




# All Green: How Different Age Groups Solved the Same Bebras Task

Carlo Bellettini<sup>1</sup>, Violetta Lonati<sup>1,2</sup>, Mattia Monga<sup>1,2</sup>✉,  
and Anna Morpurgo<sup>1,2</sup>

<sup>1</sup> Università Degli Studi di Milano, Milan, Italy

<sup>2</sup> Laboratorio Nazionale CINI 'Informatica e Scuola', Rome, Italy  
{bellettini,lonati,monga,morpurgo}@di.unimi.it

**Abstract.** In this report we investigate how pupils tackle an interactive problem-solving task recently proposed in the Bebras Challenge. The task comes across as fun and simple and we proposed it to all pupils from fourth grade up. We analyzed the data collected, produced by 18,486 participating teams while interacting with the local Bebras platform. The quantitative data were supplemented by interviews conducted after the challenge, in which students were asked to solve the question while thinking aloud. Regardless of their age, all the participants found a good challenge in the task and they spent a lot of time interacting with it. Almost all teams attempted a naive approach to the solution, misled by the superficial characteristics of the problem, and many insisted on attempting the naive approach without ever abandoning it; this behaviour clearly decreases with increasing age.

**Keywords:** K12 · Bebras challenge · observational study · problem-solving

## 1 Introduction

The Bebras International Challenge on Informatics and Computational Thinking (<http://bebras.org>) is a yearly contest organized in several countries since 2004 [1, 3], with more than three million participants worldwide. The contest, open to pupils of all school levels (from primary up to upper secondary), is based on tasks rooted on core informatics concepts, yet independent of specific previous knowledge such as for instance that acquired during curricular activities.

Each Bebras country is free to choose and adapt the tasks to the local school context. In most countries the contest is run individually, in others it is team based. Many countries also propose *interactive* tasks, *i.e.*, tasks whose solution requires to interact with the contest platform. In addition to the submitted answers, our Bebras platform [4] is able to collect data concerning the interactions with the platform itself (how much time pupils spend on each specific task, whether and when they go back and review/change their answer to an already

completed task, whether they perform actions that generate feedback from the system, and so on [5,6]). This offered us the chance to conduct an observational study about how pupils of different age groups behave to find a solution. We focused on a task that we thought could be proposed to all pupils from fourth grade up. It is based on a geometric figure; the goal is to transform the figure, through a very simple graphic processing system, by applying a limited number of operations.

Here we present an analysis of the data collected during the Italian contest, produced by 18,486 participating teams while interacting with the Bebras platform. We studied which sequences of operations were performed by teams, in order to identify repeated patterns, successful and unproductive approaches, and changes in strategies. The quantitative data were supplemented by a think-aloud protocol conducted after the challenge, in which we asked five pairs of students to solve the task while thinking aloud and then interviewed them with follow-up questions.

The paper is organized as follows. Section 2 describes the task; in Sect. 3 we report our observations and analyses of the data collected; Sect. 4 summarizes our findings and draws some conclusions.

## 2 A Fun Task that Is Harder Than It Looks

### 2.1 The Task

This study focuses on one specific task (2022-UA-01a.Filling), authored by the Ukrainian Bebras team, and included in the Bebras Challenge in 2022 by ten countries, who translated and adapted it to the local contexts and platforms. In our country, where the contest involves teams of two or three people, the task was implemented as an interactive task, and it was proposed to all age groups, from IV to XIII grade. Figure 1 shows our version of the task, translated back into English for this paper’s readers. The colors were associated to patterns to help colorblind pupils, but our text forgot to explicitly mark which is the “green color” (the solid one): please note, however, that since our contest is team based, it should be considered rather unlikely that all the members of a team are colorblind.

The interactive version we proposed was designed to allow any number of trials; it has a counter of the moves and the solvers could reset to the initial state at will. This allows the teams to explore the system and figure out its semantics, which is described in the task only in very generic terms. In particular, the task’s text does not define what a *region* is: any maximal union of connected circles with the same colour (at the beginning, there are seven regions). Similarly, the task’s text does not state the fundamental property that solvers need to infer from trials: if one fills a region with the same color of one of its neighbours, then the filled region and its neighbours with the same color are merged into one single region and, from then on, their colors will always change simultaneously.

## 2022-UA-01a\_Filling



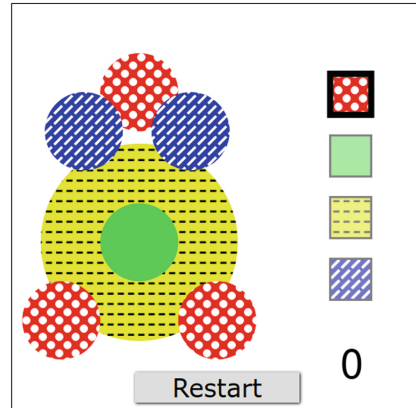
The program in the frame allows you to change the colors of the picture.

Select a color by clicking on one of the squares on the right, then click on a region of the picture: this will change the color of the region.

You can try as many times as you want; to restart just push the button "Restart".

You must color **all the picture with the green color.**

The number shows how many times you clicked. Try to reach the goal with the minimum number of clicks: it is possible to do that with just **3** clicks!



**Fig. 1.** The “All green” task; ‘green’ is the only solid color. (Color figure online)

## 2.2 The Task’s Solution

The figure can be turned into green by filling each circle with green separately; this requires six moves. By exploiting the possibility to merge circles and change their colors simultaneously, the number of moves can be reduced down to three. Here is the shortest sequence of moves leading to the desired result: fill with the blue color the largest circle (you now have five regions); fill the resulting blue region with red (you now have two regions); fill the resulting red region with green (you now have one single green region).

Such a sequence is unique, as can be proved as follows. The initial figure has four colors and three of them must be dropped. Since the number of allowed moves is three and no more than one color can be dropped in one move, then one color must disappear at each move. In the beginning, the only color that can be dropped in one single move is yellow, so one must click on the yellow circle to turn it into either red or blue. In the former case, both the red and the blue would appear in two separate regions, so neither of them could be removed in the next move. Thus, one must turn the yellow circle into blue. The next move is forced, since filling the resulting blue region with red is the only move that can reduce the number of colours, and the last one is then obvious.

The above proof is quite straightforward, but it needs the solver to understand the goal as “remove all colors except green from the picture”. This, however, encompasses a totally different mental *representation* of the problem, which the original wording of the task does not evoke.

### 2.3 How Hard Is the Task?

Bebras authors marked the task as one of medium difficulty for the age group IV-V grade, but we felt it could be proposed to all the age groups: the task comes across as fun and simple at all ages, but could be challenging enough even for grown-ups. As reported in Table 1, the task in the version used in our country turned out to be more difficult than estimated, especially for the youngest pupils. Less than 10% of the primary school teams were able to solve the task; the success rate increases with the age, regularly and markedly, but it remains at the level of challenging tasks even for the higher school grades.

**Table 1.** Overall results

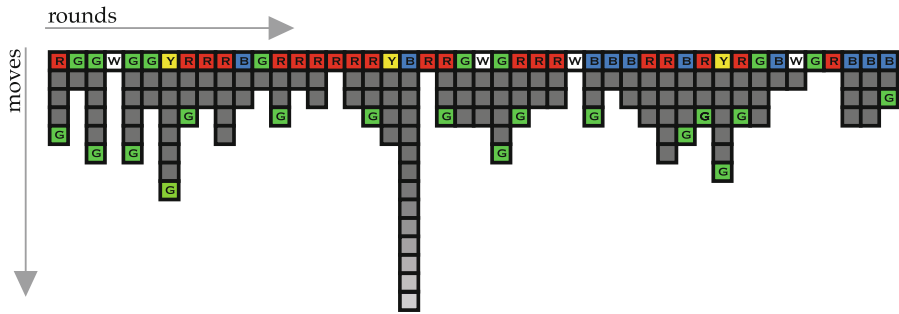
| grade   | teams  | success % |
|---------|--------|-----------|
| IV–V    | 3,008  | 8.9       |
| VI–VII  | 8,001  | 25.0      |
| VIII    | 3,648  | 35.5      |
| IX–X    | 2,074  | 50.3      |
| XI–XIII | 1,755  | 64.8      |
| Total   | 18,486 | 31.1      |

## 3 Observations

The data collected by the Bebras platform concern the interactions of teams with the platform itself, and in particular which actions they performed and when they clicked on the “Restart” button to recover the original figure and reset the counter for the number of moves. Before presenting our findings, let us introduce some preliminary definitions.

A *round* is a sequence of moves delimited by two Restart moves (obtained by one of the following actions: click on the Restart button, or open the task for the first time, or leave the task for the last time). The *length of a round* is the number of its moves. When the figure is all green, we say that it is in the *all-green configuration*; from this configuration on, the color of the figure can change but will remain uniform. A *green round* is a round that ends in the all-green configuration; there is only one *successful* green round of length 3. An *ineffective move* does not change the figure; ineffective moves are obtained whenever one clicks on a region after having selected the same color as the region (they are counted by the system as any other move). A *green-move* (resp., blue-move, or red-move) is a move that turns some region into green (resp., blue, or red).

The data collected during the context allow us to reconstruct the behaviour of each team. For instance, Fig. 2 illustrates the behaviour of an VIII grade



**Fig. 2.** Diagram summarizing the behavior of a (successful) team (VIII grade).

team who succeeded in solving the task. Each square represents a move, and each column represents a round, the leftmost being the first round carried out. The top square in each column is colored with the color applied in the first move (white, if the move is ineffective), the bottom square in each column is colored in green if the column represents a green-round. This particular team pressed the Restart button 45 times, hence producing 46 rounds. There are 16 green rounds, and the last one has length three. Four rounds start with an ineffective move, two among which followed by a restart, hence producing “empty” rounds.

All this team’s rounds are short, except for the 20<sup>th</sup> (the one fading towards the bottom of the figure). In general, rounds that either are longer than six moves or contain more than one ineffective move, can be considered as exploratory or with no intentional purpose, since i) six is the number of moves required to make the figure all green proceeding circle-by-circle and ii) an ineffective move may occur by mistake, but more than that is not reasonable in a serious attempt, considering the limit of three moves overall. Hence, we will call *trials* only the rounds that are six-moves long at most and contain one ineffective move at most.

### 3.1 Engagement

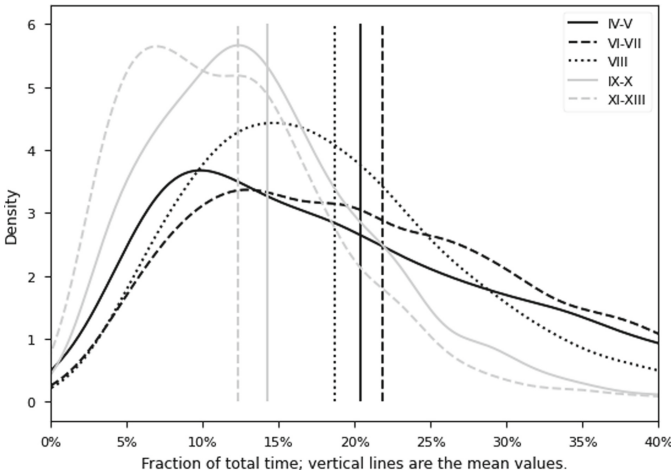
Teams spent a lot of time on the “All green” task. The overall time allowed in the context was 45 min per 12 tasks (or, for the youngest, 10 tasks), that is less than 4 min for each task on average. As reported in Table 2, teams spent, on average, 7 min and a half on the task (about 19% of the whole time) with peaks occurring on the central age levels. The pattern for the teams who solved the task, however, is slightly different: the younger successful solvers spent *more* time than average on the task, while the older spent less time. The time spent varies a lot among teams, even within a single category, as shown by Fig. 3, where we use kernel density estimation (instead of the usual histograms) for readability and visual comparison.

The active and positive engagement of participants was clearly visible also in the think-aloud protocol: participants spent several minutes on the task, and the frustration that occasionally emerged from the difficulty of finding the correct

**Table 2.** Cumulative time spent on the task (teams could leave the task and go back to it later, as many times as desired)

| grade   | time spent on the task | fraction of total time | avg n. of rounds | avg n. of trials | time spent on the task  | fraction of total time | avg n. of rounds | avg n. of trials |
|---------|------------------------|------------------------|------------------|------------------|-------------------------|------------------------|------------------|------------------|
|         | <i>all teams</i>       |                        |                  |                  | <i>successful teams</i> |                        |                  |                  |
| IV-V    | 443 s (7.4 m)          | 20.3%                  | 18.8             | 25.3             | 574 s (9.6 m)           | 25.5%                  | 16.3             | 22.0             |
| VI-VII  | 504 s (8.4 m)          | 21.7%                  | 25.0             | 26.9             | 534 s (8.9 m)           | 23.2%                  | 21.5             | 23.1             |
| VIII    | 456 s (7.6 m)          | 18.6%                  | 24.7             | 23.8             | 438 s (7.3 m)           | 18.2%                  | 21.2             | 20.7             |
| IX-X    | 357 s (6.0 m)          | 14.4%                  | 20.0             | 18.6             | 333 s (5.5 m)           | 13.4%                  | 17.7             | 16.4             |
| XI-XIII | 310 s (5.2 m)          | 12.4%                  | 17.0             | 15.0             | 278 s (4.6 m)           | 11.3%                  | 15.3             | 13.4             |
| Average | 450 s (7.5 m)          | 19.2%                  | 22.6             | 22.3             | 427 s (7.1 m)           | 18.0%                  | 19.6             | 19.4             |

solution generally did not extinguish the motivation to keep trying. Only one group decided to give up, settling for a 4-move solution.



**Fig. 3.** Fraction of total time spent on the task by all teams (kernel density estimation plots).

Another measure of engagement is provided by the number of rounds and trials carried out by teams: the averages by category are also reported in Table 2. In all age categories, the average number of trials and rounds is at least 15. The pattern is similar to the one for the time spent on the task: for the lowest two categories, the number of attempts increases when considering only the successful teams, while for the other ones it decreases. We interpret this as follows: for the younger ones, more attempts help to succeed and persistence seems useful; for the higher school levels, fewer attempts are needed to succeed; for the central grades, success does not seem immediately connected with persistence.

### 3.2 Graphs of Moves

To analyse the frequencies of moves in rounds and trials, we consider the graph of all possible configurations reached by one team. Such graph is defined as follows: the nodes of the graph are all possible configurations of the figure (*i.e.*, all possible ways the figure can be coloured); the arcs represent the possibility to switch from one configuration to another with a single click. Both the nodes and the arcs are equipped with weights that denote respectively in how many rounds/trials of the team a configuration (node) was reached and how many times a certain move (arc) was performed.

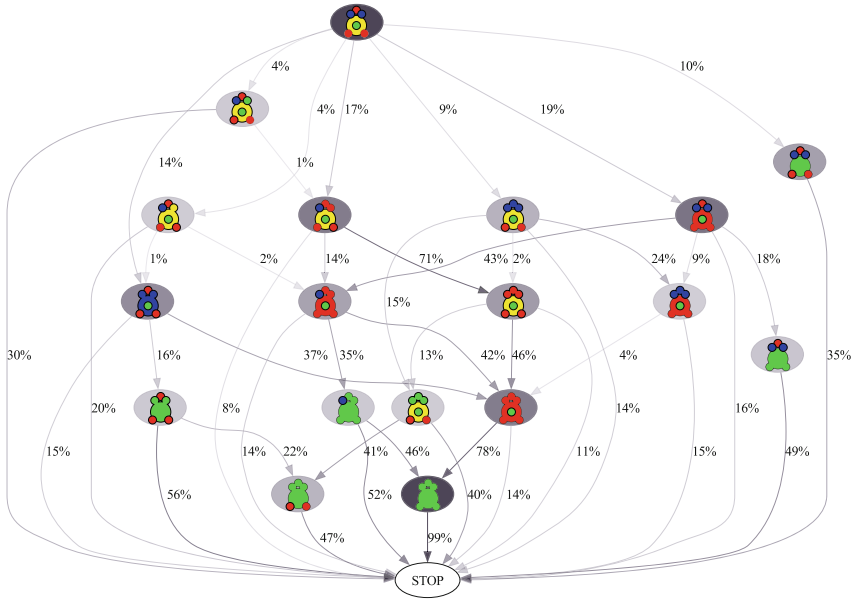
Table 3 reports the average number of nodes of such graphs, over the same age category. The number of reached configurations is lower when considering only trials, which also depends on the fact that fewer configurations are within reach with only 6 moves. The numbers in the table show a pattern similar to the one in Table 2.

**Table 3.** Average number of explored configurations, in rounds and in trials.

| grade   | conf. in rounds  | conf. in trials | conf. in rounds         | conf. in trials |
|---------|------------------|-----------------|-------------------------|-----------------|
|         | <i>all teams</i> |                 | <i>successful teams</i> |                 |
| IV–V    | 31.05            | 25.91           | 36.95                   | 32.45           |
| VI–VII  | 37.54            | 31.99           | 38.53                   | 33.34           |
| VIII    | 37.29            | 32.45           | 35.26                   | 30.87           |
| IX–X    | 32.01            | 28.21           | 29.31                   | 26.14           |
| XI–XIII | 27.32            | 24.82           | 24.38                   | 22.15           |

The graph of moves helps also identify which configurations occur the most and which are the most probable sequences of moves. For instance, Fig. 4 shows the graph of moves for all VIII grade teams. The weights of configurations and transitions are represented by the nodes' background and arcs' darkness. The figure includes only the nodes representing the most recurring configurations, namely those whose frequency is more than 4% of the most frequent configuration (which is the starting one, obtained after each restart). Moreover, to avoid dispersion of data, we merged the symmetrical nodes, that is, those that represent symmetrical pictures (*e.g.*, we basically consider equivalent filling with blue the bottom-left or the bottom-right circle). One can see that, from the initial configuration, there are four arcs towards configurations where one of the little circles is changed (the four arcs weigh 18% all together), and three arcs towards configurations where the big central circle is changed into blue, red, or green respectively (from left to right). They collect respectively 14%, 19% and 10% of all the choices.

During the think-aloud protocol we noticed in particular that many different configurations were obtained with green-moves by the youngest pupils who did



**Fig. 4.** Graph of the configurations most frequently reached by teams of VIII grades.eps

not find the solution, as they tried in many different ways to turn the figure into green by only using green-moves. Such configurations were not visited by the older students who found the solution, as they realized sooner that the green-moves were not fruitful, given the limit of three moves. The frequency of such configurations is indeed low also in the graph of Fig. 4.

### 3.3 Strategies

To analyze the strategies adopted by teams, we looked at the evolution of trials, by focusing on the first move of each trial. In all categories, one can see that a large proportion of the first few trials starts with a green-move; initial red-moves prevail in the central rounds, initial blue-moves occur more rarely and increase towards the final rounds when considering only successful teams.

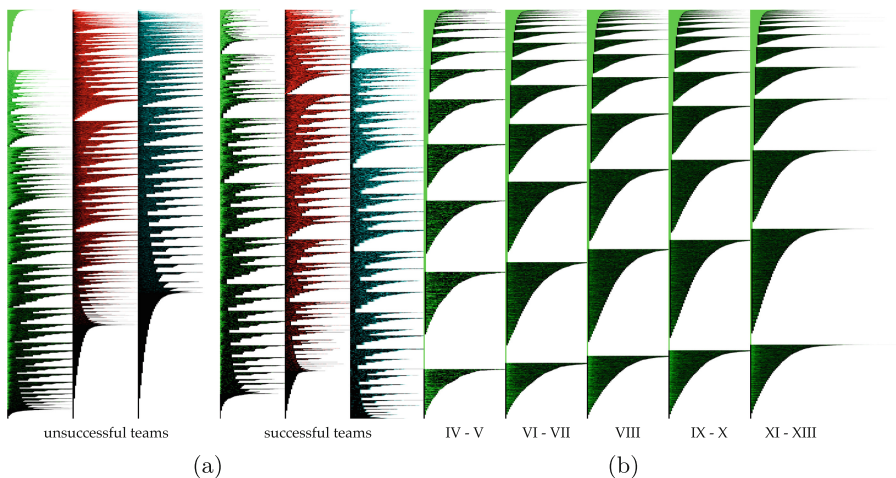
This is consistent with the strategy that we observed in the think-aloud protocol. Except for one group who solved the task very quickly, the other two groups who found the solution reasoned as follows: at first they tried to turn the whole figure into green by making green one circle at a time, hence the first few trials started with a green-move. After a while they realized that, with the above strategy, too many moves are needed and consequently switched to using different initial colours than green. Red is the most frequent colour in the original figure and this led them to try turning the whole figure into red instead of green. However this was not the successful strategy; indeed, after some trials,



they understood that the desired result could not be achieved if starting with a red-move, and concluded that they should start with a blue-move. Yet, this was not enough to find immediately the successful 3-long sequence of moves, which required some further attempts, with some revival attempts starting again with a red-move.

This approach can be seen also in the data collected by the platform. Figure 5(a) illustrates it by contrasting the occurrences of initial green/red/blue moves for VIII grade teams. The left and right portions of the figure were constructed using data exclusively from the unsuccessful and successful teams, respectively. Each team is depicted as a row. Rows are sorted from top to bottom according to the ratio of initial green-moves (resp., red/blue-moves) over the number of trials. At the top of the leftmost diagram, one can notice the relevant portion of teams who start with a green-move in most of their trials. At the top of the rightmost diagram, one can notice a small portion of teams who start with a blue-move in most of their (few) trials.

Figure 5(b) contrasts the behaviour of teams among the school levels. To simplify the figure, we only show whether the initial moves of the rounds are green or not. As in the Fig. 5(a), each row represents a team, and the teams are sorted according to the longest prefix of trials that start with a green-move. The five diagrams are scaled so that they have the same height, even if they represent populations of different sizes. This allows the reader to perceive how the percentiles change: moving towards right, there is an increase in the number of teams who abandon the green-strategy early.



**Fig. 5.** (a) Colors of initial moves, with unsuccessful/successful teams on left/right, respectively. (b) Trials starting with green-moves, grouped by category. (Color figure online)

### 3.4 Statistical Relation Between Success and Overall Performance

In order to study the statistical relationship between the probability of success in the task and the overall performances in the challenge, we considered the score each team gained in all the other tasks (*i.e.*, the “All green” task excluded). The scores were standardized to make them comparable among different categories: each score  $s$  was mapped to a standardized score  $\frac{s - \text{mean}_s}{\text{stdev}_s}$ , therefore having a mean score map to a standardized 0.0 in all the categories, and standardized scores ranges from  $-2.96$  to  $3.22$ . Then we fitted a generalized linear model to measure the effect of score on the probability of success, stratified by category. The model is the following, where  $\beta_K$  and  $\beta_S$  are two vectors of five parameters to be fitted (one for each category), which respectively measure the effect of being a team in a specific category and having performed with a standardized score  $S$  (see [2] for further details on this approach):

$$\begin{aligned} \beta_K &\sim \text{Normal}(0, 0.5) \\ \beta_S &\sim \text{Normal}(0, 0.5) \\ p_i &= \text{logit}^{-1}(\beta_K + \beta_S \cdot S) \\ y_i &\sim \text{Bernoulli}(p_i) \end{aligned} \tag{1}$$

As shown in Table 4, the teams who performed better overall (*i.e.*, those with standardized score 1.0), had a higher probability of getting the “All green” task right than those who have an overall average performance (*i.e.*, those with standardized score 0.0). The increment, however, is smaller for primary school teams.

**Table 4.** Parameters estimated by the Generalized Linear Model (1) (mean values of Monte-Carlo Markov Chains simulations)

| grade   | mean $\beta_K$ | mean $\beta_S$ | probability of success with standardized score 0.0 | probability of success with standardized score 1.0 | increment for better performers |
|---------|----------------|----------------|--|--|---------------------------------|
| IV-V    | -2.5           | 0.7            | 7.7%   | 14.6%  | 6.8%                            |
| VI-VII  | -1.6           | 0.5            | 24.0%  | 39.1%  | 15.1%                           |
| VIII    | -0.6           | 0.5            | 34.9%  | 52.2%  | 17.3%                           |
| IX-X    | 0.0            | 0.4            | 50.5%  | 67.5%  | 17.0%                           |
| XI-XIII | 0.7            | 0.6            | 65.9%  | 79.7%  | 13.8%                           |

## 4 Conclusions

In this report we described the observations we did on how different age groups tackled the same interactive problem-solving Bebras task. Regardless of their age, all the participants found a good challenge in the task, as shown by the

time spent interacting with it: overall almost  $\frac{1}{5}$  of the contest time was spent on the task, leaving only  $\frac{4}{5}$  for the remaining 9–11 tasks. Our analyses show that the ability to solve the task increases with age (as expected), regularly and markedly; for primary school pupils (in fact the age group targeted by the original authors of the task) the task turned out to be very difficult, and all the data suggest that this age group had the greatest difficulties in planning a winning strategy overcoming a failing naive approach. Except for the small minority of teams who found the solution quickly, most teams (even considering only those who succeeded in solving the task) carried out many trials. Initially, almost all teams attempted a naive approach, misled by the superficial characteristics of the problem, and many insisted on attempting the naive approach without ever abandoning it; this behaviour clearly decreases with increasing age. Moreover, while for younger kids a successful solution is associated with a higher number of visited configurations, for older ones the pattern is reversed: at some point in the process there is possibly an *eureka* moment in which they grasp a new (more productive) way of representing the problem in their mind. In this paper we mainly described the results derived from the analysis of quantitative data, with some further insight obtained by some interviews with subjects requested to try to solve the “All green” task while thinking aloud. Overall, we believe this kind of study is important to improve the design of Bebras tasks and a general understanding of interactive problem-solving.

## References

1. Dagienė, V.: Sustaining informatics education by contests. In: Hromkovič, J., Královic, R., Vahrenhold, J. (eds.) ISSEP 2010. LNCS, vol. 5941, pp. 1–12. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-11376-5\\_1](https://doi.org/10.1007/978-3-642-11376-5_1)
2. Gelman, A., Hill, J., Vehtari, A.: Regression and other stories. Cambridge University Press, Cambridge (2020)
3. Haberman, B., Cohen, A., Dagienė, V.: The beaver contest: attracting youngsters to study computing. In: Proceedings of ITiCSE 2011, pp. 378–378. ACM, Darmstadt, Germany (2011)
4. Bellettini, C., et al.: A platform for the Italian Bebras. In: Proceedings of the 10th International Conference on Computer Supported Education (CSEDU 2018), vol. 1, pp. 350–357. SCITEPRESS, March 2018. <https://doi.org/10.5220/0006775103500357>
5. Bellettini, C., Lonati, V., Monga, M., Morpurgo, A.: How pupils solve online problems: an analytical view. In: Proceedings of the 11th International Conference on Computer Supported Education (CSEDU 2019), vol. 2, pp. 132–139. SCITEPRESS, May 2019. <https://doi.org/10.5220/0007765801320139>
6. Bellettini, C., Lonati, V., Monga, M., Morpurgo, A.: Behind the shoulders of bebras teams: analyzing how they interact with the platform to solve tasks. In: Lane, H.C., Zvacek, S., Uhomoihi, J. (eds.) CSEDU 2019. CCIS, vol. 1220, pp. 191–210. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-58459-7\\_10](https://doi.org/10.1007/978-3-030-58459-7_10)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

