

Improved Mobile Robot Programming Performance through Real-time Program Assessment

Rémy Siegfried
LSRO, EPFL
Lausanne, Switzerland
remy.siegfried@alumni.epfl.ch

Severin Klingler
Dept. of Computer Science, ETH
Zurich, Switzerland
kseverin@inf.ethz.ch

Markus Gross
Dept. of Computer Science, ETH
Zurich, Switzerland
grossm@inf.ethz.ch

Robert W. Sumner
Dept. of Computer Science, ETH
Zurich, Switzerland
robert.sumner@inf.ethz.ch

Francesco Mondada
LSRO, EPFL
Lausanne, Switzerland
francesco.mondada@epfl.ch

Stéphane Magnenat
Dept. of Computer Science, ETH
Zurich, Switzerland
stephane@magnenat.net

ABSTRACT

The strong interest children show for mobile robots makes these devices potentially powerful to teach programming. Moreover, the tangibility of physical objects and the sociability of interacting with them are added benefits. A key skill that novices in programming have to acquire is the ability to mentally trace program execution. However, because of their embodied and real-time nature, robots make the mental tracing of program execution difficult.

To address this difficulty, in this paper we propose an automatic program evaluation framework based on a robot simulator. We describe a real-time implementation providing feedback and gamified hints to students.

In a user study, we demonstrate that our hint system increases the percentage of students writing correct programs from 50% to 96%, and decreases the average time to write a correct program by 30%. However, we could not show any correlation between the use of the system and the performance of students on a questionnaire testing concept acquisition. This suggests that programming skills and concept understanding are different abilities.

Overall, the clear performance gain shows the value of our approach for programming education using robots.

KEYWORDS

robotics in education; Thymio; automatic program evaluation; gamified hint system; simulation

ACM Reference format:

Rémy Siegfried, Severin Klingler, Markus Gross, Robert W. Sumner, Francesco Mondada, and Stéphane Magnenat. 2017. Improved

Mobile Robot Programming Performance through Real-time Program Assessment. In *Proceedings of ITiCSE'17, July 03-05, 2017, Bologna, Italy.*, 6 pages.

DOI: <http://dx.doi.org/10.1145/3059009.3059044>

1 INTRODUCTION

The last decade has seen a large number of initiatives and products aiming at teaching programming and computer science to children using robots. They do so by providing a simplified programming environment, typically visual, inside which children define the behavior of the robot. The main motivations are the strong interest robots raise in children, the tangibility of programming physical objects, and the sociability of interacting with them. These elements are important for the healthy physical, intellectual and social development of children. Moreover, several studies have shown that robots can be effective to teach certain computer science and software engineering notions to beginners [1, 8]. In addition, with young children, robots can do so without a strong gender bias [7]. Furthermore, robots show how information processing can be embodied within the physical world.

However, it is not known whether robots are effective at teaching computer science and software engineering in general, especially compared to a traditional software programming course. The fact that robots allow children to quickly write programs that do something indicates neither deep learning nor the acquisition of transferable skills. Moreover, a key skill that novices in programming have to acquire is the ability to mentally trace program execution [13]. We argue that programming robots poses challenges, that combined, render tracing programs difficult:

- (1) **Not steppable.** Programs cannot be executed step by step, as robots are physical real-time systems.
- (2) **Not trivially inspectable.** As the program is not steppable, the internals of execution is not easily visible.
- (3) **Not deterministic.** As robots operate in the continuous real world, having sampled and noisy sensors and imperfect motors, the execution of a program might be hard to predict from the code itself.
- (4) **Bad source code locality.** As most programming paradigms for mobile robots are concurrent, modifying a part of a program might affect other parts.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ITiCSE'17, July 03-05, 2017, Bologna, Italy.

© 2017 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-4704-4/17/07...\$15.00.

DOI: <http://dx.doi.org/10.1145/3059009.3059044>

Therefore, as observed by previous work [8], both students and teachers face difficulties identifying bugs in their code, and progressing beyond trivial exercises is problematic.

Hence, to leverage the advantages of robots for programming education, one must overcome the aforementioned challenges. One way is to improve the programming tools to make the internals of program execution more visible. Previous work has, for example, explored the use of visual feedback and augmented reality [5]. A complementary way to technical tools is to train the tracing skills of the students, by guiding them through problems of increasing difficulty.

A common way of providing adaptive guidance is to employ an intelligent hint generation system. In the context of programming education, hint systems have been successful at providing hints for code correctness [12] and coding style [2], but have not yet been applied to robot programming.

A necessary brick to provide hints and feedback to students is the ability to assess a given code regarding a set of metrics [4]. While there are well-established methods for software programming [3], evaluating the quality of a robot program and providing meaningful feedback to the student are open research questions. In this paper, we address these questions. In particular, we provide the following contributions:

- (1) an automatic evaluation framework for robot programs based on running simulations in background,
- (2) a real-time implementation on a mobile device, providing feedback and gamified hints to the student,
- (3) a user study and its analysis, evaluating the effects of this system on learning.

2 ROBOT PROGRAM EVALUATION

Static evaluation is difficult for robot programs, because a slight change in a parameter might lead to a completely different behavior. An obvious way to dynamically evaluate a robot program is to run it on a real robot in a physical environment and check whether the robot performs its task. This is not practical for automatic program evaluation, as instrumentalizing such tests would require a lot of resources and result in slow results. Another way is to execute the program within a virtual robot in a simulated environment. That is significantly faster and can be fully automatized. A potential drawback is that the quality of the program evaluation depends on the accuracy of the simulation. However, typical educative programming tasks do not rely on advanced physical phenomena, so a simple simulator providing collision and kinematic support would suffice.

2.1 Simulation-based performance evaluation

Figure 1 shows the overall structure of the robot program evaluation system. It is similar to software-only approaches [10], but with all tests taking place within the simulator. However, robot programs are difficult to separate into functional units, so the breadth of a test varies in function of its simulation conditions. For example, imagine a task in which the student must program the robot to navigate a labyrinth. A very specific case, such as following a straight corridor, could be

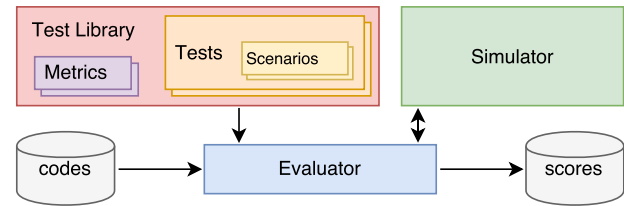


Figure 1: The automatic robot program evaluation system.

considered a unit test, and a complete task, such as navigating a given labyrinth, could be considered an end-to-end test. Hence, we use the term *test* broadly, to refer to any kind of simulation-based evaluation.

The robot program evaluation system is composed of a test library, an evaluator, and a simulator. The test library provides, for a given task, a set of tests. Each test consists of one or more *scenarios*, and a function to compose the *scenario scores* into a *test score*. A scenario is the unit of simulation, it contains

- a description of the environment (walls, obstacles),
- a script describing the actions to perform during the simulation, such as pressing a button on the robot or moving an object in the environment,
- the name of a metric to compute the scenario score.

The *metric* evaluates the performance of the robot in a given scenario; it is a function, over time, of the robot's sensor and motor values and its relation with its environment. The evaluator simulates each scenario and computes test scores using simulation results and the metrics from the library.

2.2 Didactic sequencing through gamified hints

Building on our program evaluation framework, we propose a gamified hint system. A key decision is what amount of information about the program quality to report to the student and when. If too little is provided, the system brings no benefit. If too much is provided, it would be detrimental to learning, as students could superficially solve the problem by following the hints rather than reasoning by themselves. Moreover, they could also experience cognitive overload.

For a given task, such as navigating a labyrinth, our system progressively displays hints indicating test results. A hint is an image representing one test, with its background color changing in function of the test outcome (see Figure 5). Hints can be in three states: locked, unlocked or active.

- Active hints display how well the current program performs at the corresponding tests. The background can take 3 different colors: green, orange and red corresponding to good, mediocre or bad test scores.
- Unlocked hints do not show the performance of the tests, to push students to think by themselves.
- Locked hints are invisible, and the tests are not run.

A task is split into a list of levels, each containing one or more hints, unlocked in order. The system maintains a current level. Hints for higher levels are locked, hints for the current level are unlocked, and hints for the lower levels are active.

Hints are updated each time the student runs the program on the real robot. A student passes to the next level when her or his program is good at all tests (their scores are above a threshold) up to and including the current level. Moreover, after a given time, the next level is automatically unlocked, regardless of the results of the tests.

3 IMPLEMENTATION

3.1 Thymio II, Aseba and VPL

We have implemented our robot program evaluation system on the ThymioII robot and its visual programming environment. The ThymioII robot [11] and its Aseba software were created at the EPFL, ETH Zürich, and ECAL. Both the hardware design and the software are open-source.

The robot is small (11 × 11 × 5 cm), self-contained and robust with two independently-driven wheels for differential drive. It has five proximity sensors on the front and two on the back, and two color intensity sensors on the bottom. There are five buttons on the top, a three-axis accelerometer, a microphone, an infrared sensor for receiving signals from a remote control and a thermometer. For output, there are RGB LEDs at the top and bottom of the robot, as well as mono-colored LEDs next to the sensors, and a sound synthesizer.

The Aseba programming environment [6] uses the construct `onevent` to create event handlers for the sensors. VPL is a component of Aseba for visual programming¹. This work branches from the development version of the tablet implementation, and runs on an Android tablet Nvidia Shield K1. Figure 2 shows a VPL program for following a line of black tape on a white floor. On the left is a column of *event blocks* (buttons, proximity, ground color intensity, accelerometer, microphone) and on the right is a column of *action blocks* (motors, colors of the robot). By dragging and dropping one event block and one or more action blocks to the center pane, an *event-actions pair* is created. Both event and action blocks are parametrized, enabling the user to create many programs from a small number of blocks.

The VPL tablet prototype is implemented in Qt Quick². This allows writing the test library in a comfortable declarative language, and the evaluator in JavaScript. The evaluation itself is performed using the Enki simulator³. This simulator is simple and fast yet accurate enough, and already supports the Thymio robot. Moreover, Aseba provides a library for embedding its virtual machine into Enki, allowing a seamless simulation of a valid Thymio program inside Enki. The software used in the experiment is available on github⁴.

3.2 Gamified hints for labyrinth navigation

To evaluate our system, we have implemented hints for a task in which students must program a robot to navigate inside a labyrinth. The task consists in programming the robot such that it traverses as much of the labyrinth as possible in 10

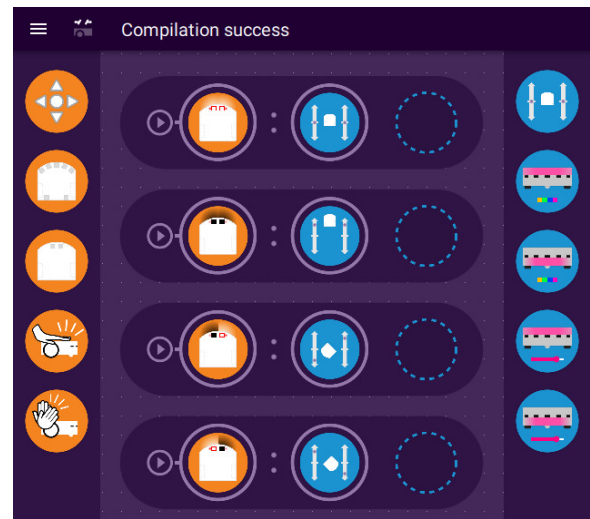


Figure 2: The Thymio VPL tablet environment.

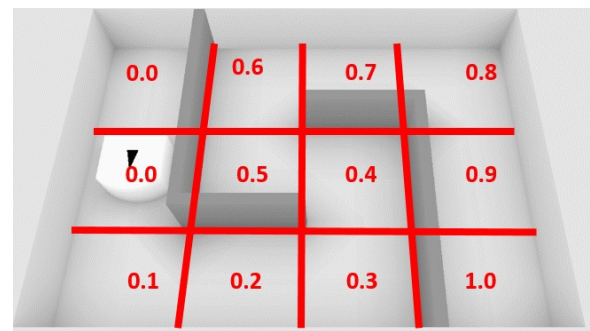


Figure 3: Labyrinth split into tiles of increasing score.

seconds. Figure 4f shows a picture of the labyrinth, which consists of a tortuous corridor with a width of 20 cm.

To compute a score of a scenario, the labyrinth is split into small tiles as shown in Figure 3. The score is given by the furthest reached tile, with the distance of the robot to the next tile taken into account to provide sub-tile precision. To test the program in different environments of increasing complexity, we create different tests and scenarios, starting with simplified versions of the labyrinth. Each scenario is simulated for a certain duration. If the robot has a high score, it indicates that the program is correct for that scenario. Only the score is available to the user through hints, the simulation itself is hidden. There are 8 scenarios, organized in 5 tests/hints, and unlockables in 4 different levels:



Hint 1: empty area, level 1 see Figure 4a

- **scenario 1:** going straight forward in an empty arena, for 5 seconds.

As there is no tile in an empty area, this test computes the score using the traveled distance.

¹<https://www.thymio.org/en:visualprogramming>.

²<https://www.qt.io/qt-quick/>

³<https://github.com/enki-community/enki>

⁴<https://github.com/aseba-community/thymio-vpl2/releases/tag/iticse2017>



Hint 2: corridor, level 2 see Figure 4b

- **scenario 2.1:** following a corridor, starting aligned to it, for 5 seconds.
- **scenario 2.2:** following a corridor, starting unaligned to it, for 5 seconds.



Hint 3: left turn, level 3 see Figure 4c

- **scenario 3.1:** passing a left turn, starting oriented to the left, for 7 seconds.
- **scenario 3.2:** passing a left turn, starting oriented to the right, for 7 seconds.



Hint 4: right turn, level 3 see Figure 4d

- **scenario 4.1:** passing a right turn, starting oriented to the right, for 7 seconds.
- **scenario 4.2:** passing a right turn, starting oriented to the left, for 7 seconds.



Hint 5: labyrinth, level 4 see Figure 4e

- **scenario 5:** navigating in the complete labyrinth, for 10 seconds.

Some hints have more than one scenario, to capture the different conditions of the corresponding test. For example, in the case of the left turn, a robot going straight forward will achieve a different score between the two scenarios, as it is initially oriented to the left in the first one. The score of a hint is the average of the scores of all its scenarios.

We give the user a maximal duration to complete each level, after which the system automatically passes to the next one. This duration is set to 5 minutes, except for level 3 which lasts 10 minutes, and for level 5 which lasts until the activity finishes. Figure 5 shows the system in action: the user is currently at level 4 (*labyrinth*) and the last run program obtained a mediocre score to the *empty*, *left turn* and *right turn* tests and a bad score to the *corridor* test.

4 EXPERIMENT AND RESULTS

To measure the efficiency of gamified hints on performance and learning, we ran a set of programming workshops for two days over a week-end. Seven sessions of 75 minutes each were proposed. Each session was composed of:

- (1) Introduction (10 minutes): the robot and the tablet software were introduced by a small demonstration.
- (2) Discovery activities (10 minutes): four easy tasks were given to the children.
- (3) Labyrinth activity (30 minutes): children were asked to program the robot to navigate in the labyrinth, like shown in the Figure 4.
- (4) Hand following activity (15 minutes): children were asked to program the robot to follow their hand.
- (5) Questionnaire (10 minutes): children were asked to fill a questionnaire, similar to Magnenat et al. [8].

The workshops were attended by 43 children including 7 girls between 9 and 12 years, with 4 to 10 children per session (average 6). No previous programming or robotics experience was required, but 22 children had already used the Thymio

robot and 15 reported previous programming knowledge, of which 11 had already used the robot. Children were provided one tablet and one robot each and shared four labyrinths. For the labyrinth activity, children were split into two groups, each given a different configuration of the software:

- The *Hints* group (24 children, 12 previous Thymio users, 8 with programming knowledge, 5 girls, ages: 9: 12, 10: 4, 11: 2, 12: 3) used the full system.
- The *Control* group (19 children, 10 previous Thymio users, 7 with programming knowledge, 2 girls, ages: 9: 10, 10:1, 11: 6, 12: 2) used a restricted version in which the levels were shown without indication of program quality, and were passed only at the end of their maximal duration.

Sessions Hints and Control were alternated during the weekend to avoid any bias due to the different assistant teams, their tiredness or their growing experience of the setup. Other activities were done with the bare VPL software for both groups, without providing any didactic sequence. We had no control on which child joined which session.

One of the authors and five university students provided assistance. The author was always present and the assistants were distributed over the two days, such that there were always two of them available. To avoid biases, the assistants were combined into three different configurations, which all supervised both types of sessions. Their role was to answer questions, and, when a child was blocked, ask her or him a related question to help progression and avoid frustration.

4.1 Research questions and results

Based on this experiment, we address 5 research questions on the effects of providing hints:

Is it feasible to perform a real-time evaluation considering the number of tests to run? We measure the time used to run the simulations of all tests. On an Android Nvidia Shield K1 tablet, the 8 scenarios, whose real-time durations add up to 53 seconds, are simulated within ≈ 1 second, including the computation of scores. Although the app freezes during this second, it is acceptable as students are often looking at the robot after clicking the run button. Therefore, we demonstrated the concrete feasibility of our approach.

Does feedback allow students to write programs faster? During the labyrinth activity, we measure the duration students took to successfully solve each level (1 to 4). As the Control group students could not pass levels before the end of the maximal test duration, the time at which they reached a level was recomputed offline during data analysis, using the algorithm as used online for the Hints group. Children in the Hints group took significantly less time than the ones in the Control group to reach the different levels (Table 1a). Yet one must be cautious as this result could be biased: children in the Hints group were more pushed toward completing each level, due to the fact that they could pass to the next level if they succeeded at the test of the current level, which was not the case for the Control group.

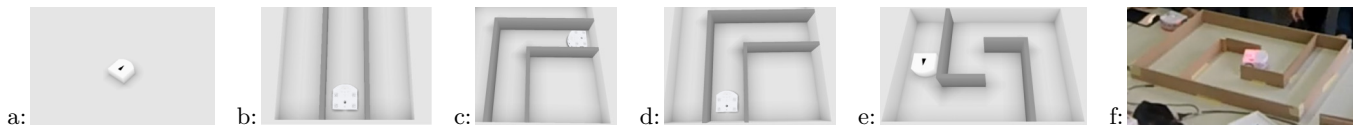


Figure 4: Different virtual environments used to test the labyrinth task (a to e); Real environment (f).

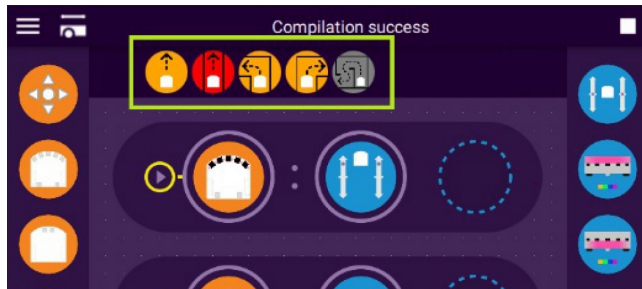


Figure 5: Hints in the Thymio VPL environment.

Does feedback allow more students to create better programs? During the labyrinth activity, we measure the best achieved score for each level (1 to 5) separately. In the Hints group, 95.65% of children achieved all tests, against 52.36% for the Control group, a statistically significant difference (Table 1b, p-value = 0.001).

Does feedback improve the building of transferable skills? During the hand following activity, we measure the duration students took to solve the task and the number of times they run a program. We observe no significant differences between the two groups (Table 1c). Therefore, we cannot answer this question with this study, but it seems unlikely that there is an improvement in transfer due to the hint system.

Does feedback improve the understanding of computer-science concepts? The answers to the questionnaire are converted into a grade, giving one point per correct answer. We see that the error rate is similar for both groups (Table 1d), so we cannot answer this question with this study. However, it seems that the hint system has no significant effect on the acquisition of computer-science concepts overall.

5 DISCUSSION AND LIMITATIONS

Our results might indicate that the scores obtained from the robot programming task and the grades in the questionnaire measure two distinct competencies. Thus, we believe that crossing the information from questionnaire assessment and performance metrics can help to better understand and assess the different programming abilities of students. Furthermore, by combining our system with human tutoring that focuses on theoretical and methodological foundations of computer science, our system could provide a motivating and stimulating playground to develop programming skills.

The proposed simulation framework is able to capture extensive statistics about the development of a program, including the evolution of program performance, the best program achieved or the frequency of program production.

Level	Hints	Control	p-value
Level 2	171	141	0.398
Level 3	336	553	0.014
Level 4	680	927	0.045
Level 5	961	1364	0.002

a. The mean level reaching time in seconds for the labyrinth activity, and the p-value of Student’s t-test (H0: means are equal), between the two groups.

Test	Hints	Control	p-value
empty area	0.89	0.70	0.065
corridor	0.72	0.74	0.498
left turn	0.88	0.55	0.003
right turn	0.94	0.60	0.003
labyrinth	0.96	0.57	0.001

b. The mean best scores for each test of the labyrinth activity, and the p-value of Student’s t-test (H0: means are equal), between the two groups.

Metric	Hints	Control	p-value
Number of runs	14.8	13.3	0.347
Task completion time	742	604	0.681

c. The mean number of runs and task completion time for the hand following activity, and the p-value of Student’s t-test (H0: means are equal), between the two groups.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Hints	0.42	0.00	0.05	0.32	0.00	0.63	0.32	0.11
Control	0.29	0.00	0.00	0.46	0.08	0.63	0.21	0.04
p-value	0.85	0.99	0.73	0.83	0.65	1.00	0.89	0.88

d. The error rate for Hints and Control groups; p-values of Pearson’s chi-square test (H0: both conditions follow the same distribution). n = Hints: 24, Control: 19.

Table 1: Quantitative results of the user study.

These statistics are useful for many different applications that go beyond a hint system. A distinct advantage of our solution over direct source code analysis is that our approach is agnostic to the underlying programming language as all metrics are based on the simulated behaviour of the robot.

From our qualitative observations during the experiments, the gamified hint system seemed to increase the students

involvement in the task. We observed that the performance feedback provided by the hints prompted students for a more critical assessment of their current solutions as they could directly observe which previously passing tests were no longer correct with their new solution attempt.

We have explored potential effects of factors such as gender, previous programming knowledge or previous exposition to Thymio, but could not draw relationships of similar importance as the use of Hints. The only statistically significant one is that children with previous programming knowledge completed the labyrinth significantly faster than the ones without (average 15 min 55 s against 21 min 32 s, $p=0.026$).

The main limitations of our work are the limited number of participants (43), the simplicity of the hand following task used to test transfer, and the short duration of the workshop. All these factors must be improved to find an answer to the unresolved questions of this paper. Moreover, the current implementation of the simulator only simulates a 2-D environment and its physics engine is simplistic. As such, it is not applicable to flying robots, for example.

6 FUTURE WORK

A direct extension of our work would be to combine the results obtained from the simulation with other data extracted from the programming environment. For instance, we could count the number of times a program is run on the robot or directly analyze its source code. We believe that, with enough data, it would be possible to automatically identify students who are struggling with a given task.

We plan to investigate the effect of directly providing the feedback from the hint system to teachers. Indeed, during an informal survey after the experiment, several assistants reported that it was easier to aid students when using the hint system, as it gives a high-level overview of the capabilities of their program. In addition, the progress made visible by the hint system allowed assistants to quickly identify students in need of guidance. Directly providing this feedback to educators through a tablet computer similar to the work of Maldonado et al. [9] has the potential to facilitate personalized interventions and improve learning.

7 CONCLUSION

In this paper, we presented an innovative way to guide students learning to program using robots. We demonstrated a gamified hint system based on the automatic assessment of robot programs. Our system is able to evaluate programs in real-time on a tablet device. By running the program in a virtual machine inside a fast robot simulator, it can be tested in different scenarios without affecting the user experience.

We ran a user study in which students had to program a robot to navigate a labyrinth. We demonstrated that our hint system increases the percentage of students writing correct programs from 50% to 96%, and decreases the average time to write a correct program by 30%. However, our findings on the effect on the learning outcome are inconclusive. We could not show any transfer of programming skills to the new

task of following the hand. Moreover, we did not find any correlation between the use of the system and the performance of students on a questionnaire testing concept acquisition.

Additional studies considering more tasks and a longer exposition to the system are needed to further investigate the relationship between the user guidance and the acquisition of computer science concepts. Nevertheless, the clear performance improvement in the task where students were guided shows the value of this approach for programming education using robots.

8 ACKNOWLEDGEMENTS

This work was supported by ETH Research Grant ETH-23 13-2, by Mobsya, by the Swiss NCCR “Robotics”, and by the Swiss CTI grant 17479.2 PFES-ES. The authors thank Moti Ben-Ari and the anonymous reviewers for their feedback.

REFERENCES

- [1] Marina Umaschi Bers, Louise Flannery, Elizabeth R. Kazakoff, and Amanda Sullivan. 2014. Computational Thinking and Tinkering: Exploration of an Early Childhood Robotics Curriculum. *Computers & Education* 72 (2014), 145–157.
- [2] Rohan Roy Choudhury, Hezhengm Yin, and Armando Fox. 2016. Scale-Driven Automatic Hint Generation for Coding Style. In *Intelligent Tutoring Systems*. Springer, 122–132.
- [3] Petri Ihantola, Tuukka Ahoniemi, Ville Karavirta, and Otto Sepälä. 2010. Review of recent systems for automatic assessment of programming assignments. In *10th Koli Calling International Conference on Computing Education Research*. ACM, 86–93.
- [4] Hieke Keuning, Johan Jeuring, and Bastiaan Heeren. 2016. Towards a Systematic Review of Automated Feedback Generation for Programming Exercises. In *21st Conference on Innovation and Technology in Computer Science Education*. ACM, 41–46.
- [5] Stéphane Magnenat, Morderchai Ben-Ari, Severin Klinger, and Robert W. Sumner. 2015. Enhancing Robot Programming with Visual Feedback and Augmented Reality. In *20th Conference on Innovation and Technology in Computer Science Education*. ACM, 153–158.
- [6] Stéphane Magnenat, Philippe Rétonnaz, Michael Bonani, Valentin Longchamp, and Francesco Mondada. 2010. ASEBA: A Modular Architecture for Event-Based Control of Complex Robots. *IEEE/ASME Transactions on Mechatronics* PP, 99 (2010), 1–9.
- [7] Stéphane Magnenat, Fanny Riedo, Michael Bonani, and Francesco Mondada. 2012. A Programming Workshop using the Robot “Thymio II”: The Effect on the Understanding by Children. In *Advanced Robotics and its Social Impacts (ARSO)*. IEEE.
- [8] Stéphane Magnenat, Jiwon Shin, Fanny Riedo, Roland Siegwart, and Morderchai Ben-Ari. 2014. Teaching a Core CS Concept Through Robotics. In *19th Conference on Innovation & Technology in Computer Science Education*. ACM, 315–320.
- [9] Roberto Martinez Maldonado, Judy Kay, Kalina Yacef, and Beat Schwendimann. 2012. An interactive teacher’s dashboard for monitoring groups in a multi-tabletop learning environment. In *International Conference on Intelligent Tutoring Systems*. Springer, 482–492.
- [10] Amit Kumar Mandal, Chittaranjan Mandal, and Christopher MP Reade. 2006. Architecture of an Automatic program evaluation system. *CSIE Proceedings* (2006).
- [11] Francesco Mondada, Michael Bonani, Fanny Riedo, Manon Briod, Léa Pereyre, Philippe Rétonnaz, and Stéphane Magnenat. 2017. Bringing robotics into formal education using the Thymio open source hardware robot. *IEEE Robotics and Automation Magazine* (2017). Accepted.
- [12] Kelly Rivers and Kenneth R Koedinger. 2015. Data-driven hint generation in vast solution spaces: a self-improving python programming tutor. *International Journal of Artificial Intelligence in Education* (2015), 1–28.
- [13] Juha Sorva. 2013. Notional Machines and Introductory Programming Education. *Transactions on Computing Education* 13, 2 (2013), 8:1–8:31.