Aspects of Designing a Successful Bebras Challenge

Christian Datzko[1][0000−0003−1175−314X] and Susanne Datzko[2][0000−0002−8060−0596]

[1] University of Applied Sciences and Arts Northwestern Switzerland – School of Education, Brugg-Windisch, Switzerland, christian@datzko.ch
[2] Swiss Federal Institute of Technology, Zürich, Switzerland, susanne@datzko.ch

Abstract. The Swiss Bebras Challenge in its 11th iteration prompted for a more thorough analysis of the tasks as well as the task set than in the years before where this was done more informally. In addition to the challenge itself an after-challenge survey was run for the participants as well as the coordinating teachers. Combined with the performance data of the participants various aspects were analyzed some of which are presented in this paper. This is also compared to the performance of the 2019 challenge.
The performance and participation of the students was analyzed with regard to gender. In addition two specific aspects of tasks were explored: the role of the method in which the answer must be given, and the role of the length of the text a student has to read in order to solve a task. Also the setup of the challenges in its different versions for the different target ages and the anticipated difficulties of the individual tasks therein were evaluated.

Keywords: Bebras · Swiss Bebras Challenge · Gender · Interactive Bebras Tasks · Text Length of Bebras Tasks · Task Set Generation

1 Introduction

The Bebras International Challenge of Informatics and Computational Thinking (short: Bebras) is an "international initiative aiming to promote Informatics (Computer Science, or Computing) and computational thinking among school students of all ages" [3]. Since its start in 2004 [1] it has grown to reach more than 54 countries and more than 2.98 million students annually by mid-2021 [2].

After a pilot run in 2009 using the German challenge, Switzerland started to offer the challenge in 2010 in French, German and Italian. The annual participation in Switzerland has grown to almost 31 000 participants annually in 2020 [12].

The challenge takes place online. Teachers enroll their students and let them participate in class [13]. The challenge itself consists of 15 tasks for each age group except for 12 tasks for the age group 10–12, and 9 tasks for the age group 8–10. Of these ⅓ are easy tasks, ⅓ are of medium difficulty and ⅓ are hard to

solve. The tasks are independent of each other and graded individually [9, p. 4] and [10, p. iii].[3]

Although the Swiss Bebras Challenge is a rather typical Bebras challenge some aspects are notably different:

- As described by Datzko [8] the tasks are first adapted from the international community into German in cooperation with Germany, Austria, and Hungary (D-A-CH-HU) and then translated into French and Italian. Therefore the same task in German, French and Italian should be very similar.
- In 2020 an after-challenge survey was run among all participants and all teachers that signed up students ("coordinators").
- The federal and locally oriented system in Switzerland makes it particularly hard to reach out to teachers.
- Until 2019 the Swiss Bebras Challenge was run mostly by volunteers; in 2020 this system was replaced by a paid management and a steering committee of people from different universities.

In this paper the following hypothesis were checked:

- The gender of the participants has no influence on their performance in the Swiss Bebras Challenge 2020.
- Interactive tasks are more attractive to the participants.
- The longer a task the more difficult it is.
- The task set generation process created a comparable task set than in the year before.

The sources and collection of data are explained in appendix A.

2 Gender

The distribution of the students by gender holds no surprises: since the teachers sign up their students and in most cases complete classes or learning groups, the distribution of students by gender participating in the challenge is more or less the same as the distribution of students by gender in general.

When preparing the tasks some general aspects are considered [14]:

- *"Is is necessary to gender the task?*
- *Can the female gender be used instead of the male gender?*
- *If a female gender is used, can it be less 'stereotypical' (shopping, jewelry rearrangement, etc.)?*
- *[. . .] can the gender balance be improved?"*

[3] In this document the term "age groups" with a rough age range is given instead of the school years to enhance the international understanding of the target age of the students. Teachers in Switzerland, however, register the students based on the school year that they're in so it could be that some students are older and some students are even younger than the age group stated.

The performance of the genders is generally quite evened out. Figure 1 shows the data for the five different age groups. For older age groups a small tendency towards a better performance of male students can be seen but the difference is so small this could as well be a random artifact.

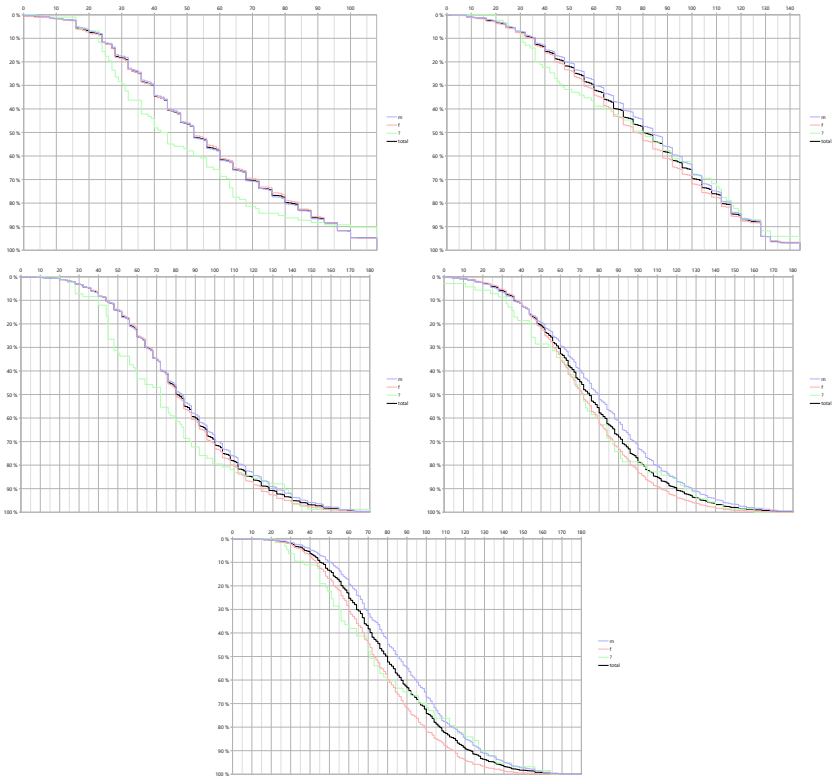

Fig. 1. Performance of students in the five different age groups (age group 8–10, 10–12, 12–14, 14–16, and 16–19) in terms of total points achieved (x-axis) as well as the accumulated relative number of students (y-axis). The total number of students with unknown gender is very low, so their graphs have to be handled with care.

In order to minimize data and to ensure a fully anonymous after-challenge survey the gender was not asked for. Therefore no statement about gender differences for these aspects can be made.

3 Answer Types

The challenge system offers different answer types. The classical multiple choice answer type (and its variant where the choices are images instead of text) have

been over time complemented with what the Bebras community calls "interactive tasks". This interactivity, however, could simply be a clickable SVG image; constructive interactivity where for instance parts of images have to be dragged onto their places is just a subset of these. The complete list of answer types, most of which are used in all task sets in the Swiss Bebras Challenge 2020, is:

- Multiple choice with text.
- Multiple choice with images.
- Multiple choice with text in a drop-down menu.
- Click-on-object interactive.
- Drag-and-drop interactive.

An analysis of the ranks that the students created for each task (see figure 2) showed that interactive tasks are much more popular than multiple choice tasks (68.9% vs. 35.7%). Even more, multiple choice tasks with drop-down menus were viewed worse than other multiple choice tasks (21.9% only). There is no big difference between multiple choice with text to select and multiple choice with images to select. Among the interactive tasks those with drag-and-drop interactivity were viewed even better (83.1%).

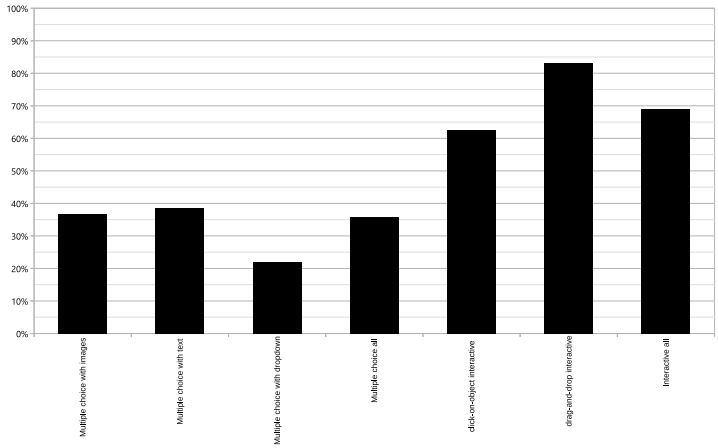

Fig. 2. Popularity of tasks depending on their answer type.

4 Text Length

In table 1 the average lengths of the tasks within each age group are given. While for the 2019 challenge the length of each task increased with the age groups, for 2020 the average task length is about the same for every age group. However, it

has to be noted that the age group 8–10 only needs to solve 9 tasks, and the age group 10–12 needs to solve 12 tasks while all other age groups need to solve 15 tasks. So the average time spent reading all tasks of that age group estimated after Hunziker [11][4] is almost the same for all age groups except for the oldest age group in 2019 while in 2020 the time spent reading decreases. For comparison: the total time spent available for solving the challenge is 40 minutes.

	Year	8–10	10–12	12–14	14–16	16–19
Average Words per Task	2019	89.9	97.3	109.3	111.7	133.9
	2020	104.0	109.0	109.5	111.3	108.3
Average Time Spent Reading	2019	10.8 min.	9.3 min.	10.3 min.	9.3 min.	12.2 min.
	2020	12.5 min.	10.5 min.	10.3 min.	9.3 min.	9.8 min.

Table 1. Average Number of Words per Task in German

This shows that the students spend about ¼ to ⅓ of the time of the challenge only for reading. In addition to that graphs have to be read, and interactive panes have to be understood leaving not much time in order for the task itself to be solved.

The numbers of words for each task have been correlated with the estimated difficulties, with the actual difficulties, and with the popularity of the task. The correlations are oriented so that while the number of words is increasing the difficulties are decreasing and the popularity is also decreasing. The results are shown in table 2.

Number of Words vs.	Year	8–10	10–12	12–14	14–16	16–19
Rated Difficulty	2019	0.70	0.13	-0.01	0.47	0.25
	2020	-0.04	-0.22	0.04	0.19	0.43
Percent Correct	2019	0.69	0.22	-0.06	0.24	0.05
	2020	0.50	0.18	0.44	0.52	0.51
Popularity	2019	*No data available.*				
	2020	0.35	0.15	-0.63	0.41	0.32

Table 2. Correlations between the number of words in each task of an age group vs. the rated difficulty, the rate at which the task was solved correctly, and the popularity of the task

For both years it can be seen that there is a notable linear dependency between the number of words in a task and its rate at which the task was solved correctly. One can say that tasks with more words are more difficult than tasks

[4] Basically since the reading speed over time increases with age the time spend for reading the same amount of words decreases with age.

with fewer words. Therefore it really is worth the extra effort to try to formulate tasks as short as possible while still maintaining a sufficiently complete definition of the task. However, for the rating of the tasks the number of words didn't seem to have much of an influence except for the youngest and the oldest age category in 2019.

Good examples for this area of tension are the tasks 2020-SK-04 "Where can it fly" and 2020-CH-01c "Formal Languages".

The task 2020-SK-04 was had an unclear formulation. About 40% of the students understood that they should determine where bees could fly *within* 30 minutes while about 15% of the students understood that they should determine where bees could fly *in exactly* 30 minutes. This made the task rather unpopular (rank 8 out of 12 for age group 10–12 and rank 8 out of 15 for age group 12–14). In the end both interpretations of the text were counted as correct.

The task 2020-CH-01c uses about 20% more words than the average task. In addition it offers two explanatory and one illustrative graphics, and the four alternatives itself are four graphics, also. Still the task was perceived as unclear. Quite a few students stated explicitly in the survey that they simply guessed the solution. Each possible answer was selected by between 14% and 33% of the students which supports this claim.

39.2% of the students participating in the after-challenge survey thought that the tasks were too much to read, while only 28.6% of the coordinators participating in the after-challenge survey thought the same thing. So it could be interpreted that the tasks were not generally too long but rather on the long end of the appropriate spectrum.

The correlation with the popularity holds a surprise: while the correlation is positive for all age groups except for age group 12–14, that particular age group has a very high negative correlation. Likely at least for that age group and those 15 tasks other aspects played a much greater influence on how popular the task was. Otherwise the correlations are lower than the perceived difficulty of the tasks making the length (and reading time) not so much a factor for liking a task. For 2019 no data is available because no survey was conducted.

5 Task Sets

The general process of generating the task sets is described by Datzko [8]. In 2019 a teacher for upper primary, lower secondary and upper secondary school was in charge of finalizing and compiling the Swiss task set. Also in 2019 the results of the meeting with the D-A-CH-HU were used as a base for selecting the tasks. In 2020 a university professor was in charge of finalizing and compiling the Swiss task sets based on the selection by a committee in Switzerland with the results of the meeting with the D-A-CH-HU as an additional input for that committee.

Switzerland does not test the tasks before the actual challenge like most other Bebras countries as well. Van der Vegt [15], [16] for instance found out that it is very difficult to predict the difficulty of a task within a given task set as well as

overall. This usually leads to quite large bands of percentages of correct solutions for each difficulty in each age group.

The rank error is calculated by comparing the estimated rank r_e of each task with its final rank r_f within each age group. For this the a priori task difficulties are translated into numbers (easy = 1, medium = 2 and hard = 3). The tasks are ranked based on the percentage of correct answers (both with and without counting unanswered participations) and then cast into their respective ranks. The distance between the estimated rank and the perceived ranks is averaged using the root-mean-square deviation:

$$d = \frac{\sqrt{\sum_{i=1}^{n}(r_{e,i} - r_{f,i})^2}}{n}$$

In addition to that the estimated rank r_f of all tasks in a task set is correlated with the percentage of correct answers for each task (both with and without counting unanswered participations).

The challenge system presents the tasks grouped by their difficulties, within which the tasks are presented in a random order. This to some extend favors the easier tasks so that students usually work on them first and therefore have a higher probability to attempt to solve them and potentially are less stressed out which increases the probability to solve the task correctly.

The results in table 3 show that there are some differences between the task sets from 2019 and from 2020. While for both years the task sets are not predicted perfectly as expected, the 2020 difficulty predictions for 2020 were a bit more off. A more thorough analysis presented in figures 3 and 4 show that the bands in which the tasks rated easy, medium, and hard performed differ more in 2020.

Age	Rank Error		Correlation	
Group	2019	2020	2019	2020
8–10	0.00	0.47	0.93	-0.47
10–12	0.35	0.26	-0.09	0.57
12–14	0.24	0.19	0.45	0.67
14–16	0.12	0.19	0.84	0.51
16–19	0.19	0.27	0.66	0.07

Table 3. Rank errors and correlations per age group 2019 and 2020

A maybe typical display of the deviation from the predicted to the actual difficulties of the tasks within a task set is the second-oldest age category. While the differences in the year 2019 were zero or one difficulty category, for 2020 one task (2020-CH-03a) was predicted as hard while it was actually easy. The figures 5 and 6 show the details.

For 2020 some of the issues with the task set can be found in the results of the general question of the survey shown in figure 7. Unclear formulation and text length are issues to attend (even more with increasing age). Also the difficulty

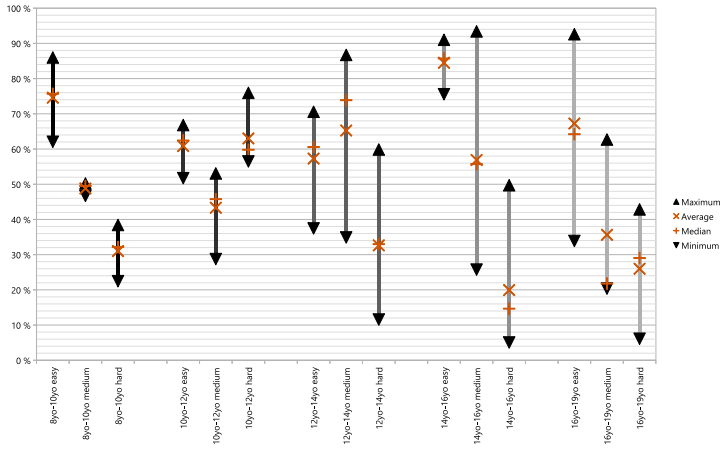

Fig. 3. Bands of percentage of correct answers for each age group 2019

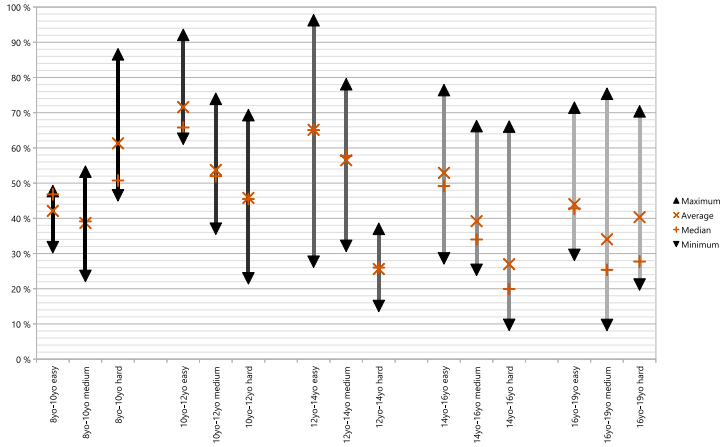

Fig. 4. Bands of percentage of correct answers for each age group 2020

of the task sets for the older students (almost no students received full credit) contributed.

6 Discussion, Conclusion and Further Considerations

The gender of the students has no influence on their performance. This is reassuring that the efforts taken in order to make the challenge interesting for all students are paying off. With the data collected no statement can be made

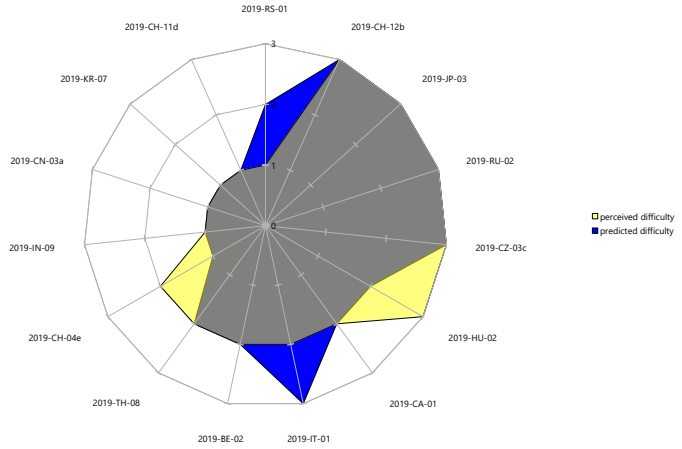

Fig. 5. Differences between the estimated and perceived difficulties of each task for the task set for the age group 14–16 from 2019 ordered by perceived difficulty.

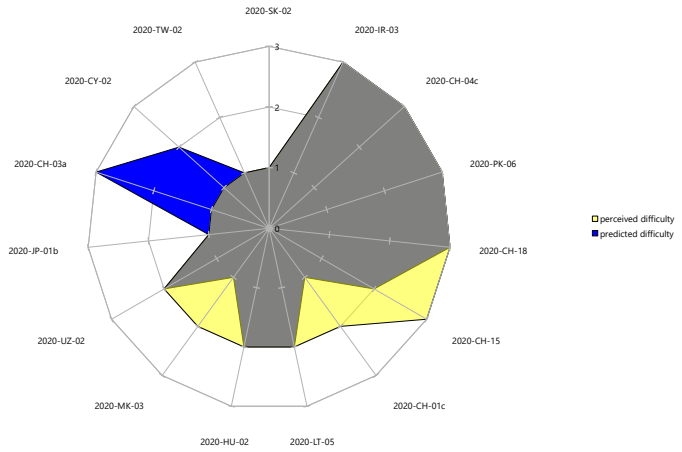

Fig. 6. Differences between the estimated and perceived difficulties of each task for the task set for the age group 14–16 from 2020 ordered by perceived difficulty.

whether the gender made a difference about which tasks were liked more or less. A future survey for instance in the form of interviews that maybe goes more in-depth rather than trying to achieve a broad participation could be made.

The length of the text of a task, however, makes a difference. Of course, for some tasks more text is necessary to explain the task a student has to solve. Reducing unnecessary text, however, is worth the effort. After all the Bebras

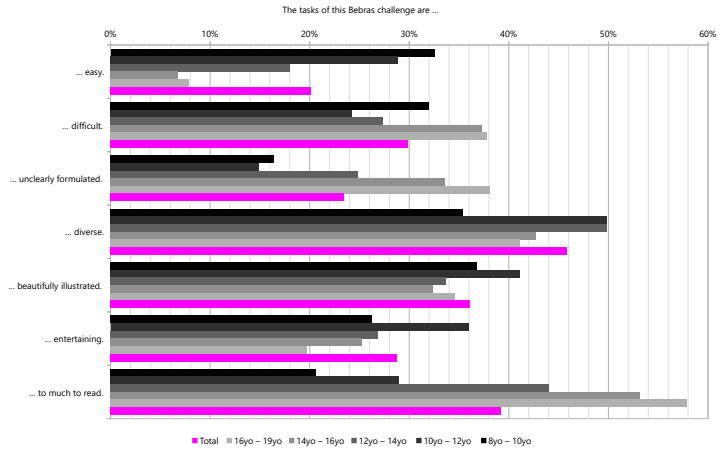

Fig. 7. Survey general question in age groups

Challenge is supposed to be about aspects of computational thinking and not about the ability to read and understand texts. It's not just all about the text length, though. A good compromise between text length and comprehensibility must be found. In some cases this could mean using a text that's a little bit longer.

Interactive tasks are much more popular. Especially tasks that are interactive in a constructive way are very much liked by students. Although they take a lot of effort to program this effort is definitely worth the gain in popularity of the challenge as a whole. However, making a task interactive just so that it becomes an interactive task is not the best idea. Again, the Bebras Challenge should focus on Computational Thinking instead of "just" being appealing.

The data for the task sets in 2019 and 2020 is probably not good enough to extract any general trend. In this case an analysis over more years should be made to have some general findings. What can be said, though, is that task sets are inherently difficult to create. This confirms the research of Van der Vegt [15], [16]. Without testing the tasks with students before the actual challenge it's probably not very likely that the situation will improve. So it could be questioned whether the traditional rating of "hard", "medium", and "easy" for tasks should be eliminated from the task development process and replaced with a different system. Also the system of ordering tasks by "difficulty" and scoring them accordingly should be questioned.

In general running a Bebras challenge is a year-long endeavor involving a lot of expertise in different areas and a lot of manpower, especially in a country like Switzerland where the challenge is offered in several languages. Judging from the feedback from the students and teachers, however, it's a good investment.

Acknowledgments

The authors would like to thank Eljkaim Schrijvers and his team from cuttle.org for their support. They would also like to thank Juraj Hromkovič and Regula Lacher for their encouragement to make a thorough analysis of the Swiss Bebras Challenge 2020 of which this paper is an excerpt. Also the Swiss Bebras team allowed the authors to use the data. For the survey Fabian Frei and Juraj Hromkovič gave some additional input. The authors would also like to thank the anonymous reviewers for their constructive feedback.

References

1. Bebras International Challenge of Informatics and Computational Thinking: History (2021), https://www.bebras.org/?q=history
2. Bebras International Challenge of Informatics and Computational Thinking: Statistics (2021), https://www.bebras.org/?q=statistics
3. Bebras International Challenge of Informatics and Computational Thinking: What is Bebras (2021), https://www.bebras.org/?q=about
4. Bundesamt für Statistik: Hauptsprachen nach Kantonen – Ständige Wohnbevölkerung ab 15 Jahren (2019), https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/sprachen-religionen/sprachen.assetdetail.11607238.html
5. Bundesamt für Statistik: Obligatorische Schule: Lernende nach Grossregion und Schulkanton (2020), https://www.bfs.admin.ch/bfs/de/home/statistiken/kataloge-datenbanken/tabellen.assetdetail.11787885.html
6. Bundesamt für Statistik: Sekundarstufe II, Berufliche Grundbildung: Lernende nach Grossregion, Schulkanton, Geschlecht, Staatsangehörigkeit und Charakter der Schule (2020), https://www.bfs.admin.ch/bfs/de/home/statistiken/kataloge-datenbanken/tabellen.assetdetail.11787905.html
7. Cuttle.org: Cuttle – science contests (2021), https://cuttle.org/
8. Datzko, C.: The genesis of a bebras task. In: Pozdniakov, S.N., Dagienė, V. (eds.) Informatics in Schools. New Ideas in School Informatics. pp. 240–255. Springer International Publishing, Cham (2019)
9. Datzko, C., Datzko, S., Hromkovič, J., Lacher, R. (eds.): Aufgaben und Lösungen 2019 – Alle Stufen. SVIA-SSIE-SSII Schweizerischer Verein für Informatik in der Ausbildung, Zürich (2019), https://www.informatik-biber.ch/documents/2019/Informatik-Biber-2019-Alle-Stufen-mitLoesungen.pdf
10. Datzko, S., Frei, F., Hromkovič, J., Lacher, R., Pellet, J.P. (eds.): Aufgaben und Lösungen 2020 – Alle Stufen. SVIA-SSIE-SSII Schweizerischer Verein für Informatik in der Ausbildung, Zürich (2020), https://www.informatik-biber.ch/documents/2020/Informatik-Biber-2020-Alle-Stufen-mitLoesungen.pdf
11. Hunziker, H.W.: Im Auge des Lesers – foveale und periphere Wahrnehmung: vom Buchstabieren zur Lesefreude. Transmedia Verlag, Zürich (2006)
12. Informatik-Biber Schweiz: Chronik (2021), http://www.informatik-biber.ch/de/geschichte/
13. Informatik-Biber Schweiz: Teilnehmen (2021), http://www.informatik-biber.ch/de/teilnehmen/
14. Vasiga, T.M.: Why HTML is useful and ODT is painful but salvageable for data extraction & unconscious gender imbalance in bebras tasks (2017), unpublished

15. van der Vegt, W.: Predicting the difficulty level of a Bebras task. In: Olympiads in Informatics. vol. 7, pp. 132–139 (2013), https://ioinformatics.org/journal/INFOL127.pdf
16. van der Vegt, W.: How hard will this task be? Developments in analyzing and predicting question difficulty in the Bebras challenge. In: Olympiads in Informatics. vol. 12, pp. 119–132 (2018), https://ioinformatics.org/journal/v12_2018_119_132.pdf

A Data Set

This paper uses data collected in the challenge system [7] as well as data collected in the after-challenge survey.

A.1 Data from the Challenge System

For the Swiss Bebras Challenge every participant solves 9 to 15 tasks as described above. For each of these tasks the given answer is stored. In addition, besides the name, the teacher may enter a gender of the participant or choose not to enter a gender. The analyzed data does not include coordinators or administrators (which might test run the challenge themselves) but it does include some test accounts they might have set up for themselves which would be too tedious to weed out for every user account. This has been done so since 2010 so the numbers are comparable to each other. For this paper the participants were anonymized and only more general data (like the canton of the school for which the participant was registered) was used.

Since the data is usually deleted during the year to re-purpose the challenge for testing purposes and to make place for user names for the next challenge, the data for 2019 is taken from an unpublished analysis done by Christian Datzko in 2019/2020.

The lengths of the tasks were taken from the German task version. The same tasks in French tend to be a little bit longer and the same tasks in Italian tend to be a little bit shorter, but in general the relative lengths of different tasks stay the same. Table 4 shows the data of the participation in 2019 and 2020 in the different languages and age groups.

In order to put the absolute numbers into context this was compared to the population statistics published by the Bundesamt für Statistik [4], [5], and [6].[5]

Figure 8 shows that the Bebras challenge is run all over Switzerland.

A.2 Survey

After the participation every student was invited to participate in a purpose-created after-challenge survey. The intent was to collect information about what students liked or disliked about the challenge and why. In there four questions were asked:

[5] For the years 2019 and 2020 where no data is available yet a constant growth rate is assumed to extrapolate the data for these years.

Language	Age Group	2019				2020			
		Male	Female	Unknown Gender	Total	Male	Female	Unknown Gender	Total
German	8–10	1 666	1 404	8	3 078	2 167	1 789	65	4 021
	10–12	3 024	2 686	69	5 779	3 686	3 278	28	6 992
	12–14	3 074	3 079	7	6 160	3 357	3 397	25	6 779
	14–16	2 675	3 163	5	5 843	3 865	4 687	29	8 581
	16–19	433	381	2	816	467	269	27	763
	Total	10 872	10 713	91	21 676	13 542	13 420	174	27 136
French	8–10	397	412	0	809	171	159	20	350
	10–12	268	247	8	523	203	194	36	433
	12–14	165	138	3	306	210	149	50	409
	14–16	471	522	4	997	439	540	30	1 009
	16–19	274	241	2	517	555	694	33	1 282
	Total	1 575	1 560	17	3 152	1 578	1 736	169	3 483
Italian	8–10	0	1	0	1	25	27	0	52
	10–12	108	96	0	204	73	80	8	161
	12–14	91	80	0	171	39	44	0	83
	14–16	36	38	0	74	25	9	0	34
	16–19	54	12	1	67	34	11	0	45
	Total	289	227	1	517	196	171	8	375
Total	8–10	2 063	1 817	8	3 888	2 363	1 975	85	4 423
	10–12	3 400	3 029	77	6 506	3 962	3 552	72	7 586
	12–14	3 330	3 297	10	6 637	3 606	3 590	75	7 271
	14–16	3 182	3 723	9	6 914	4 329	5 236	59	9 624
	16–19	761	634	5	1 400	1 056	974	60	2 090
	Total	12 736	12 500	109	25 345	15 316	15 327	351	30 994

Table 4. Absolute participation in Switzerland 2019 and 2020.

- A general question about the Bebras challenge ("The tasks of this Bebras challenge are easy / difficult / unclear formulated / diverse / beautifully illustrated / entertaining / too much to read.").
- A short description of every task of the task set to choose the most popular tasks.
- A question what a popular task defines ("These are my favorite tasks because the subject is exciting / they are challenging to solve / they are quick and easy to solve / I like the pictures / the problem is interesting / I could understand them quickly / they are interactive.").
- An open text field for additional comments.

The survey was not linked to individual participants but rather run anonymously on a different server. But the survey was offered by the challenge server on the exit page of the challenge. This made a high participation rate of 42.0% possible. However, as expected the return rates vary. Table 5 shows the detailed information.

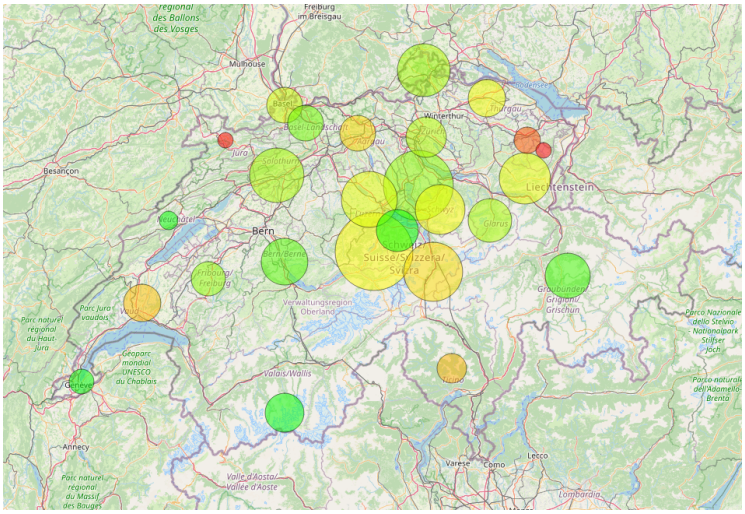

Fig. 8. Participants per student population of each canton 2020 on a map of Switzerland; the size of the circle indicates the relative participation and the color indicates a relative increase from −100% (red) via 0% (yellow) to +100% or more (green). The map is created with an OpenStreetMap in the background by using an adapted script originally written by Wilfried Baumann.

Since just 58 coordinators filled in the after-challenge survey, just some general notions can be read from this.

Language	Age Group	Participants		Return Rate
		Survey	Challenge	
German	8–10	1 542	4 021	38.4%
	10–12	3 325	6 992	47.6%
	12–14	3 934	6 779	58.0%
	14–16	2 230	8 581	26.0%
	16–19	185	763	24.3%
	Total	11 216	27 136	41.3%
French	8–10	129	350	36.9%
	10–12	198	433	45.7%
	12–14	315	409	77.0%
	14–16	414	1 009	41.0%
	16–19	488	1282	38.7%
	Total	1 544	3 483	44.3%
Italian	8–10	20	52	38.5%
	10–12	140	161	87.0%
	12–14	66	83	79.5%
	14–16	30	34	88.2%
	16–19	26	45	57.8%
	Total	282	375	75.2%
Total	8–10	1 691	4 423	38.2%
	10–12	3 663	7 586	48.3%
	12–14	4 315	7 271	59.4%
	14–16	2 674	9 624	27.9%
	16–19	699	2 090	33.4%
	Total	13 042	30 994	42.0%

Table 5. Survey participants 2020.