the background to the pairing of the contestants. Of the 51 teachers, 18% said that pupils were ordered to compete in pairs: 8% stated capacity reasons (not enough computers available), while the remaining 10% reported that this was a requirement or educational intention of the school. In three-quarters of the schools where competing in pairs was optional, pupils preferred to compete in pairs, and in a quarter of these schools as individuals. 2% of schools did not allow the contestants to choose their teammate.

In the next question, we related the choice to compete in pairs to other alternative and modern pedagogy activities in the school. 55% of teachers answered that their school was characterised by a traditional approach. The others reported a leaning towards alternative approaches such as Waldorf or Montessori school program, project oriented education, formative assessment, or alternative directions of mathematics and writing teaching at the national level – Hejný mathematics (Hejný, 2012), Comenia script writing (Bartošová *et al.*, 2012), etc.

6. Discussion and Conclusion

6.1. *Interpretation of the Results*

We investigated the impact of competing in pairs or individuals in the Bebras Challenge on pupils aged 8 to 12. We interpret our findings that the correctness of the answers of pairs is higher than individuals. This difference is visible in all types of answers (multiple choice, interactive, building a program). It is more apparent in programming tasks than in the other task types. We think that programming tasks provide immediate feedback on the correctness of the answer and allow pairs to discuss how to fix their programs.

Pairs answer more tasks but the difference is quite small (about every sixth pair solved one more task). It can be explained that working in pairs leads to a greater willingness to work longer and less instances of giving up on solving tasks, thus it brings better results. This explanation can be supported by the finding that pairs spend slightly more time solving a given set of tasks than individuals, about 2% more.

Pairs seem to be more confident in answering, and thus to have greater self-efficacy. This is documented by the observed slightly higher proportion of pairs than individuals (every 12th one) that answer correctly on the first try and also the lower proportion of pairs who finally cancel their answer (every third one).

We noted that a higher proportion of pairs than individuals change their answer multiple times (every third one). This might indicate the use of the test environment to label the currently discussed solution when searching for an answer.

When focusing on programming tasks in which pupils build a program and iterate their solution based on feedback, pairs solve successfully these tasks in less time (about 10% less) and need fewer iterations (about one iteration less per every second solution). We interpret this to mean that pairs solve this type of task more effectively than individuals.

In terms of the components of CT, pairs perform slightly better in tasks involving abstraction, algorithmization, and generalisation (from 4% to 5%). No difference is found between pairs and individuals for the other observed components, i.e., decomposition and evaluation. This could be a subject of a further research.

It seems that the advantage of competing in pairs decreases with increasing age of pupils. This is demonstrated in the correctness of the answers, in the number of solved tasks and in the time spent on the test. It would mean that solving computer science tasks in pairs is most beneficial for the youngest pupils.

6.2. *Discussion of the Results*

Our results are largely consistent with those of other authors in pair programming research. We concur with Iskrenovic-Momcilovic (2019), Hannay *et al.* (2009), and Bellettini *et al.* (2022) who claim that pairs perform significantly better than individuals.

In the literature, there are different opinions regarding the speed of pair problem solving. Some authors report that pairs solve tasks more slowly (Bodaker and Rosenberg-Kima, 2022). Others argue that pairs solve tasks faster (Salleh *et al.*, 2011). In our research, we observed a trend of slower task solving in pairs compared to individuals as age increases.

Regarding successful solutions of tasks, speed depends on the presence of feedback in the task. In programming tasks, pairs work faster than individuals. We believe this is because programming tasks provide immediate feedback after any answer submission. Thus, contestants do not need long discussions about its correctness. On the contrary, pairs work as fast as individuals in tasks without feedback. The absence of feedback probably leads to more challenging pair discussion, which also explains the high rate of volatile answers for pairs in RQ3c. This is consistent with other studies, which claim that challenging pair discussions are present in complex programming tasks with insufficient feedback of the programming environment (Hannay *et al.*, 2009; Arisholm *et al.*, 2007).

Pairs are more likely to change the already marked answer while solving the task. However, after submitting the final answer and moving on to other tasks, they do not return to the task as often. Because pairs solve more tasks, it cannot be assumed that they are more under time pressure than individuals. It is likely that pairs are more confident in their answers or agree less often on the need to change their answer. This may be because they have greater self-efficacy and a more complex decision-making process than individuals. Wei *et al.* (2021) reached similar results in the area of pair programming. Another reason for this might be that individuals can easily use the possibility of the testing environment to move across tasks. Pairs must first agree on such a decision before moving on to another task.

Our results show that solving computer science tasks in pairs is most beneficial for the youngest pupils. A similar conclusion was reached by Iskrenovic-Momcilovic (2019) who notes that programming in pairs produces better results for beginners in comparison to programming as an individual student.

In the data, there are significant differences between Grade 6 and younger pupils. For pairs, there is an increase in the time required to solve the test and an increase in the proportion of volatile answers. At the same time, both 5th and 6th graders had an identical test settings. A possible reason for this may be the transition of pupils from primary school to lower secondary school, which happens two months before the contest. During this transition, the structure of each class usually changes, as some pupils move to another school. The pupil's relationship with the teacher also changes – in primary school, one teacher teaches all subjects; in secondary school, many teachers teach different subjects. According to Israel *et al.* (2016); Lai and Wong (2021), the new environment, classmates and classroom climate can have a negative impact on the ability to collaborate, and therefore on the quality of the collaborative pedagogical process. However, we could not test this hypothesis with data from older pupils. Thus, we are unable to determine whether the increase in the test-solving time of pairs compared to individuals is a smooth age trend or a one-time jump caused by the transition to a differently organised educational level.

6.3. *Limitations of the Research*

One limitation of the research is that the data only comes from the categories for grades 3 to 6, as older pupils cannot compete in pairs in the Bebras Challenge in Czechia. According to our investigation, this is due to the low interest of teachers who were mostly against pair contests for older pupils.

Another limitation is the use of data collected while solving several different tests (one each from the years 2021, 2022, and two from 2023). The tasks in these tests may vary in difficulty and in suitability for teamwork. This may have a negative impact on the quality of the results when looking at trends across years.

There is a limitation regarding the detection what impact the changes in contestants' opinion had on the final correctness of the answer (RQ3c). It lies in relatively low number of contestants in some of the examined answer scenarios (e.g., no risk). Although tens of thousands of contestants were included, in some calculations there were only a few hundred individuals and only tens of pairs.

We are also aware that only less than 10% of pupils competed in pairs.

There might be a limitation in the research related to the components of CT (RQ1b). It may lie in the inaccuracies in determining the presence of components of CT in used tasks.

The contest system does not allow to record the timestamp of switching between tasks; it only records the time of answering. Consequently, the time it took contestants to solve the task when they switched between tasks and "only" thought on another one without giving any new answer could not be reliably determined.

Another limitation may be the effect of pupils' transition to lower secondary school, as already discussed.

If one of the members of a pair is dominant, it can be expected that this pair will behave close to individual contestant. We have no data about it so result of the research can be limited by this.

6.4. *Possible Follow-up Research*

Further research in this area could look at how older pupils solve tasks in pairs. For a deeper insight into the issue, qualitative research based on observations of pupils solving tasks in pairs would be useful. This would make it possible to determine how pupils interact when solving computer science tasks and also clarify uncertainties about the causes of volatility among pairs.

International comparisons involving countries with different prevalence of collaborative learning could be useful. Another possibility is to conduct longitudinal research in a country where pair learning is being introduced. This could shed light on the extent to which more frequent teamwork of pupils increases its effectiveness. Another option is to consider how to modify Bebras tasks so that their assignments or types of answers are more suitable for group work, especially for older pupils.

6.5. *Contribution of the Research*

Our study extends the area of knowledge about cooperative work from programming tasks to other computer science topics. It shows that when solving short tasks, the trend of decreasing advantage of teamwork with increasing age of pupils is manifested. Teamwork is an essential skill for later employment and should be developed at schools. Hence we should consider whether current Bebras tasks are an appropriate tool for collaborative learning among older pupils.

Bebras tasks are a well-designed didactic source of computer science tasks. Thus it is suggested they are used not only in contests, but also in the school curriculum. The question is which teaching method to use to exploit their potential for developing computational thinking. Instead of solving the test independently, a two-phase discussion method may be appropriate. Pupils would first discuss the solution of a selected task in groups. After that, with the teacher's moderation, they would discuss the solutions proposed by different teams within the whole class. In this way, pupils would develop their ability to debate computer science problems, reflect on the reasoning of others, and think critically.

References

Arisholm, E., Gallis, H., Dyba, T., Sjoberg, D.I.K. (2007). Evaluating Pair Programming with Respect to System Complexity and Programmer Expertise. *IEEE Transactions on Software Engineering,* 33(2), 65–86. DOI: 10.1109/TSE.2007.17

Bartošová, I., Maněnová, M., Třečková, E. (2012). The New Comenia Script to Schools. *Procedia – Social and Behavioral Sciences*, 69, 2228–2236. DOI: 10.1016/j.sbspro.2012.12.192

Bebras.org (2024). *Bebras Task Workshops*. https://www.bebras.org/workshops.html

Bellettini, C., Lonati, V., Monga, M., Morpurgo, A. (2022). How is Two Better Than One? An Observational Study on the Impact of Working in Pairs When Solving Bebras Tasks. In: Bollin, A., Futschek, G. (Eds.), *Informatics in Schools. A Step Beyond Digital Education. ISSEP 2022. Lecture Notes in Computer Science, vol 13488*. Springer, Cham, 54–65. DOI: 10.1007/978-3-031-15851-3_5

Berki, J., Drábková, J. (2020). *Základy informatiky pro 1. stupeň ZŠ* (Basis of informatics for primary school). Textbook. Technical University of Liberec, Liberec, Czechia. `https://imysleni.cz/ucebnice/zaklady-informatiky-pro-1-stupen-zs`

Bodaker, L., Rosenberg-Kima, R.B. (2022). Online pair-programming: elementary school children learning scratch together online. *Journal of Research on Technology in Education,* 55(5), 799–816. DOI: 10.1080/15391523.2022.2036653

Dagienė V., Futschek G., Stupurienė G. (2019). Creativity in solving short tasks for learning computational thinking. *Constructivist Foundations* 14(3), 382–396. `https://constructivist.info/14/3/382`

Datzko-Thut, S. (2024). Team participation 2023. Report from Czechia, France, Germany, and Switzerland. In: *International Bebras Task Workshop 2024.* Vrnjačka Banja, Serbia.

Datzko-Thut, S., Pohl, W. (2024). Teamvorteil im Informatik-Biber? In: Thomas, M., Weigend, M. (Eds.), *Informatik in Teams. 11 Münsteraner Workshop zur Schulinformatik*. University of Münster, Münster, Germany, 93–100.

de Jesus, Â.M., Silveira, I.F. (2022). A Collaborative Learning Framework for Computational Thinking Development through Game Programming. *Informatics in Education,* 21(2), 253–281. DOI: 10.15388/infedu.2022.14

Filipi, Z., Mainz, D., Fadrhonc, J. (2020). *Práce s daty* (Work with data). Textbook. University of West Bohemia, Plzeň, Czechia. `https://imysleni.cz/ucebnice/prace-s-daty-pro-5-az-7-tridu-zakladni-skoly`

Hannay, J.E., Dybå, T., Arisholm, E., Sjoberg, D.I.K. (2009). The effectiveness of pair programming: A meta-analysis. *Information and Software Technology*. 51(7), 1110–1122. DOI: 10.1016/j.infsof.2009.02.001

Hejný, M. (2012). Exploring the Cognitive Dimension of Teaching Mathematics through Scheme-oriented Approach to Education. *Orbis Scholae*, 6(2), 41–55. `https://karolinum.cz/data/clanek/5036/OS_2_2012_final.41-55.pdf`

Hopcan, S., Polat, E., Albayrak, E. (2022). Collaborative Behavior Patterns of Students in Programming Instruction. *Journal of Educational Computing Research,* 60(4), 1035–1062. DOI: 10.1177/07356331211062260

Huang, J., Parker, M.C. (2023). Developing computational thinking collaboratively: the nexus of computational practices within small groups. *Computer Science Education*, 33(3), 342–374. DOI: 10.1080/08993408.2022.2039488

Iskrenovic-Momcilovic, O. (2019). Pair programming with scratch. *Education and Information Technologies*, 24, 2943–2952. DOI: 10.1007/s10639-019-09905-3

Israel, M., Wherfel, Q.M., Shehab, S., Ramos, E.A., Metzger, A., Reese, G.C. (2016). Assessing collaborative computing: development of the Collaborative-Computing Observation Instrument (C-COI). *Computer Science Education*, 26(2–3), 208–233. DOI: 10.1080/08993408.2016.1231784

Kalaš, I., Blaho, A., Moravčík, M. (2018). Exploring Control in Early Computing Education. In: Pozdniakov, S., Dagienė, V. (Eds.), *Informatics in Schools. Fundamentals of Computer Science and Software Engineering. ISSEP 2018. Lecture Notes in Computer Science, vol 11169*. Springer, Cham, 3–16. DOI: 10.1007/978-3-030-02750-6_1

Keith, K.P., Sullivan, F.R., Pham, D. (2019). Roles, Collaboration, and the Development of Computational Thinking in a Robotics Learning Environment. In: Kong, S.-C., Abelson, H. (Eds.), *Computational Thinking Education*. Springer, Singapore, 223–245. DOI: 10.1007/978-981-13-6528-7_13

Laal, M., Ghodsi, S.M. (2012). Benefits of collaborative learning. *Procedia – Social and Behavioral Sciences*, 31, 486–490. DOI: 10.1016/j.sbspro.2011.12.091

Lai, X., Wong, G.K.-W. (2021). Collaborative versus individual problem solving in computational thinking through programming: A meta-analysis. *British Journal of Educational Technology*, 53(1), 150–170. DOI:

10.1111/bjet.13157

Li, W., Liu, C.-Y., Tseng, J.C.R. (2023). Effects of the interaction between metacognition teaching and students' learning achievement on students' computational thinking, critical thinking, and metacognition in collaborative programming learning. *Education and Information Technologies*, 28(10), 12919–12943. DOI: 10.1007/s10639-023-11671-2

Ministry of Education, Youth and Sports of the Czech Republic (2008). *Memorandum of Understanding M MT – Intel*. Ministry of Education, Youth and Sports of the Czech Republic, Praha. https://msmt.gov.cz/ict/memorandum-msmt-intel

Ministry of Education, Youth and Sports of the Czech Republic (2021). R*ámcový vzdělávací program pro základní vzdělávání* (Framework Education Programme for Elementary Education). Ministry of Education, Youth and Sports of the Czech Republic, Praha. https://www.edu.cz/wp-content/uploads/2021/07/RVP-ZV-2021.pdf

Mrkvička, T., Roskovec, T., Rost, M. (2021). A Nonparametric Graphical Tests of Significance in Functional GLM. *Methodology and Computing in Applied Probability*, 23, 593–612. DOI: 10.1007/s11009-019-09756-y

Myllymäki, M., Mrkvička, T. (2024). GET: Global Envelopes in R. *Journal of Statistical Software*, 111(3), 1–40. DOI: 10.18637/jss.v111.i03

Power, R. (Ed.) (2020). *E-Learning Essentials 2020*. Power Learning Solutions, Sydney, NS, Canada. https://pressbooks.pub/elearning2020/

Procházka, J., Lapeš, J., Tocháček, D. (2020). *Robotika s LEGO WeDo pro 1. stupeň základní školy* (Robotics with LEGO WeDo for primary school). Charles University, Praha, Czechia. https://imysleni.cz/ucebnice/edukacni-robotika-s-lego-wedo-2-0-pro-1-stupen-zakladni-skoly

Rodríguez, F.J., Price, K.M., Boyer, K.E. (2017). Exploring the Pair Programming Process. In: *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. Association for Computing Machinery, New York, 507–512. DOI: 10.1145/3017680.3017748

Salleh, N., Mendes, E., Grundy, J. (2011). Empirical Studies of Pair Programming for CS/SE Teaching in Higher Education: A Systematic Literature Review. *IEEE Transactions on Software Engineering*, 37(4), 509–525. DOI: 10.1109/TSE.2010.59

Scherer, R., Siddiq, F., Sánchez Viveros, B. (2020). A meta-analysis of teaching and learning computer programming: Effective instructional approaches and conditions. *Computers in Human Behavior*, 109, 106349. DOI: 10.1016/j.chb.2020.106349

Schubert, S.E. (2001). The impact of modelling in informatics education on collaborative learning with school Intranets. In: Taylor, H., Hogenbirk, P. (Eds.), *Information and Communication Technologies in Education. IFIP — The International Federation for Information Processing, vol 58*. Springer, Boston, 247–258. DOI: 10.1007/978-0-387-35403-3_20

Selby, C., Woollard, J. (2014). *Refining an understanding of computational thinking*. University of Southampton, Southampton. https://eprints.soton.ac.uk/372410/

Šimandl, V., Dobiáš, V. (2024). Soutěžit sám nebo ve dvojici? Aneb jak si vedly dvojice žáků v soutěži Bobřík informatiky (Competing alone or in pairs? How pairs of pupils did in the Bebras Challenge). In: Drábková J. (Ed.), *Didinfo 2024: New Perspective in Informatics Education, International Proceedings on Teaching Informatics*. Liberec, Technical university of Liberec, 25–33. http://www.didinfo.net/images/DidInfo/files/Didinfo_2024_inter.pdf

Vaníček, J. (2012). *Informatika pro 1. stupeň ZŠ* (Informatics for primary schools). Textbook. Computer Press, Brno, Czechia.

Vaníček, J., Dobiáš, V., Šimandl, V. (2023). Understanding loops: What are the misconceptions of lower-secondary pupils? *Informatics in Education,* 22(3), 525–554. DOI: 10.15388/infedu.2023.20

Vaníček, J. (2019). Early Programming Education Based on Concept Building. *Constructivist foundation*, 14(3), 360 – 372. https://constructivist.info/14/3/360.vanicek

Vaníček, J., Řezníček, P. (2004). *Informatika pro základní školy, 1. díl* (Informatics for lower-secondary schools). Textbook. Computer Press, Brno, Czechia.

Vaníček, J., Šimandl, V., Dobiáš, V. (2022). Bebras Tasks Based on Assembling Programming Code. In: Bollin, A., Futschek, G. (Eds.), *Informatics in Schools. A Step Beyond Digital Education. ISSEP 2022. Lecture Notes in Computer Science, vol 13488*. Springer, Cham, 113–124. DOI: 10.1007/978-3-031-15851-3_10

Vaníček, J. (2024). Sestavy bobřích úloh jako pomůcka učitele pro výuku informatiky (Bebras tasks sets as a teacher's helper for computer science education). In: Drábková J. (Ed.), *Didinfo 2024: New Perspective in Informatics Education, International Proceedings on Teaching Informatics*. Liberec, Technical university of Liberec, 16–24. http://www.didinfo.net/images/DidInfo/files/Didinfo_2024_inter.pdf

Vygotsky, L.S. (1980). *Mind in Society: Development of Higher Psychological Processes*. Harvard University

Press, Cambridge.
Wei, X., Lin, L., Meng, N., Tan, W., Kong, S.-C., Kinshuk. (2021). The effectiveness of partial pair programming on elementary school students' Computational Thinking skills and self-efficacy. Computers & Education, 160, 104023. DOI: 10.1016/j.compedu.2020.104023
Wing, J. M. Computational thinking. Communications of the ACM [online]. 2006, 49(3) [cit. 2024-02-21]. DOI: 10.1145/1118178.1118215
Wu, B., Hu, Y., Ruis, A. R., Wang, M. (2019). Analysing computational thinking in collaborative programming: A quantitative ethnography approach. *Journal of Computer Assisted Learning*, 35(3), 421–434. DOI: 10.1111/jcal.12348

V. Šimandl is an assistant professor in the Department of Computing in the Faculty of Education at the University of South Bohemia in České Budějovice in the Czech Republic. His research activities focus on the area of Informatics education in lower secondary schools and programming. He has been organizing the Czech Bebras Challenge for 11 years.

J. Vaníček is an associate professor and the head of the Department of Computing in the Faculty of Education at the University of South Bohemia in České Budějovice in the Czech Republic. His area of interest is Informatics education in primary and lower secondary schools and early age programming. He is an author of 7 textbooks about information technology and programming. Between 2017 and 2020 he was a head of the strategic project PRIM developing new Czech national informatics curricula. He has been organizing Bebras Challenge for 16 years and he is a representative of the Czech Republic in the International Bebras Committee.

V. Dobiáš is an assistant professor in the Department of Computing in the Faculty of Education at the University of South Bohemia in České Budějovice in the Czech Republic. His research activities focus on the digital divide and computational thinking.