Students’ Performance on Programming-Related Tasks in an Informatics Contest in Finland, Sweden and Lithuania

Valentina Dagiene  
Vilnius University  
Vilnius, Lithuania  
valentina.dagiene@mi.vu.lt

Linda Mannila  
Abo Akademi University  
Turku, Finland  
linda.mannila@abo.fi

Timo Poranen  
University of Tampere  
Tampere, Finland  
timo.t.poranen@uta.fi

Lennart Rolandsson  
Royal Institute of Technology  
Stockholm, Sweden  
leanntr@kth.se

Pår Söderhjelm  
Lund University  
Lund, Sweden  
par.soderhjelm@teokem.lu.se

ABSTRACT

The ways in which informatics is covered in K-12 education vary among European countries. In Finland and Sweden, informatics is not included in the core curriculum, whereas, for example, in Lithuania, all students are exposed to some informatics concepts starting in the fifth grade. Bebras is an annually arranged international informatics contest for K-12 level, resulting in a large collection of data about contestants and their results. In this paper, we analyze contest data from the Finnish, Swedish and Lithuanian 2013 contests, focusing on students’ performance on tasks related to algorithmic thinking. Our findings suggest that despite coming from different educational systems, students perform rather similarly on the tasks. The same tasks are difficult and the thinking behind picking an incorrect answer seems rather similar throughout the countries. The analysis also points out that there is a lack of easy questions – this needs to be fixed in order to not risk scaring students away.

Categories and Subject Descriptors

K.3.2 [Computer and Information Science Education]: Computer Science Education

General Terms

Algorithms, Performance

Keywords

Informatics Education, Contests, Algorithmical Thinking, Programming, Data Structures

1. INTRODUCTION

The lack of programming and computer science (from now on referred to as informatics) in K-12 education is increasingly recognized as a serious lack in the national curriculum of many countries [7]. While many European countries included informatics in their curricula as early as in the 1970s, many of these efforts were dropped for different reasons; often the change was due to insufficient awareness of the importance of informatics and the misunderstanding that using IT as a tool is all that needs to be taught. As a result, many students graduate from secondary school with quite a lot of experience from using computers and software, but little, if any, knowledge of the underlying principles.

Both digital literacy and informatics are essential components of a modern education. In today’s world familiarity with programming and other informatics topics is as critical to all citizens as familiarity with traditional scientific disciplines was in the previous century. To be prepared for the jobs of the 21st century, students must not only be digitally literate but also understand key concepts of informatics. “All of today’s students will go on to live a life heavily influenced by computing, and many will work in fields that involve or are influenced by computing. They must begin to work with algorithmic problem solving and computational methods and tools in K-12” [2]. This is also addressed in a recent report by the joint Informatics Europe & ACM Europe Working Group on Informatics Education, where one recommendation states that: ‘all students should benefit from education in informatics as an independent scientific subject, studied both for its intrinsic intellectual and educational value and for its applications to other disciplines’ [6]. The emphasis on informatics as a science as opposed to merely the use of IT as a tool can also help to promote a more equal gender balance in the field [1].

Recently, the development in two European countries in particular has received much attention: In Estonia, the Tiger Leap Foundation formed by the Ministry of Education launched the ProgeTiger program in 2012, aiming at teaching children informatics starting in the first grade [10]. Similarly, in the UK, a new national curriculum will come into force in fall 2014, where ICT has been replaced by a computing program covering both primary and secondary school [11].

In this paper we analyse students’ results in the international Bebras informatics contest for three Baltic countries: Finland, Sweden and Lithuania, where informatics is not included as a subject in the curriculum. The aim is to analyze how students manage with problems requiring algorithmic thinking, both in general and based on age and country.
2. BEBRAS - AN INTERNATIONAL K-12 INFORMATICS CONTEST

2.1 The Contest in a Nutshell

Several programming contests and olympiads [5, 8] are arranged globally on a regular basis. In Finland, Sweden, and Lithuania, the same system is used: a national contest. A majority of the contest tasks have to be chosen answer and the age-group.

The Bebras contest addresses all primary, middle and upper secondary schools dividing pupils into five age groups: Mini (grades 3-4, age 8-10), Benjamin (grades 5-6, age 10-12), Cadet (grades 7-8, age 12-14), Junior (grades 9-10, age 14-16), and Senior (grades 11-12, 16-18). In Finland and Sweden the contest is introduced one year earlier, hence the age groups there correspond to grades 2-3 (Mini), 4-5 (Benjamin), 6-7 (Cadet), 8-9 (Junior) and 10-12 (Senior).

The main goals of the Bebras contest are to raise all students’ awareness of informatics and evoke interest in the field, as well as to motivate students to understand its fundamentals and become fluent with the technology, e.g. to be able to communicate with a machine. The contest should help children get interested in informatics and to stimulate thinking about contributions of informatics to science at the very early stage of their education. Since informatics is not a subject in its own right in many countries [7], this kind of contest might be one way, or even the only way, to introduce children to what informatics really is.

Finland organised its first national contest in 2010 and Sweden in 2012. The number of participants has increased in both countries from the beginning (1472 in Finland, 1625 in Sweden), and in 2013 there were 4434 participants from Finland and 1798 from Sweden. Participation numbers usually increase rapidly, sometimes even double on an annual basis, during the first years. Table 1 gives the participation statistics for different age groups in Finland, Sweden and Lithuania. Compared to Finland and Sweden, Lithuania has a very high participation rate: in 2013 over 25 000 students (0.9% of the population) participated in the contest.

As we can see, the participation rate for girls and boys is close to 50/50 for Mini, Benjamin and Cadet in all three countries. In Finland, the proportion of girls drops at Junior-level (69% boys, 31% girls), whereas in Lithuania the distribution stays more equal until Senior-level (68% boys, 32% girls). In Sweden, participation remains equal up to Senior-level, when there is a heavy drop in girl participation: not even one tenth of the participants are girls.

An additional detail worth noting is the decrease in number of participants for Seniors in Finland (from 1281 for Juniors to 170), whereas a similar drop cannot be seen for neither Lithuania or Sweden. One can hence expect Finnish Seniors to represent a more selected group of pupils than what is the case for the other countries.

2.2 Contest Organization

Each country has free hands in implementing the national contest as long as it is arranged during the official contest week (second week of November). Most countries use web-applications for the contest. In Finland, an initial version of an online contest system was developed in 2009 [9]. Later the Finnish system has been further developed by Slovenia and Sweden. Lithuania has also developed its own contest management system with students/teachers/tasks database and assessment functions. Both the Finnish and the Lithuanian system make it possible to practice on contest questions from previous years before taking part in the actual contest.

The number of questions (usually between 15-21) varies depending on the age group and country. Students have fixed time to answer all questions: 45 minutes in Finland/Sweden and 55 minutes in Lithuania. The contest systems store students’ name, gender, teacher’s name, school data and used time in addition to answers; hence, the Bebras contest results in a collection of anonymised datasets depicting detailed information on gender, the time used, the chosen answer and the age-group.

The scoring system can also vary between countries. In Finland, Sweden and Lithuania, the same system is used: a correct answer gives full points, an incorrect answer decreases points, and skipping a question leaves the score unchanged. The number of points awarded also depends on the difficulty level of the task at hand (although this differentiation was not used in the following analysis).

Teachers are informed using the local contest website (www.majava-kilpailu.fi, www.bebras.se, www.bebras.lt) and through direct contacts from the organizers via email or other channels. The aim for all three countries is for Bebras to become a nation-wide contest and to convey the message “this is informatics” to teachers, so that they get a picture of what informatics is and what kind of content as well as problems are involved.

Prior to the contest, teachers register their school in the contest system, and receive instructions as well as contest keys for their students. The contest is open daily during the contest week, and teachers decide the time for when their classes take part in the contest. All participation takes place under teacher supervision.

2.3 Development of Bebras Tasks

Contest tasks are prepared during an annual international task workshop. The workshop produces a task pool, from which each country is obliged to choose tasks for their national contest. A majority of the contest tasks have to be from that pool, but countries are allowed to use a few of their own tasks or tasks from previous years. The workshop decides which tasks should be mandatory, e.g. must be used in all countries to make it possible to produce comparable results and to emphasise important educational goals (usually 2-3 tasks per age-group).

Contest tasks are divided into six categories [4] and each task is given a difficulty level (easy, medium, hard) that varies when used in different categories. Tasks are either multiple choice (four-choice questions with one correct answer) or interactive (using drag-and-drop techniques, assembling constructions, picking items, writing, etc.).

2.4 Selection of Tasks for the 2013 Contest

Tasks for the 2013 contest were selected based on a process involving many researchers and teachers. In Finland and Sweden, the same tasks were used for all age groups. Lithuania...
Table 1: Number of participants in Finland, Sweden and Lithuania in 2013

<table>
<thead>
<tr>
<th></th>
<th>Finland (X% boys, Y% girls)</th>
<th>Sweden (X% boys, Y% girls)</th>
<th>Lithuania (X% boys, Y% girls)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mini</td>
<td>826 (52%, 48%)</td>
<td>262 (49%, 51%)</td>
<td>2 176 (55%, 45%)</td>
</tr>
<tr>
<td>Benjamin</td>
<td>852 (50%, 50%)</td>
<td>201 (56%, 44%)</td>
<td>7 022 (54%, 46%)</td>
</tr>
<tr>
<td>Cadet</td>
<td>1 294 (55%, 45%)</td>
<td>451 (55%, 45%)</td>
<td>6 550 (57%, 43%)</td>
</tr>
<tr>
<td>Junior</td>
<td>1 281 (69%, 31%)</td>
<td>413 (54%, 46%)</td>
<td>6 490 (60%, 40%)</td>
</tr>
<tr>
<td>Senior</td>
<td>170 (78%, 22%)</td>
<td>471 (91%, 9%)</td>
<td>3 671 (68%, 32%)</td>
</tr>
<tr>
<td>Total</td>
<td>4 423 (58%, 42%)</td>
<td>1 798 (63%, 37%)</td>
<td>25 909 (58%, 42%)</td>
</tr>
</tbody>
</table>

nia has used the same task set as Germany/Austria/Switzerland/Netherlands for several years and for the contest in 2013 they also added a set of tasks used by Finland/Sweden (Lithuania offers the largest number of tasks in their contests, 18 for Mini and 21 for all others categories, whereas in Finland/Sweden there are 10 for Mini and 15 for the others). As a result there is a large overlap between the tasks included in the contest in Finland/Sweden and Lithuania.

As algorithms and programming are key concepts in informatics, and usually also the aspects missing in today’s education, it is common practice for the contest to include a larger proportion of tasks related to algorithms and programming concepts than to e.g. tasks dealing with computer usage. This was also the case for the 2013 contest in Finland, Sweden and Lithuania. In Finland/Sweden the majority of tasks for all age groups were related to algorithmic thinking: 7 in Mini, 11 in Benjamin, 12 in Cadet, 12 in Junior, and 12 in Senior. Lithuania used all of these tasks and several additional tasks as well. All questions used are available on national contest websites and contest management systems.

In the following, we will provide some background information on what tasks were used in our analysis and the rationale for choosing these specific tasks.

3. ANALYSIS OF TASKS ON ALGORITHMS AND PROGRAMMING

3.1 Tasks of Specific Interest

All three countries used 27 tasks categorized as related to algorithmic thinking. Some tasks were used in two or three age groups. In our analysis we have focused on differences in performance between students based on age and country. After involved discussions on what tasks to analyse in order to cover as many algorithm/programming concepts as possible, we jointly selected a set of tasks of special interest. These tasks are listed in Table 2 together with brief descriptions. The data were anonymised by removing all personal and school related information prior to the analysis.

In order to give the reader an idea for how programming concepts are included in the Bebras contest, the complete wordings for two of the selected tasks are presented in Figure 1 and Figure 2 respectively.

In the next section, we will present the results from studying the performance on the selected tasks by students of different age groups in the three countries.

4. RESULTS

4.1 Country-Specific General Performance

The average scores over the tasks common to all three countries are shown in Figure 3 (the same qualitative differ-
Table 2: Selected tasks related to algorithms / programming concepts

<table>
<thead>
<tr>
<th>Task name</th>
<th>Age group</th>
<th>Mandatory</th>
<th>What students can learn from the task</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the forest</td>
<td>Mini, Benjamin</td>
<td></td>
<td>finding a path; graph; tracing; finding a solution backwards</td>
</tr>
<tr>
<td>Zebra Tunnel</td>
<td>Benjamin</td>
<td>yes</td>
<td>to follow instructions; algorithm analysis; data structures: FIFO (queue) and LIFO (stack)</td>
</tr>
<tr>
<td>The highest tree</td>
<td>Benjamin, Cadet</td>
<td></td>
<td>search algorithm; local optimisation; global optimum</td>
</tr>
<tr>
<td>Spinning toy</td>
<td>Benjamin, Cadet,</td>
<td>for all</td>
<td>binary tree representation; tree traversal; operations abstraction</td>
</tr>
<tr>
<td></td>
<td>Junior, Senior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signal fire</td>
<td>Cadet, Junior, Senior</td>
<td>yes for Cadets</td>
<td>graphs; shortest path problem; breadth-first search</td>
</tr>
<tr>
<td>Necklace</td>
<td>Junior</td>
<td>yes</td>
<td>shortest path</td>
</tr>
<tr>
<td>Dice</td>
<td>Junior</td>
<td>yes</td>
<td>following a list of commands; procedure; imperative programming</td>
</tr>
<tr>
<td>Different paths</td>
<td>Junior</td>
<td></td>
<td>graphs; dynamic programming</td>
</tr>
<tr>
<td>River inspection</td>
<td>Junior, Senior</td>
<td>yes for Seniors</td>
<td>flow problem; planar directed graph; maximal cut; sweeping line</td>
</tr>
<tr>
<td>Visiting Friends</td>
<td>Senior</td>
<td>yes</td>
<td>top-down analysis; modulo operations; patterns</td>
</tr>
<tr>
<td>Apples in the basket</td>
<td>Senior</td>
<td>yes</td>
<td>patterns, invariants</td>
</tr>
</tbody>
</table>

ence between Finland and Sweden was obtained when the average was taken over all tasks). Students in Sweden and Finland perform similarly, except in the Benjamin and Senior age groups, where the average score in Finland is ~40% and ~70% higher, respectively (measured over all tasks). The discrepancy among the Seniors can be explained by the different selection of participating students (cf. Section 2.1), whereas we see no obvious reason for the discrepancy among the Benjamins. One may speculate that the Finnish students’ greater reading ability could have an impact in this age group, as problem texts are longer than in Mini and students are not yet perfect readers.

The students in Lithuania have similar scores as Sweden in Mini and Benjamin, but lower scores than both Sweden and Finland in the Cadet and Junior age-groups (and Senior, but there the different selection makes the comparison less useful). This is somewhat surprising, given that the Lithuanian students are in average one year older. One should, however, remember that the participation rate is much higher in Lithuania than in Finland and Sweden. Hence, although we do not have statistic evidence, we conjecture that the lower scores for Cadets and Juniors can be explained by Swedish and Finnish being biased towards academically stronger schools.

The average score on each task compared among the three countries is shown as a scatter plot in Figure 4 for Sweden vs Finland (all tasks) and Sweden vs Lithuania (common tasks). As can be seen from the diagram, the contest covers a whole spectrum of difficulties, from tasks that almost everyone solved to tasks that had a negative average score, i.e. worse results than if the cohort had answered randomly. It is clear that, apart from a few exceptions, the same tasks are difficult in the three countries.

4.2 Performance on Same Task in Different Age Groups

As mentioned above, some tasks are included in several age groups to make it possible to investigate how student understanding and performance change throughout the educational system. One such task in the contest is “Spinning toy”, which – as shown in Table 2 – is included in Benjamin, Cadet, Junior and Senior. The average score on the task for boys and girls in the three countries are shown in the diagram in Figure 5.

As the diagram shows, the general trend is quite evident: performance improves with age. Whereas the results are overall quite similar for all countries, the diagram gives rise to some interesting questions: Why are Lithuanian girls performing substantially worse than the others starting at Cadet-level? Why is there no progression from Junior to Senior for Swedish boys? And are these findings merely related to the task at hand or do they indicate a more general trend for e.g. a certain type of tasks for given groups of students? These are examples of questions that merit further investigation and should be comprehensively discussed in our communities.
4.3 Dealing with incorrect answers

In addition to looking at how well students do at giving correct answers, it is also interesting to look deeper into tasks where the most common answer is not the correct one. We have chosen three tasks related to graphs for this kind of analysis.

In the task “Different paths”, the participants had to count the number of shortest paths between two points on a grid, in which some parts were inaccessible. As the diagram in Figure 6 indicates, all three incorrect answers were more common than the correct one. Students, however, seem to think rather alike in the three countries, as the proportions for the respective answers are quite similar.

In the task “Highest tree”, the participants were given a description of a local search algorithm for finding the highest tree in a forest (always go to the highest tree you currently see) and were asked which tree the algorithm would find in a given example, where 13 is the global maximum but 10 is the local maximum found by the algorithm. In Figure 7, we see the chosen answers for two age groups. As the diagram shows, the task was rather tricky for Benjamins, whereas the proportion of students choosing the correct answer increased in the Cadet age group.

The diagram in Figure 8 shows the chosen answers for “Signal fire” for three age groups. This task appears to have been quite difficult throughout all age groups, with 6 minutes being the most common answer (when the correct answer is 5 minutes). The only exception is Finnish Seniors, out of which more than 60% answered correctly. This can however be explained by the Finnish Seniors representing a
selected group of students, as discussed in Section 2.1. Students seem to struggle with abstract thinking: they should trace the graph mentally without real experimentation. Another difficulty can be to notice peculiarities of the graph and relying on what seems correct just by looking, whereas one should check by counting. The 8 minutes incorrect choice can be explained by students confusing the shortest path with the longest path.

5. DISCUSSION AND FUTURE WORK

An international contest like Bebras involves several countries, cultures and languages. Clearly, these are all factors that make it challenging to create unambiguous and clear tasks descriptions, which mean exactly the same in all languages and that are interpreted in precisely the same manner by all students.

In general, the task-wise scores are rather similar in the three countries (disregarding the Senior students of Finland). This is an interesting result, as it suggests that the average ability to solve a certain task is largely independent of the school system. However, Tomcsányi and Vaníček [12] noted that even small differences in translations, wording or pictures might affect the answers. Also, in some countries students are more willing to guess than in other countries; for example, in our comparison the students of Lithuania consistently gave fewer empty answers.

The results of analysing tasks where an incorrect answer is more common than the correct one also point out another similarity between students from different countries: not only is a certain task usually equally “difficult” in the three countries, but there is also a fairly constant fraction of students that thinks in the same way and therefore picks the same incorrect answer.

Our analysis has shown that, apart from the Mini category, there was a lack of easy tasks. This is problematic both for the contest itself (one risks scaring away students and teachers) and for the current analysis, because the results for the more difficult tasks contain more noise as a large fraction of the students might be merely guessing. Unfortunately, it is not easy to estimate how difficult a task will be for a particular age group when developing the task [13].

The experience of doing this investigation, combining data from three countries, has given rise to many ideas on how to further develop the way in which tasks are created and translated. For instance, the translation process could be made more transparent, so that changes made to the task description during translation in some country immediately would be accessible for all other countries as well. The large and multifaceted data collected in Bebras contests make it possible to analyze several interesting aspects related to e.g. students’ understanding, difficulties and misconceptions based on different factors, such as age, gender and geographical location. In this paper, we have looked into some of these questions for a limited set of data, but many questions still remain unanswered. Some of these have already been mentioned in the paper (e.g. in Section 4.2).

In addition, tasks where an incorrect answer is more common than the correct one merit further investigation – is the choice of an incorrect alternative associated with some commonly known misunderstandings? The gender aspect is also interesting - what would a detailed analysis of the performance of boys and girls tell us?

6. REFERENCES