Anticipating breakdowns with Kaspar: can children predict potential problems when interacting with a humanoid social robot?

Sílvia Moros, Patrick Holthaus, Gabriella Lakatos, Catherine Menon, Vignesh Velmurugan, and Luke Wood

Robotics Research Group, University of Hertfordshire, United Kingdom
s.moros2@herts.ac.uk,
WWW home page: https://robotics.herts.ac.uk/

Abstract. This paper reports the effectiveness of primary school children at predicting likely robot failures that may eventuate during a programming activity session of the Kaspar humanoid robot in a classroom setting. The paper also explores the relationship between the predicted failures, the actual failures that occurred during the programming activity session, and the impact of these predicted and actual failures on the children's enjoyment of the activity. We found that children could not accurately predict those failures of the robot which did organically occur during the session, but neither the predictions nor the failures affected their level of enjoyment.

Keywords: child-robot interaction, failure prediction, failure acceptance

1 Introduction

Robots are becoming increasingly present in our lives, including in classroom environments. Schools have been proposing or considering the use of robots as as a support in their curriculum for years, for example in lessons such as mathematics, physics, computer science and also the occasional biotechnology or history lesson [1, 2, 10, 3]. However, using robots has its own drawbacks: there may be ethical and safety issues, personalization is difficult, and to date robots are used for at most a few days instead of a full academic year [14]. On top of that, introducing another component to the classroom, such as a physical robot, increases the risk of errors and failures, given that robotic failures can occur both in terms of software and hardware. Recovery from robotic failures has been a topic of research in different contexts [11], [9]; in the HRI context one focal point of research has been the issue of how to recover trust, specifically, if a robot fails.

To our knowledge, research into robotic failures in a classroom context is still relatively new and unexplored. Some of this exploration has focused on children's perceptions of the failures and whether these affected their opinion of the robot interaction [7], or in assessing whether children could detect issues with a robot powered by ChatGPT in a mathematics class lesson [5]. These uncharted areas range from aspects such as children's perception of the failure, their opinion on the activity despite said failures [7], the preconceptions about robotic failures and more. Investigating these areas is therefore a crucial step in determining the feasibility of introducing robots as permanent features in the classroom. Therefore, in this study we address the following questions:

RQ1 Do children have an accurate estimation of the possible issues and failures which may occur with the robot during a robotic interaction activity?

RQ2 Do predicted or actual robot failures impact the children's enjoyment of, or willingness to repeat, the activity impacted?

Regarding RQ1, we hypothesized that the children would be relatively accurate at predicting possible robot failures during the programming activity session. This is because all the participating children have experience with the programming language (Scratch) used within the session. Additionally, some children already had a similar experience by programming non-humanoid robots with Scratch in the classroom in years prior. However, this was the first time for all participating children to programme a humanoid robot.

Regarding RQ2 we also hypothesized that children's enjoyment and willingness to repeat the programming activity would not be affected by predicted or actual failures. This is based on our previous experience in running and observing outreach activities with robots. We hypothesize that even making the children consciously think about failures will not affect their overall enjoyment.

2 Methods

To address the Research Questions we developed a study to take place within the UK Robotics Week, which is an outreach event where researchers conduct robotics activities within local primary schools. In this experiment, the children have an opportunity to see the humanoid robot Kaspar first (see 2.2) and are then asked to complete the pre-activity questionnaire. After this they engage in the programming activity session with Kaspar. Following this study, students are asked to complete a post-activity questionnaire.

2.1 Participants

Participants were children belonging to Years 3, 4 and 5 from a local primary school in Hertfordshire, United Kingdom. A total of approximately 170 children engaged in the programming activity session, however experimental data was obtained only from those for whom we had gained parental consent. This consisted of a cohort of 90 children, 50% female and 50% male. The children belonged to Years 3, 4 and 5, which corresponds to ages between 7 and 11 (MD = 8,91, SD = 0.9). Parental consent was obtained from parents/guardians/tutors before collecting the data in accordance with the ethics approval by the University of Hertfordshire's Health, Science, Engineering and Technology Ethics Committee with Delegated Authority (ECDA), number SPECS/SF/UH/05395.

2.2 The Kaspar Robot

The robot used for this study was the Kaspar robot. Kaspar is a humanoid robot developed by the University of Hertfordshire (Figure 1). Kaspar was developed to help children diagnosed with Autistic Spectrum Disorder (ASD) in areas such as emotion recognition, joint playing with other children and more [4, 13] Since its development in 2005, it has engaged with over 500 children, in several contexts including helping children with Speech and language disabilities [6] and with neurotypically developing children as an educational robot [8]. It is a stationary, minimally expressive robot that sits on a table with slightly folded legs. It communicates through verbal and physical gestures, including arm and head movements. Fully programmable with semi-autonomous features, Kaspar can respond to touch and has controllable arms, neck, torso and eyes.

Fig. 1. The Kaspar robot

2.3 Procedure

Each programming activity session lasted 1h and 20 minutes and could accommodate up to 20 children; the sessions were repeated until all children could participate, with a total of 9 sessions overall. Children were placed in pairs in front of a Kaspar robot and were given instructions on how to program emotions on it using Scratch, following the same outline as in [8]. All children had previous experience with both the Scratch interface and the Scratch language. The Scratch programming language is a block-based programming language and the children could use the blocks to trigger specific motors in the robot, thus expressing emotions through movement. The children could also use blocks to add sound to these emotions if they wished.

It is important to note that the robots were not purposefully modified in any way so that there were failure points; all failures or issues, if any, happened organically during the session. However, due to the nature of the activity, any failures or issues were rectified by the researcher or a member of the teaching team if the children could not rectify them themselves; this allowed the children to experience the full session without it being curtailed by any robotic failures.

2.4 Pre-activity Questionnaire

The pre-activity questionnaire asked children one question only: to predict which failures they thought they might encounter from a provided list of options Q1: "Which of these issues you think can happen in today's session? Mark all that apply.". A free-text "Other" option was also provided for additional issues or commentary. The complete list of issues and failures within the questionnaire was as follows:

- I will not know how to program Kaspar and tell it what to do.
- Kaspar will need to go to the bathroom.
- Kaspar will move, but NOT when it should.
- Kaspar will move, but it will NOT make any sound when I tell it to.
- Kaspar will NOT do what I told it to do.
- Kaspar will work but then break down.
- Kaspar will launch into space.
- Kaspar will NOT move at all
- Other

These questions have been chosen to illustrate the following categories, which include failures on both the part of Kaspar and the part of the (child) programmer:

- Kaspar fails to operate at all (e.g. "Kaspar will NOT move at all")
- Kaspar oeprates but with degraded functionality (e.g. "Kaspar will move but it will NOT make anyy sound when I tell it to"
- Kaspar operates but does not fully obey programmed instructions (e.g. "Kaspar will move but NOT when it should")
- Kaspar fails to operate because of a user or programming error (e.g. "I will not know how to program Kaspar and tell it what to do"
- Failures which indicate a lack of understanding of the robot (e.g. "Kaspar will launch into space")

2.5 Post-activity Questionnaire

After the programming session, the children were provided with the post-activity questionnaire, which asked them to report any robotic failures or issues which had occurred during the session, Q2: "Has any error happened in today's session? If yes, which one/s?". In this case no list of failures was provided for participants to select, but instead additional space was provided so participants could expand on the issue or failure. The post-activity questionnaire also asked two additional questions using a 5-point Likert scale as shown in in Figure 2: Q3: "Did you have fun programming the Kaspar robot?" and Q4: "Do you want to program the Kaspar robot again?"

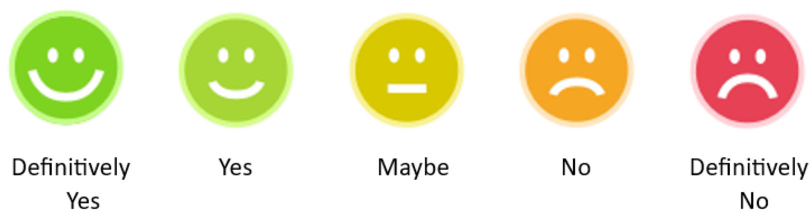

Fig. 2. Likert Scale used in Q3 and Q4

.

3 Results

3.1 Pre-activity Questionnaire

The frequency of each one of the answers is plotted in Figure 3. Only 12 children used the free-text entry. 3 of those answers were largely positive about their expectations, such as "It [Kaspar] will listen to me" (twice) or "Kaspar will move when I tell it to". 5 children predicted new errors, with text such as: "It [Kaspar] will shut down", "It will jump off the table", "Walk around", "It will make an explosion" and "Its face will fall off". Lastly, the remaining 4 children used this space to write about other topics not related to robotics or programming or their writing was not decipherable.

3.2 Post-activity Questionnaire

For the first question, "Has any error happened in today's session", 44 of the respondents (48,89%) answered "Yes".
Analysis of the free text provided showed that many of these answers did not correspond exactly to the pre-provided options in the pre-activity questionnaire,

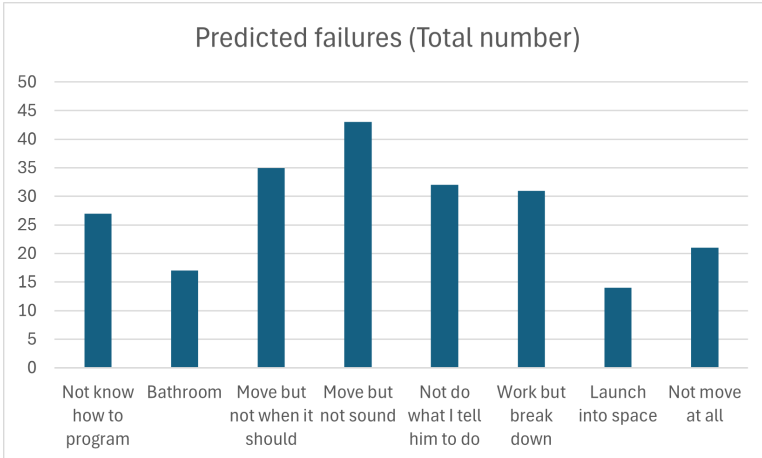

Fig. 3. Frequency of the issues and failures identified in the Pre-activity Questionnaire

demonstrating that the children were not simply repeating these from memory but instead providing additional detail.

For example, there were 10 answers not belonging to any category, which we included in the "Others" category. 7 of them clearly related specifically to the software, like "Little glitch" (x2), "It was glitchy", "It got deleted", "Everything got deleted" and "It was coming off the document". 3 of these responses identified issues unrelated to robot failures, such as "I was scared of Kaspar looking at me", "Yes" and "There wasn't an error on this".

By contrast, other answers could be clearly related to the pre-activity questionnaire options, even where the text entered was marginally different. For example, three children expressed issues related to their lack of programming knowledge on how to program Kaspar, with statements such as "We forgot to relax the Kaspar" (3). These statements have been categorised as "I will not know how to program Kaspar".

Other responses were directly hardware related, stating issues with parts of Kaspar stopping or not moving: these include the mouth, not smiling or closing it (4); the arms (5); its hand (3) and its eyes not winking (2), or body features not working without specifying which (1). These errors can be interpreted as the pre-activity questionnaire option "Kaspar will work but then break down". Some of the responses were more general, stating that Kaspar stopped (3). We have also included these in the "Kaspar will work but then break down" category.

Six children stated directly that Kaspar didn't move when instructed to and we have categorised these responses as the pre-activity questionnaire option "Kaspar will NOT do what I told it to". Two children reported Kaspar moving on its own, which we have categorised as"Kaspar will move but NOT when it should". Lastly, two other responses cited the sound as the issue, stating that "We didn't manage to make the sound out of it" and "The sound didn't work", and this was categorised as "Kaspar will move but it will NOT make any sound".

These results are shown in Table 1.

Table 1. Post-activity questionnaire analysis, using the pre-activity questionnaire categories

Category	Number of responses
I will not know how to program Kaspar	3
Kaspar will need to go to the bathroom	0
Kaspar will move, but NOT when it should	2
Kaspar will move, but it will NOT make any sound	2
Kaspar will NOT do what I told it to do	6
Kaspar will work but then break down	18
Kaspar will launch into space	0
Kaspar will NOT move at all	0
Other (*software*)	7
Other (*not related*)	3
No description of issues	3

A breakdown of the answers to the post-activity questionnaire question: "Did you have fun programming the Kaspar robot" can be seen in Figure 4, where we have assigned 1 to the answer "Definitively No" and 5 to "Definitively Yes", resulting in a mean score of 4.67, with a standard deviation of 0.67. The median was 5, "Definitively Yes".

The children who reported any kind of issue or failure had the same median but the mean score is 4.52. The children who did not report errors had a mean score of 4.8.

A similar breakdown for the last question, "Do you want to program the Kaspar robot again?" can be seen in Figure 5, again using a Likert scale with answers converted from 1 to 5 for analysis. In this case, the average score was 4.43 with a standard deviation of 0.86; the median was again 5, "Definitively Yes".

4 Discussion and Limitations

4.1 Discussion

In the pre-activity questionnaire children were asked to predict the possible issues or points of failure they might expect to encounter when programming Kaspar. Children were able to identify and select more than one type of issue, but the most selected answer was "Kaspar will move but it will NOT make a

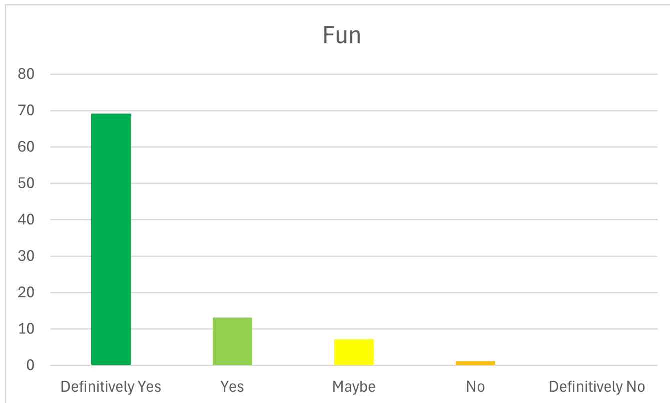

Fig. 4. Answers to the question “Did you have fun programming the Kaspar robot?”

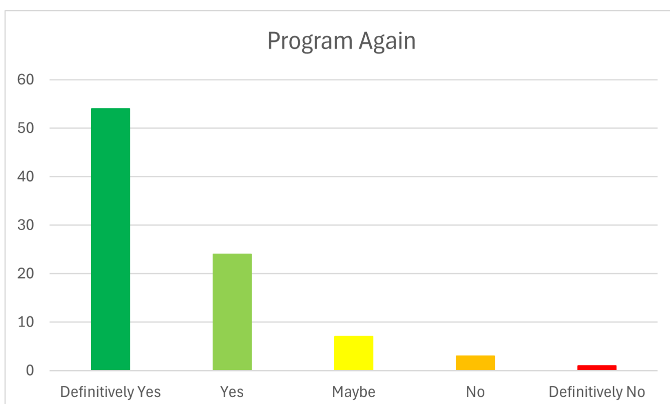

Fig. 5. Answers to the question “Do you want to Program the Kaspar robot again?”

sound". Since the students have had experience with Scratch used to program other robots but not with Kaspar specifically, this might be due to assumptions of how the Kaspar robot works, as the children couldn't know in advance whether Kaspar could be programmed to make sounds or not.
As stated in section 2.4, we added two options that were not possible: "Kaspar will need to go to the bathroom" and "Kaspar will launch into space". While it is true that neither of those options were widely selected (17 and 14 times respectively), some children still chose them. While this could be the children providing intentionally nonsensical answers, it could also be because some children genuinely believed that Kaspar could need a bathroom break, possibly because of its humanoid form.
Lastly, and based on the overall number of answers per category, it is evident that most children thought Kaspar would work in some way; the option "Kaspar will NOT move at all" was selected in only 21 responses out of 90. This may be interpreted to indicate that though the children are aware that issues and failures might occur during the programming session - perhaps as a result of their experiences with programming and other robots - nevertheless they still expect the robot to work to some extent.

Based on the responses of the post-activity questionnaire, we can conclude that the children were reasonably effective at detecting the issues and failures which occurred organically during the programming session. It is worth noting that the distinction between hardware and software failure, shown in the results, is a distinction drawn by us in post-analysis: the children themselves were largely oblivious to the difference between software and hardware failures during the session. Nevertheless, during our observation of the session it was apparent that a few of the children did recognize the differences, especially the most obvious ones: for example when the code was accidentally deleted or when the robot stopped moving when it previously moved while running the same code. They were all capable of recognizing when the robot's actions were not coherent to what they wanted, but they were not capable of analyzing whether this was due to their programming or due to the robot malfunctioning.

However, while the children detected the issues within the session, they were not accurate predictors, based on the analysis of the pre and post questionnaire. Children believed that the sound would be the biggest point of failure, with 47 out of 90 replies (52,22%), while in reality only 2 out of 44 children (4,5%) experienced issues with it. The main issue, as seen in Table 1, was the robot working and then breaking down or stopping with 18 out of 44 responses (40,9%). The second most encountered issues in the post-activity questionnaire, which were software issues, with 7 out of 44 (15,9%), were not explicitly identifed by any children in the pre-activity questionnaire, as no children wrote it down in their papers. Moreover, none of the children who explicitly stated that they were in some way responsible for the issues, such as "We forgot to press "Relax"", had chosen the "I will not know how to program Kaspar"" option in the pre-activity questionnaire.

Regardless, the fact that the children detected issues, and predicted others, did not appear to impair their enjoyment of the session, as the answers for the questions "Did you have fun programming Kaspar" and "Do you want to program Kaspar again" were both overwhelmingly positive, even when the children self-reported issues. This suggests that robotic failures in learning sessions do not impact the ability for learners to enjoy and get value from the session [12].

4.2 Limitations

As this was a study conducted in a real classroom, there was no way to prevent some spillage of the answers; i.e., that the children may have influenced the answer of other children by talking about the robot.

Another limitation could have been the excitement that children usually feel when playing with robots as this might have made them more forgiving of robotic failures. However, all children had experience with Scratch, so they have encountered software issues in the past; and, except Year 3 children, the rest had this programming robots experience before, albeit with other robots.

5 Conclusions and Future Work

In this study we have assessed whether children can be good predictors of failures when using a humanoid robot in a programming environment.

Children were generally keen observers and were quick to identify points of failure, but they were not usually able to discern whether these were software or hardware related. Moreover, they were not good predictors, as their predictions of what would the main issues be were not accurate; in general, children assumed sound would be the major failure point. This inaccuracy is partially explained by them never having programmed with Kaspar, and it reflects on the difficulty of making accurate predictions of failures by looking at the robot alone. Interestingly, none of them thought that the software can fail, despite having programmed non-humanoid robots with the software before. Nevertheless, this result is valuable because it points out that children typically do not assume software can fail.

Robotic failures or errors did not seem to have an impact on the children's enjoyment of the session, or willingness to engage in it again. This is encouraging because it suggests that robots have "grace" from children when they malfunction, and that children want to interact with them again despite the existence of issues. However, if robots are to be used as permanent features in classrooms, it is important to understand what children conceive of as failures, as well as to properly communicate the distinction between software and hardware failures and requirements. Mechanimsms to rectify any organically-occurring errors, both physically (such as fixing the parts) or emotionally (such as fixing the trust or willingness to interact) also need to be established.

In the future we propose to continue exploring children's understanding of the differences between software and hardware, as well as the impact of the failure points in other areas such as trust recovery.

Bibliography

[1] Anwar S, Bascou NA, Menekse M, Kardgar A (2019) A systematic review of studies on educational robotics. Journal of Pre-College Engineering Education Research (J-PEER) 9(2):2

[2] Chalmers C (2018) Robotics and computational thinking in primary school. International Journal of Child-Computer Interaction 17:93–100

[3] Chang CW, Lee JH, Chao PY, Wang CY, Chen GD (2010) Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. Journal of Educational Technology & Society 13(2):13–24

[4] Dautenhahn K, Nehaniv CL, Walters ML, Robins B, Kose-Bagci H, Mirza NA, Blow M (2009) Kaspar–a minimally expressive humanoid robot for human–robot interaction research. Applied Bionics and Biomechanics 6(3-4):369–397

[5] Helal M, Holthaus P, Wood L, Velmurugan V, Lakatos G, Moros S, Amirabdollahian F (2024) When the robotic maths tutor is wrong-can children identify mistakes generated by chatgpt? In: 2024 5th International Conference on Artificial Intelligence, Robotics and Control (AIRC), IEEE, pp 83–90

[6] Lakatos G, Gou MS, Holthaus P, Wood L, Moros S, Litchfield V, Robins B, Amirabdollahian F (2023) A feasibility study of using kaspar, a humanoid robot for speech and language therapy for children with learning disabilities. In: 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), IEEE, pp 1233–1238

[7] Moros S, Wood L (2023) You're Faulty But I Like You: Children's Perceptions on Faulty Robots. In: International Conference on Robotics in Education (RiE), Springer, pp 157–168

[8] Moros S, Wood L, Robins B, Dautenhahn K, Castro-González Á (2020) Programming a humanoid robot with the Scratch language. In: Robotics in Education: Current Research and Innovations 10, Springer, pp 222–233

[9] Salem M, Lakatos G, Amirabdollahian F, Dautenhahn K (2015) Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. In: Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction, pp 141–148

[10] Sapounidis T, Tselegkaridis S, Stamovlasis D (2024) Educational robotics and stem in primary education: a review and a meta-analysis. Journal of Research on Technology in Education 56(4):462–476

[11] Schulz T, Soma R, Holthaus P (2021) Movement Acts in Breakdown Situations - How a Robot's Recovery Procedure Affects Participants' Opinions. Paladyn, Journal of Behavioral Robotics: Special Issue Trust, Acceptance and Social Cues in Robot Interaction 12(1):336–355, DOI 10.1515/pjbr-2021-0027

[12] Stower R, Kappas A, Sommer K (2024) When is it right for a robot to be wrong? children trust a robot over a human in a selective trust task. Computers in Human Behavior 157:108,229
[13] Wainer J, Dautenhahn K, Robins B, Amirabdollahian F (2014) A pilot study with a novel setup for collaborative play of the humanoid robot kaspar with children with autism. International journal of social robotics 6:45–65
[14] Woo H, LeTendre GK, Pham-Shouse T, Xiong Y (2021) The use of social robots in classrooms: A review of field-based studies. Educational Research Review 33:100,388