International Journal of Child-Computer Interaction 43 (2025) 100709

Contents lists available at ScienceDirect

International Journal of Child-Computer Interaction

journal homepage: www.elsevier.com/locate/ijcci

Research paper

A rapid and inclusive instrument for assessing children's basic understanding of physical computing

Andrea Bonani [a], Rosella Gennari [b], Alessandra Melonio [c,*], Pierpaolo Vittorini [d]

[a] Direzione Istruzione e Formazione italiana, via del Ronco 2, 39100 Bolzano, Italy
[b] Free University of Bozen-Bolzano, via Buozzi 1, 39100 Bolzano, Italy
[c] Ca' Foscari University of Venice, Via Torino 155, 37010 Mestre, Venice, Italy
[d] University of L'Aquila, P.le S. Tommasi 1, 67100, Coppito, L'Aquila, Italy

ARTICLE INFO

Keywords:
Physical computing
Assessment
Understanding
Rapid questionnaire
Inclusive questionnaire
School
Teacher
Children

ABSTRACT

While there are many initiatives that strive to empower children with physical computing, there seems to be no validated questionnaire for rapidly measuring different children's understanding of the basics of physical computing. This paper presents the design of such an instrument—PCBUQ. It is rapid in that it consists of few items. It is inclusive because designed for different young children. It is for the basics of physical computing in that it considers physical input and output devices, basic patterns and programs that use them for interacting with the physical world. Data gathered from experts, primary and middle schools were used to validate PCBUQ. The first items assess children's capability of classifying physical devices as input (e.g., buttons), and output devices (e.g., LED, speaker). The other items evaluate whether children can interpret problematic scenarios and infer how to resolve them with adequate input and output devices, patterns and programs. PCBUQ was found to have adequate reliability. The reported statistical analyses highlight the items that strongly and weakly correlate with the construct under analysis, their difficulty and discrimination. Results are discussed to guide future physical computing initiatives for children and their assessment.

1. Introduction

1.1. Motivation and rationale

Over the past ten years, computational thinking (CT) has become the subject of numerous K-12 education interventions on computing, with global efforts to bring it into schools and different learning contexts (Angeli & Giannakos, 2020; Barr, Harrison, & Conery, 2011; Hayat Alajlan & Almassaad, 2023; Iwata, Pitkänen, Laru, & Mäkitalo, 2020; Kong, Abelson, & Kwok, 2022; Liu, 2023; Weintrop et al., 2016). The variety of interventions reflects the different views that exist on CT, which Denning and Tedre frame along six thematic dimensions (machines, methods, computing education, software engineering, design, computational science), to conclude that CT is "the mental skills and practices for designing computations" and "explaining and interpreting the world as a complex of information processes" (Denning & Tedre, 2019). Lodi and Martini instead frame CT in a historical perspective, and historically compare Wing's and Papert's views, "rectifying their original meaning" (Lodi & Martini, 2021). In particular, they point out that Papert, and then Papert and Harel have argued for the importance of engaging learners in constructing computational artefacts in ways that are meaningful to them, thus envisioning computing as an empowerment opportunity for all, by coupling CT with practice (Papert & Harel, 1991). Recently, researchers in the area of Participatory Design considered how to expand the domain of computing education for fostering CT to include contextual, human-centred and societal challenges (Dindler, Smith, & Iversen, 2020; Iversen, Smith, & Dindler, 2018). Resonating with a constructionism approach to computing, their approach "shifts focus from programming skills as an end in themselves towards providing children and young people with the means necessary to take part in technological development" and "include a concern for how children and youth are empowered through constructive, analytical and critical engagement with technology" (Dindler et al., 2020; Iversen et al., 2018).

This approach voices the need of differentiating computing education to reach further learners, including traditionally under-served populations of learners, such as females, and all those specialising in subjects other than computer science, and non-technical teachers. To achieve this goal, it is necessary to differentiate computer science

* Corresponding author.
E-mail addresses: andrea.bonani@scuola.alto-adige.it (A. Bonani), gennari@inf.unibz.it (R. Gennari), alessandra.melonio@unive.it (A. Melonio), pierpaolo.vittorini@univaq.it (P. Vittorini).

https://doi.org/10.1016/j.ijcci.2024.100709
Received 29 March 2024; Received in revised form 25 November 2024; Accepted 27 November 2024
Available online 9 December 2024

education and adopt new styles of teaching that promote broader technology skills beyond screen-based coding. Making, fabrication and other physical-computing activities have been employed with this goal (Bjørn, Menendez-Blanco, & Borsotti, 2022).

1.2. Background and related project

The work reported in this paper refers to empowerment initiatives that engage children in computing practices for developing digital technology for concrete problems, responsibly and creatively, with dedicated physical-computing toolkits (Gennari, Matera, Melonio, & Rizvi, 2024). Physical computing involves using programmable devices to create practical, interactive artefacts using software and hardware that can perceive and respond to events in the physical world.

The growth in popularity and diversity of physical computing initiatives has resulted in an increasing number of tools, platforms and courses designed explicitly for children (Hodges, Sentance, Finney, & Ball, 2020). One of the most popular devices available is the BBC Micro:bit (Austin et al., 2020). There are web-based development environments for it that support several different programming languages, e.g., block or text based (Ball et al., 2019). There are also interactive courses available for exploring and teaching physical computing with it or similar devices. An example is the physical computing unit of code.org, which aims at making learners explore the role of physical devices in computing; other courses focus on electronic circuits and how to control them with coding (Microsoft, 2024; Raspberry Pi Foundation, 2024).

The research presented in this paper originated from the 3P project, a collaborative initiative between computer science and interaction design researchers and a teacher specialising in computing education. The project aimed at empowering primary and middle schools through physical computing and design activities (Bonani, Gennari, Melonio, & Rizvi, 2022; Gennari, Melonio, & Rizvi, 2023). Briefly, it organised courses for teachers with Micro:bit and the IoTgo rapid-prototyping toolkit to acquire an understanding of physical computing to then design different smart things, which are every-day's things enhanced with physical computing. Each course started with IoTgo for exploring how physical inputs and outputs work, either on the Micro:bit or easy to connect to it, e.g., a PIR motion sensor. It continued with IoTgo boards, of progressive complexity. The basic boards offered teachers different simple scenarios for children, each with a problem or issue (e.g., a sickly person in a park) and a goal to achieve with a physical-computing solution, e.g., Gennari et al. (2023). An example is shown in the left side of Fig. 1. The boards also show resolution patterns and programs with an infinite loop and conditionals, with different conditions, with or without alternatives. See the right side of the aforementioned figure. Teachers were then invited to elaborate on their own problematic scenarios, for their classes, which became part of the materials for their interventions. The course then challenged participants to maintain the same patterns, and modify the given programs to address the problematic scenarios, with different physical inputs and outputs. In the end, the course asked participants to create their own smart things, for a scenario of interest, and explore more complex problems, resolution patterns and programs.

In case they wished so, teachers could then bring similar interventions at school with the course materials or materials available on online platforms, e.g., BBC (2024), Gennari et al. (2023) and Microsoft (2024). Their interventions were recommended to take 15 hours at school, organised in 3-h sessions, in different weeks. Before moving onto more complex problems and resolution algorithms, and into the design of novel solutions, however, it was important for teachers to have a formative assessment point to investigate whether the basics of physical computing were acquired, a prerequisite for the success of any physical-computing interventions for the empowerment of children. There was no rapid and standardised tool for assessing children's understanding of the basics of physical computing, in many mainstream schools and a short time. Thereby, an ad-hoc one was created, pilot tested, and assessed.

1.3. Focus and outline of the paper

This paper offers a rapid questionnaire, PCBUQ, resulting from the 3P project, for rapidly assessing children's understanding of the basics of physical computing. The first items ask to classify physical-computing devices according to how they work. Then other items ask to critically interpret problematic scenarios, and infer the relevant physical devices for tackling them. The final items ask to interpret problematic scenarios, and infer resolution patterns and programs for them, relying on an infinite loop and conditionals.

The paper starts discussing the most relevant related work concerning physical computing interventions and assessment instruments for children. The PCBUQ instrument, which is the focus of the paper, is presented next. Its development process is detailed, considering its goal, format and content in relation to a relevant taxonomy for computing education interventions for non-computing experts. A field study with primary and middle school children was run for standardising the instrument. This part is detailed next in the paper. Results are discussed in the conclusions to the paper, considering the limitations of the reported study and future work.

2. Related work

2.1. Physical computing

Physical computing is one of the most prevalent methods of integrating computing into K-12 education. Physical computing activities require CT, namely, abstraction, algorithmic thinking, automation, decomposition, debugging, and generalisation (Psycharis, Kalovrektis, Sakellaridi, Konstantinos, & Mastorodimos, 2018). Moreover, it combines computation with a physical system, and a smooth integration with other design-oriented disciplines, as in the 3P project reported above.

Physical computing was born from Papert's concept of constructionism, which sustains that learners can understand new theories about their world if they construct these new theories from their own experiences, and that the best way of doing that is through the creation of a shareable object by learners (Blikstein, 2013). The nature of these shareable objects has evolved greatly over the years, stepping from the early physical computing kits with programmable bricks, such as Lego Mindstorms RCX, to embeddable micro-controllers such as BBC Micro:bit or Arduino (Austin et al., 2020; Galadima, 2014; Lego, 2024).

Physical computing toolkits have been used in educational contexts to learn about computational concepts by engaging learners in the making of interactive projects, to apply those concepts in authentic environments (Kim & Min, 2021; Neutens & wyffels, 2016; Przybylla & Romeike, 2014; Stager & Martinez, 2013). As reported by Katterfeldt and colleagues, learning is more engaging when one interacts with and programs tools and materials to create a computational artefact. Classes that use physical computing tools provide opportunities for learners to communicate differently by using and sharing physical artefacts (Katterfeldt, Dittert, & Schelhowe, 2015). Additionally, physical computing enables learners to make abstract programming concepts concrete and tangible, besides providing them with hands-on experiences (Hodges et al., 2020; Przybylla & Romeike, 2014).

In the physical-computing and making literature, multiple examples exist of successful integration of physical computing into empowerment initiatives for children and young people. Several situated studies have shown that incorporating physical computing tools such as Micro:bit into specific classes can significantly empower them to critically reflect on the developed technology (Gennari, Matera, Melonio, Rizvi, & Roumelioti, 2020; Lechelt, Rogers, & Marquardt, 2020; Wu & Su, 2021). Like the one reported in this paper, several initiatives guide children in crafting their own digital solutions, in the form of smart things, with various physical-computing components, allowing for different degrees of autonomy (Gennari et al., 2020; Gennari, Melonio, & Rizvi, 2022;

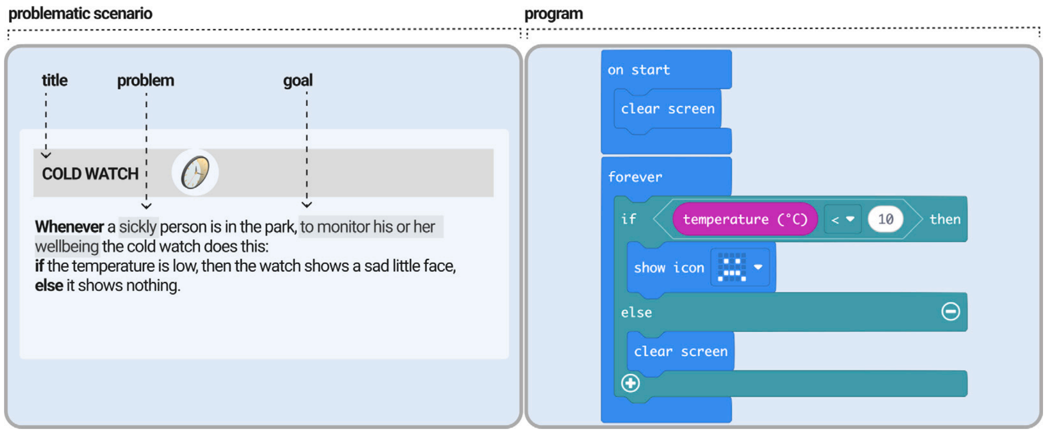

Fig. 1. An example of a simple problematic scenario, resolution pattern and program.

Wang, Zhao, Kleek, & Shadbolt, 2023). Finally, the benefit of "getting physical" is a gained view of computing as a tool that "encourages creativity, promotes learning by doing and engaging the whole mind and body learning" (Cápay & Klimová, 2019).

2.2. Assessment

Much research has been published on assessing computational concepts, in relation to CT (Babazadeh & Negrini, 2022). Various CT-based methods have been used to measure the efficacy of CT interventions (Chiazzese, Arrigo, Chifari, Lonati, & Tosto, 2019). These methods range from specific quantitative instruments like questionnaires, to qualitative instruments like interviews and product analyses.

Among quantitative instruments are Visual Blocks Creative Computing (TVBCC) and CT-Test. They are comprehensive questionnaires, with 40 and 28 items respectively. They assess specific programming concepts such as loops andsequences (Brackmann et al., 2017; Román-González, Moreno-León, & Robles, 2017; Sáez-López, Román-González, & Vázquez-Cano, 2016). Others use large sets of tasks to assess CT skills, such as Bebras tasks (Chiazzese et al., 2019), which consist of small problems or challenges, relying on unplugged activities. As for children's learning of computing, the VELA Assessment instrument assesses middle-school children's learning of foundational programming concepts (Grover, 2020).

To empower learners beyond specific computing concepts, alternative forms of assessment are required, e.g., interviews, observations, thinking-aloud or eye-tracking methods (Greenstein, 2012; Miliou, Adamou, Mavri, & Ioannou, 2023). In 2021, Liu conducted an integrative review of CT assessment, which shows the prevalence of assessment items and artefact analysis, which tend "to regard CT as a learning outcome", and not as a process to evaluate step by step (Liu, Luo, & Israel, 2021). They focus specifically on neglected assessment methods that consider visual behaviours and verbalisation. In general, these and similar assessment methods draw on learners' perspective. They present they own challenges, though, e.g., they tend to be time-consuming, requiring specific skills that are not available normally in school settings (e.g., eye-tracking) and, if qualitative, difficult to generalise. They support the need of rapid, formative assessment instruments.

Assessing physical computing interventions for children is a complex issues per se, differently tackled. For instance, Bieraugel and Neill developed a self-assessment tool to measure learners' confidence with digital fabrication technologies, but faced challenges with complex knowledge (Bieraugel & Neill, 2017). Bowler and Champagne used question prompts to encourage reflections in young people about technological artefacts in makerspaces, noting that preparation time and time for reflection need to be factored in Bowler and Champagne (2016). When physical computing kits are used in empowerment projects, the evaluation of their use by children is often situated and qualitative, e.g., Lechelt et al. (2020). In other cases, researchers have developed specific and contextualised assessment tasks with programming paradigms and languages for children (Microsoft, 2024; Raspberry Pi Foundation, 2024). The code.org course offers circa 10 classes on physical computing, with a specific physical computing board. At the end of each class, there are 2-4 tasks related to the class material, such as to read a problem and program a resolution from scratch, without pictorial representations.

However, inclusive assessment instruments for children require a specific design, which takes care of ethical considerations and their expression means. Younger learners tend to need structured information and support to master their self-assessment (Greenstein, 2012). In particular, in a mainstream school setting with many different children and strict timing, questionnaires need to cater for them all, and the administration has to consider time constraints (di Mascio, Melonio, Tarantino, & Vittorini, 2014). Not only questionnaires need to be rapid to administer and process, but, given the variety of children in mainstream schools, such instruments need to consider that some of them may be unable to read fluently. Therefore, to be inclusive, they are recommended to adopt a visual format, so that textual items are accompanied by pictorial representations; they should also pay attention to characteristics such as brevity, word simplicity, structured format, gender issues, besides pictorial representations (Harter & Pike, 1984).

Overall, to the best of our knowledge, there is no rapid and inclusive questionnaire for assessing the understanding of the basics of physical computing in mainstream classes, which is relevant for the success of physical-computing interventions for their empowerment. This paper offers it.

3. Methods and material

3.1. Goal and objectives

The Bloom's taxonomy, in the original (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1964) and revised versions (Anderson, Krathwohl, & Bloom, 2001), has been employed in diverse manners in the computing education field, e.g., for the instructional design of classes, their evaluation, structuring this or grading the evaluation in difficulty levels.

Bloom's Taxonomy (Bloom et al., 1964) outlines a framework for categorising educational goals, divided into stages, to help educators guide learners' progress. Krathwohl revised and specialised Bloom's taxonomy for computer programming education. This paper leverages the revised Bloom's taxonomy by Anderson et al. (2001) to conceptualise cognitive stages, and it draws on the discussion of Bloom's classification by Thompson, Luxton-Reilly, Whalley, Hu, and Robbins

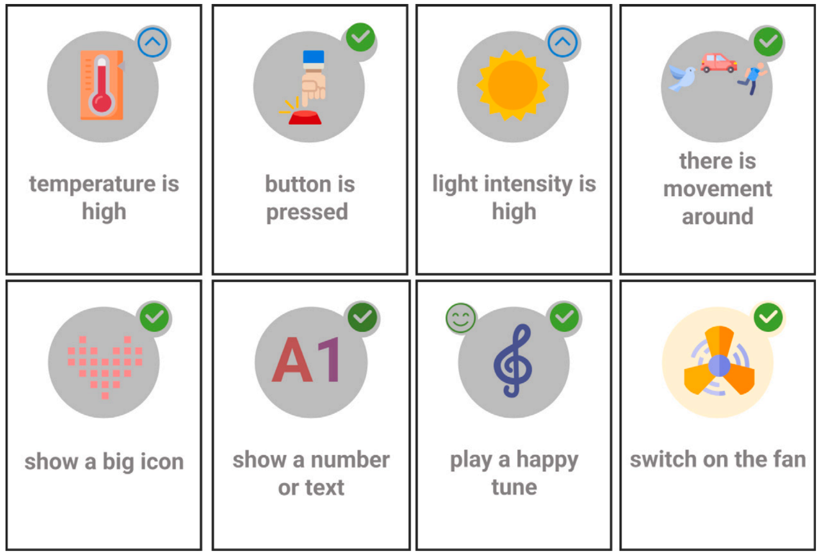

Fig. 2. Examples of visual and textual representations in PCBUQ.

(2008). Specifically, Thompson et al. provided concrete examples, enabling computer science educators to effectively use Bloom's taxonomy for programming assessment and evaluating introductory computing educational activities. The revised Bloom's taxonomy by Krathwohl et al. envisions different cognitive categories, starting from remembering, stepping through understanding, and ending with creating. In particular, understanding is defined as constructing meaning from instructional messages, and it includes the following:

- classification,
- interpretation,
- inference.

By referring to it, the questionnaire presented in this paper has the overall *goal* of assessing children's *understanding of the basics of physical computing*, to then create their own novel solution for a scenario. Understanding that requires the capability of *classifying* physical devices into input devices (e.g., buttons, touch sensors) and output devices (e.g., LED, speaker). It also takes to be able to *interpret* a problematic scenario and *infer* how to resolve it with input devices, output devices and a program that follows a suitable pattern, as exemplified in Fig. 1. Therefore the questionnaire frames its goal along the following *objectives* to assess:

- classifying devices as per their function,
- interpreting problematic scenarios, and inferring resolutions.

The study in this paper aimed at assessing the validity and reliability of the questionnaire.

3.2. Development process

3.2.1. Item generation, content and format

The Physical Computing Basic Understanding Questionnaire (PCBUQ) was designed by considering that the main target population included children who are expected to be capable of reading but have varying fluency, vocabulary and reading comprehension skills. In line with what recommended in the literature on assessment, outlined in the related work section, that means that items of PCBUQ were designed with pictorial representations besides textual information. Fig. 2 shows such choices in the representation of devices and the states they are in, with pictorials and textual descriptions.

Again in line with recommendations from the literature, text in items was kept simple and at the minimal level. Moreover, the questionnaire was designed for children who had experienced at least the basics of physical computing, as explained in the background part of the paper—the working of physical input and output devices, infinite loops and conditionals. Since MakeCode is a popular visual programming environment, also used in the 3P project courses, items of the questionnaire used it. This could be replaced by any other physical-computing programming environment for primary and middle school children, as elaborated in the paper conclusions.

The preliminary version of PCBUQ had 7 items. Items targeted the aforementioned understanding objectives of the Bloom's taxonomy: classification, interpretation and inference. It was validated by an expert panel.

3.2.2. Expert panel

The preliminary 7-item questionnaire was submitted to a group of experts: two teachers and three researchers were invited by e-mail to participate in the validation process. Specifically, the expert researchers evaluated whether the items met the criteria of clarity, comprehensibility and suitability to the objective, freely expressing their suggestions and observations.

The aim of the *clarity* criterion was to assess whether the wording and pictorials of the item used a language and representations that would be immediately clear to the target children. The aim of the *comprehension* criterion was to assess whether what the item addressed could be easily understood by children. For instance, anaphora resolutions were removed whenever unnecessary or confusing. See an example in the text of Item 3 in Fig. 3, which uses a second occurrence of "map" instead of a pronoun for enhancing comprehension.

The aim of the *suitability* criterion was to assess whether the item was adequate for achieving the proposed objective. The original Item 1, in particular, represented different input and output devices like in Fig. 2, for a total of 36. It asked children to classify them all, as either input devices, by crossing them, or output devices, by circling them. Item 1 was revised and split into two items, each with 3 possible answers, out of which only 1 was correct. The first asked to identify an input device, and the second to identify an output device. See the revised first item in Fig. 4.

The validation occurred in two rounds. In the first round, each expert evaluated whether each item was clear, comprehensive and suitable. Consensus was considered reached when >70% of the experts agreed. Items for which the minimum consensus was not reached were rewritten in the second round (McMillan, King, & Tully, 2016). The resulting questionnaire had 8 items, and it was tested in pilot studies.

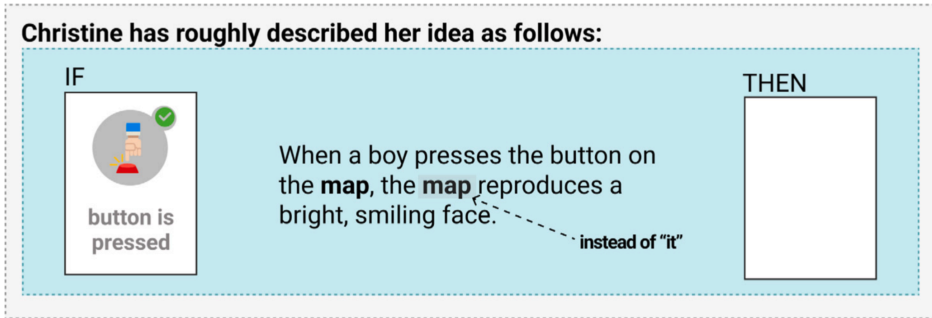

Fig. 3. Revision on text of Item 3 of the questionnaire due to the expert review.

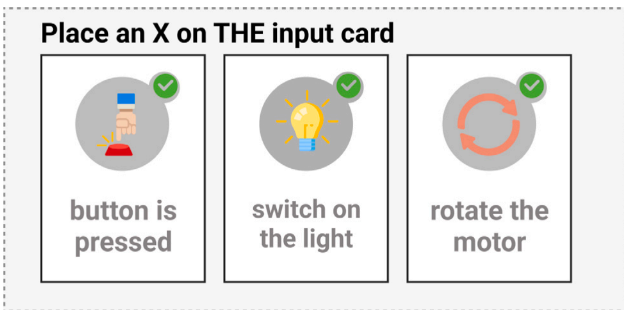

Fig. 4. Item 1 of the questionnaire after the expert review.

3.2.3. Pilot studies

A series of convenience-sampling pilot studies were implemented to refine the 8-item questionnaire. All participants had a basic understanding of how to design and program their own solutions with physical computing.

The *first pilot study* included four 8–13 year olds. After answering individually the questionnaire, they were asked about its questions by a researcher, expert of the field. All children stated that they understood the questions and how to answer them. According to the researcher, the initial questions were too easy and required more items to choose from. Specifically, Items 1 and 2, revised in light of the expert review, had only 3 possible answers, and the researcher observed that children could easily guess the single correct answer, e.g., see Fig. 4. They were thus revised by increasing the number of answers to 4, with a further device to classify, e.g., by referring to Fig. 5, "temperature is high" was missing and added as further correct answer.

The *second pilot study* involved a primary school (sixteen 10–11 years olds) and a middle school (thirteen 12–14 years olds), and the same researcher of the first study. It was managed as in the previous study. All children stated they understood the questions and liked its visual format. However, the researcher observed that the younger children had comprehension issues with Item 5; the format and text of Item 5 were thus changed to make it more understandable for younger children, e.g., by providing possible answers instead of asking children to write them down, thus promoting recognition rather than recall. The initial and final formats are illustrated in Fig. 6. Moreover, another revision was deemed necessary for Items 3 and 4, which initially had 3 answers instead of the final 6, as all children had guessed the correct answers. See for example the resulting Item 3 in Fig. 7.

3.3. The revised questionnaire

In developing items, researchers addressed the major issues identified by the expert panel and during the pilot studies, as briefly outlined above. This process led to the creation of a concise instrument with 8 items. Like for the preliminary questionnaire, the final items of the revised one were organised into expected progressive levels in relation to the understanding objectives from the Bloom's taxonomy—classification, interpretation and inference.

Table 1
Items of the revised questionnaire, their level and understanding objectives from the revised Bloom's taxonomy.

Item	Level	Understanding objectives
1, 2	Basic	Classifying devices into input and output
3–5	Intermediate	Interpreting a problematic scenario and inferring resolution input and output devices
6–8	Advanced	Interpreting a problematic scenario and inferring a resolution pattern and program

Specifically, the revised questionnaire starts from what are expected to be basic level items, demanding to *classify* physical devices into input and output ones (Items 1–2). See Fig. 5 for an example. It continues with items that require to *interpret* a problematic scenario and *infer* the correct physical device for tackling it computationally (Items 3–5), of intermediate level. See Fig. 7 for an example. It ends with items that are considered of an advanced level, each requiring to *interpret* a problematic scenario and *infer* a resolution pattern and program (Items 6–8). See Fig. 8 for an example. See Table 1 for a recap of the structure of the questionnaire.

3.4. Methodology

The revised questionnaire was analysed in terms of reliability, difficulty, discrimination and structure as follows.

3.4.1. Reliability analysis

Researchers measured the *reliability* of the questionnaire through Cronbach's alpha (Cronbach, 1951). Cronbach's alpha ranges from 0 to 1, with higher values indicating that the questionnaire is reliable in measuring the construct under evaluation and vice-versa. Even though no standardised thresholds is available, the thresholds of 0.5, 0.7 and 0.9 are commonly used to indicate a low, acceptable, good, or too-high alpha value[1] Nunnally and Bernstein (1994). Two actions can be taken regarding a low Cronbach's alpha: (i) items with low correlation with the questionnaire total score should be discarded or revised, and (ii) items that may add more variability to the questionnaire should be added. On the other hand, an alpha that is too high suggests that some items may be redundant, i.e., they ask the same thing in different ways.

[1] It is worth mentioning that the thresholds are also domain-dependent. For instance, in medicine, an instrument is considered reliable if it shows an $\alpha \approx 0.9$, whereas, in other research fields, an $\alpha \approx 0.7$ is considered a value for good reliability.

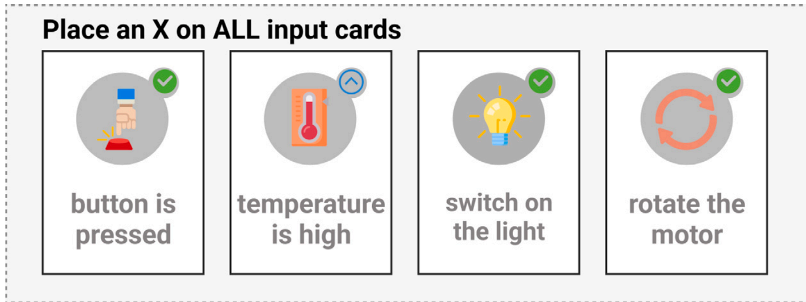

Fig. 5. Item 1 of the questionnaire after the pilot studies.

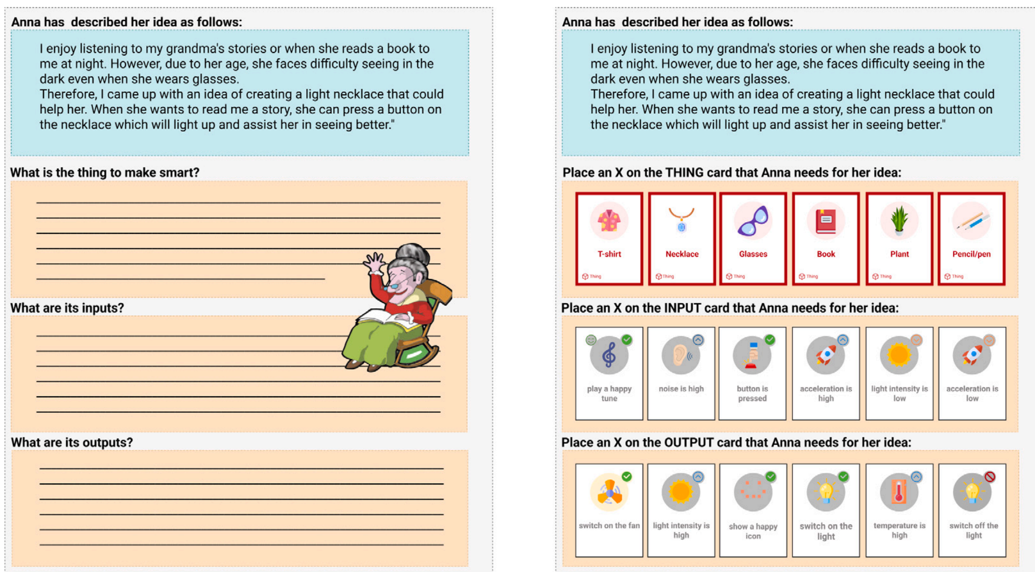

Fig. 6. An example of the outcome of the pilot studies for Item 5: the initial one (left), and the revised one (right).

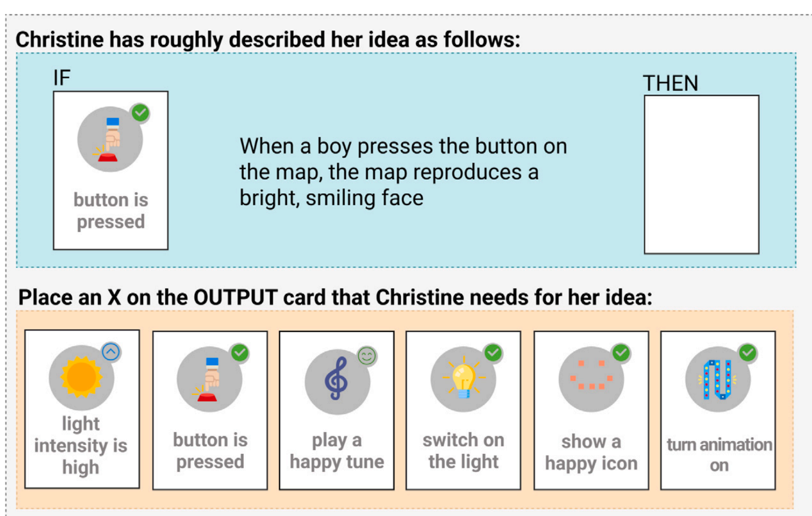

Fig. 7. Item 3 of the questionnaire after the pilot studies.

3.4.2. Difficulty and discrimination analysis

To measure the characteristics of the questionnaire's items, several approaches can be followed. Item Response Theory (IRT) is among the most consolidated (Hambleton & Swaminathan, 1985). Through several models and algorithms, different parameters can be estimated. This research focused on the *difficulty* β and *discrimination* δ parameters. The higher the β, the higher the item's difficulty; the higher the δ, the higher the item's ability to differentiate students.

3.4.3. Structure analysis

A confirmatory factor analysis (CFA) was chosen to evaluate if children's answers fit the questionnaire *structure*, which divides items

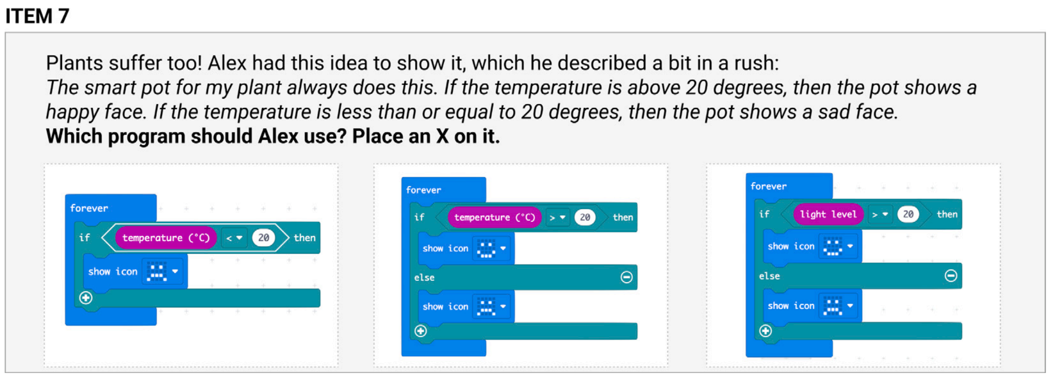

Fig. 8. Item 7 of the questionnaire after the pilot studies.

into levels according to understanding objectives of the Bloom's taxonomy (Thompson, 2004). See Table 1. To this aim, the questionnaire model was defined: (i) a latent variable was introduced for each row of Table 1; (ii) each latent variable was considered inferable from the respective items. A CFA then explores if: Items 1 and 2 measure the expected basic level on classification; Items 3, 4 and 5 evaluate the intermediate level about interpretation and inference; Items 6, 7 and 8 assess the advanced level about interpretation and inference. Therefore, the following was calculated:

- the correlation of each item with its related latent variable. The correlation coefficient may range from −1 to 1: the highest is the (absolute) value, the strongest (positive or negative) is the correlation;
- the Root Mean Square Error of Approximation (RMSEA) to measure how far the data are from the theoretical questionnaire structure. The lower the RMSEA, the better the data fits the theoretical model. Values less than 0.05 are good, values between 0.05 and 0.1 are acceptable, and values greater than 0.1 are poor (Fabrigar, MacCallum, Wegener, & Strahan, 1999).

3.4.4. Final analysis specifications

To deal with missing data, the fully conditional specification method was adopted, where each incomplete variable is imputed by using specific models (e.g., random forest, linear/logistic regressions) (Van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2007). All analyses were executed using R 4.3.1, with packages "psych" for the reliability analysis, "mirt" for the IRT analysis, and "lavaan" for CFA (Philip Chalmers, 2012; Revelle, 2022; Rosseel, 2012).

4. Field study

4.1. Participants and consent

According to the results of the first pilot study, which included a small number of children of a larger age range (8–13 years old), younger children could only tackle the first two classification items of the questionnaire. In view of this and their very different developmental capabilities, the questionnaire was administered to 163 children, aged 10–14 year old, attending mainstream primary and middle schools. A total of 75 children, aged 10–11 year old, were from primary school. Additionally, 88 children were from middle school, distributed among the 6th grade (17 children, aged 11–12), 7th grade (44 children, aged 12–13), and 8th grade (27 children, aged 13–14). Only 116 complete questionnaires were returned, which were retained for the data processing detailed below.

4.2. Procedure

The questionnaire was administered after children had experienced the basics of physical computing, outlined in the introductory section. In brief, they had experienced diverse physical input and output devices, infinite loops, and conditionals. These had a condition that depends on the state a physical input device is in (*if a button is pressed*), and a reaction of a physical output device (e.g., *then the LED matrix shows a happy icon*). Moreover, children were introduced to alternative conditions (e.g., with *if–else* control-flow structures), besides sequencing. Examples are in Fig. 8.

The intervention for the basics of physical computing was run by teachers during their class hours. Timing depended on school-related constraints, and the organisation of an intervention could not be controlled. However, in total, each intervention lasted circa 15 h, as recommended. The teacher with expertise in computing education, who had managed the 3P courses for teachers with researchers, was also present in classes at school. Moreover, the questionnaire was always administered by teachers with the support of the external expert teacher for a correct administration, again during their class hours, at the conclusion of the intervention. Additionally, teachers read the questionnaire aloud to the entire class to ensure that all children understood the items. Both teachers and researchers were present while it was being completed by the children, to identify and address potential misunderstandings or doubts the children might have had regarding its items. Their presence ensured clarity and supported children in completing the questionnaire confidently and accurately. The questionnaire was administered in paper format, anonymously. Completing the questionnaire took between 20 and 40 min circa. Data were then organised in electronic tabular format, and cleaned by two researchers of 3P. Then, another researcher, expert in psychometric measurements, processed them with the aforementioned methodology.

5. Results, discussion and revisions

Results are first presented and discussed. The main revisions to the questionnaire are thus derived and explained per item.

5.1. Results and discussion

The value of Cronbach's alpha calculated on the 116 sample units is $\alpha = 0.58$. It can be considered acceptable, even if with space for improvements. A detailed analysis is reported in Table 2. In the first three columns, the table reports the item number, the reliability if the item is dropped (column α) and the item correlation with the questionnaire (column R). The results suggest that the "critical" item is Item 5 (marked in grey): it is the least correlated item and the one that — if removed — could improve the reliability of the questionnaire.

The table also reports the difficulty and discrimination parameters of the IRT analysis, defining the items' probability functions depicted

Table 2
Analysis details. Column α shows the reliability if the corresponding item is dropped; column R shows the item correlation with the questionnaire; column β reports the difficulty of the corresponding item and δ the discrimination; column *LV* contains the names of the latent variables and *Corr.* the correlation coefficient of the item with the respective latent variable.

Item	Reliability		IRT		CFA	
	α	R	β	δ	*LV*	*Corr.*
1	0.54	0.57	−0.04	0.94	B_{1B}	0.68
2	0.56	0.50	1.02	0.89		0.54
3	0.56	0.44	1.89	0.93	B_{2I}	0.49
4	0.55	0.48	1.35	0.95		0.61
5	0.60	0.40	0.39	0.22		0.08
6	0.52	0.56	1.32	1.44	B_{2A}	0.56
7	0.53	0.51	3.32	2.35		0.56
8	0.53	0.56	0.04	1.26		0.46

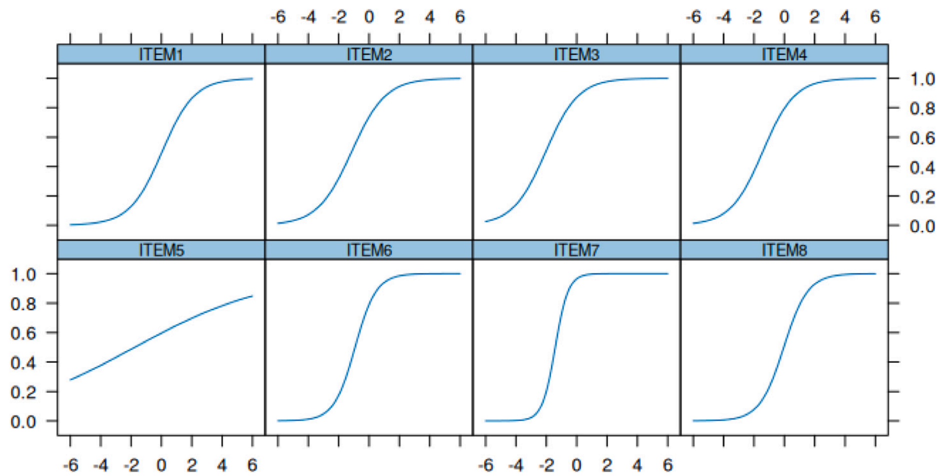

Fig. 9. Items' probability functions.

in Fig. 9. Each curve shows, for each item, how the estimated student's ability (on the x-axis) is related to the probability of answering correctly the corresponding item (on the y-axis). The parameters indicate that the first and the last items are the easiest (in green), whereas the seventh is the hardest (in violet).

According to this analysis, Item 2 tends to be more difficult than Item 1, both targeting basic classification; Item 1 targets the classification of input devices, with a motor and light as distractors; 2 addresses the classification of output devices, with an accelerometer and a slider as distractors. Finally, regarding the advanced items, Items 6, 7 and 8 have a loop, and at least a conditional statement. A difference between Item 7 and the others is in the condition—Item 7 refers only to analogical sensors (i.e., for temperature and light). Moreover, the phrasing of the condition is also different: Item 7 negates the *if*-condition instead of using *else*. The analysis shows that their difficulties differ significantly: Item 7 is the hardest, and Item 8 is among the easiest. However, Item 8 like Items 6 and 7 have high discrimination, with Item 7 being the most discriminant (in violet). Most critically, Item 5 is neither the easiest nor the hardest item according to the IRT analysis, and it is the item least capable of discriminating performances.

The CFA returned a very good RMSEA of 0.017, which means that the data fit the questionnaire's theoretical structure well. Details are in the last two columns of Table 2 and in Fig. 10. Item 5 is again the critical item, because it almost does not correlate with the expected latent variable (B_{2I}, for interpretation and inference, intermediate level). Moreover, Items 1 and 2 show, respectively, a good and average correlation with their B_{1B} latent variable (classification, basic). Average correlations are found in the other items for their latent variables, B_{2I} and B_{2A} (interpretation and inference, intermediate and advanced). Focusing on the latent variables, Fig. 10 shows that the lowest correlations are between the basic (B_{1B}) and intermediate/advanced (B_{2I}, B_{2A}) latent variables and the highest is between the intermediate (B_{2I}) and advanced (B_{2A}). A low correlation indicates discriminant validity between the two latent variables, i.e., they represent distinct constructs, and vice versa. This means that the first two items distinctly measure the basic abilities, whereas the other items measure more intertwined abilities that require both interpretation and inference.

5.2. Revision decisions

The final questionnaire, PCBUQ, was obtained as follows. In line with the results of all analyses, Item 5 was dropped. It is the only item that, once dropped, increases reliability, from 0.58 to 0.60; it is neither the easiest nor the most challenging; it is the least capable of discriminating learners' performances; it is the item least correlated with the corresponding latent variable.

According to the CFA (see Fig. 10), all other items could be considered to match the expected level (basic, intermediate, advance) of understanding of the revised Bloom's taxonomy. In particular, B_{2I} and B_{2A}, targeting interpretation and inference, showed to have good correlation.

According to the difficulty results reported in Table 2, Item 2, which requires to classify outcome devices, tended to be more difficult than Item 1. It was kept as is in PCBUQ to maintain variability.

Item 7 turned out to be hardest according to the IRT difficulty analysis. It is worth reminding that Item 7 differed from Items 6 and 8 in the usage of analogical sensors, and the phrasing of alternative condition. Thus, in PCBUQ, to slightly simplify its difficulty, Item 7 was modified with *else* for phrasing the alternative condition, but it kept analogical sensors to maintain variability. Item 8, though simpler than the other Items 6 and 7, had a good discrimination ability, comparable to that of Item 6, and hence it was maintained as is in PCBUQ.

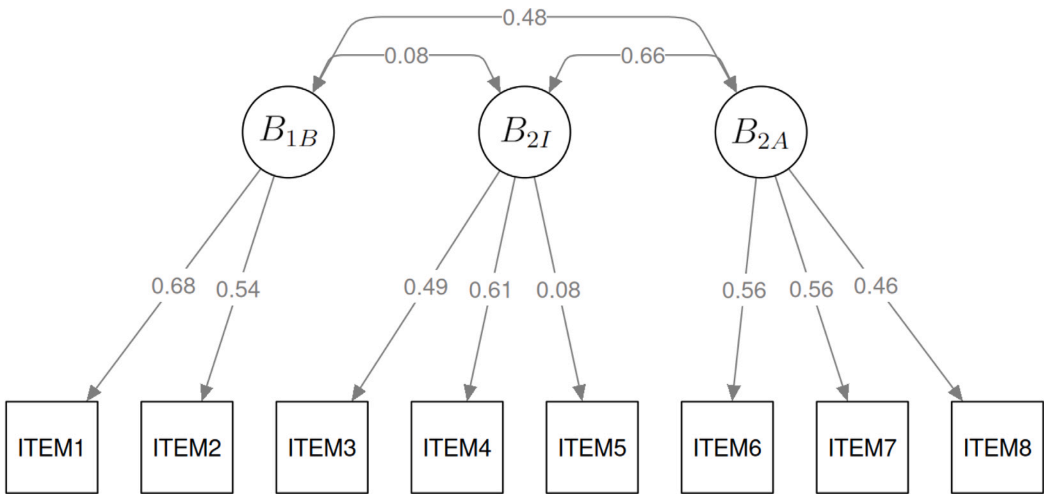

Fig. 10. CFA for the structure of the questionnaire, shown in Table 1. On the top the latent variables: B_{1B} = classification, of basic level; B_{2I} = interpretation and inference, intermediate; B_{2A} = interpretation and inference, advanced. On the bottom, the questionnaire's items. On the centre, the arrows describe how strong items are related with the corresponding latent variable (the higher the value, the stronger the relationship).

6. Conclusions

The work in this paper stems from the 3P project for the empowerment of primary and middle schools with physical computing, carried out in Italy. The curriculum is standardised across all primary and middle schools in Italy, ensuring that learners had a common baseline of educational experiences. 3P offered open courses and material for teachers to carry out interventions in their classes. The basic intervention starts with simple problematic scenarios for children, a resolution pattern and programs. It continues and challenges to maintain the same pattern, and modify it by experimenting with diverse physical input and output devices. Before tackling more complex challenges, a formative assessment step is recommended and a specific instrument was missing, in line with the critical review by Liu et al. which shows that most assessment methods tend "to regard CT as a learning outcome", and not as a process to evaluate step by step (Liu et al., 2021). The paper thus reports on the creation of the PCBUQ questionnaire for it, for investigating children's understanding of the basics of physical computing, to administer before children tackle more complex challenges. Specifically, the questionnaire was designed to be: (1) inclusive, that is, clear, suitable and adequate for diverse children attending mainstream schools (e.g., with textual-pictorial items), (2) rapid to administer and process in time-constrained class hours by teachers on their own, to obtain formative feedback on their interventions.

The questionnaire development process, detailed in the paper, served to refine items and classify them into understanding-related levels of the Bloom's taxonomy for non-computing experts, shown in Table 1. The final statistical analyses served to assess it, as shown in Table 2, and suggested to drop Item 5 and slightly revise Item 7 in terms of phrasing. In brief, the reported analyses show that PCBUQ can be considered a satisfactorily reliable instrument capable of discriminating performances of learners of the basics of physical computing in terms of classification, interpretation and inference.

The reported work indicates that similar formative questionnaires can be analogously designed for different intervention curricula to be **inclusive** and **rapid** for teachers to administer, keeping reliability at an acceptable level. The advantage of such questionnaires with respect to more advanced methods (e.g., eye-tracking, think-aloud) is that they could be used by teachers, independently of researchers or experts in assessment methods, to understand those who are most likely to be in need of further scaffolding.

In particular, the external teacher, who followed interventions in schools and the administration of the questionnaire, noted that the difference in difficulties of what were envisioned as basic items (1, 2) and advanced items (6, 7, 8) in Table 1 could be explained with the variability in length of the interventions, besides teachers' confidence in their knowledge. In particular, the teacher observed that not all schools managed to spend time on the exploration of analogical devices. This factor may have affected the results of the questionnaire in relation to those items that tackle their understanding, besides the different way of phrasing Item 7. A suggestion for future versions of this and other similar questionnaires is to clearly differentiate the items that require digital and analogical devices, use the same language structure for them, and study the type of devices (analogical versus digital) as a criterion for clustering items into difficulty levels.

A limitation of the reported work was that the last three items of the questionnaire were designed by referring to a block-based environment that teachers had used in class. On the one hand, they refer to basic control-flow structures (loops and conditionals), which are present in other programming environments for physical computing for children; these items could be adapted accordingly, still paying attention to differences in devices as recommended by a reviewer, e.g., the available input and output devices.

Furthermore, the intermediate and the advanced-level items of the questionnaire refer to a problematic scenario to interpret for inferring solutions. This was consistent with the overall goal of the interventions reported in the paper, which was to empower learners to design their own smart solutions for every-day's scenarios with physical computing, and not to scaffold future computer-science coursework (Weintrop & Wilensky, 2019). On the other hand, the questionnaire was designed and tested with young learners, who had started learning of physical computing, and then designed their solutions for problematic scenarios with it. Moreover, as commented by a reviewer, other interventions at school not only require to design an artefact for a problematic scenario, like the interventions reported in the paper, but they also target electronics knowledge, e.g., the usage of breadboards. This is not addressed by the questionnaire presented in the paper.

Thereby, it remains open the issue of investigating the design of similar formative instruments for more experienced learners and advanced physical-computing interventions, e.g., addressing more advanced electronics knowledge (e.g., of breadboards), or computing knowledge (e.g., of data structures).

Selection and participation

The questionnaire and overall data processing were submitted to the privacy and ethical boards of the leading university. The approval was received in Fall 2022. Moreover, the questionnaire was administered

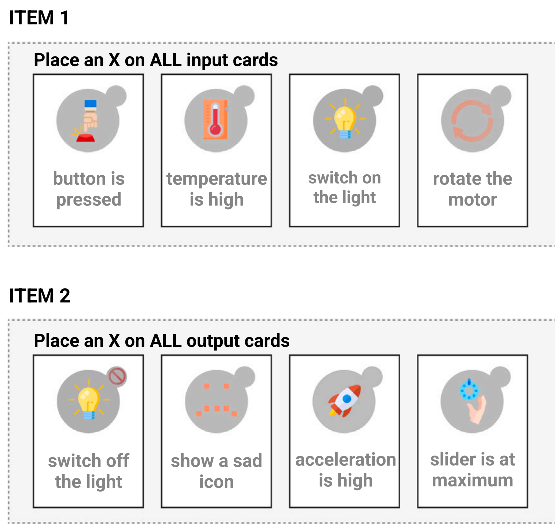

Fig. A.11. Item 1 and 2 of the final questionnaire.

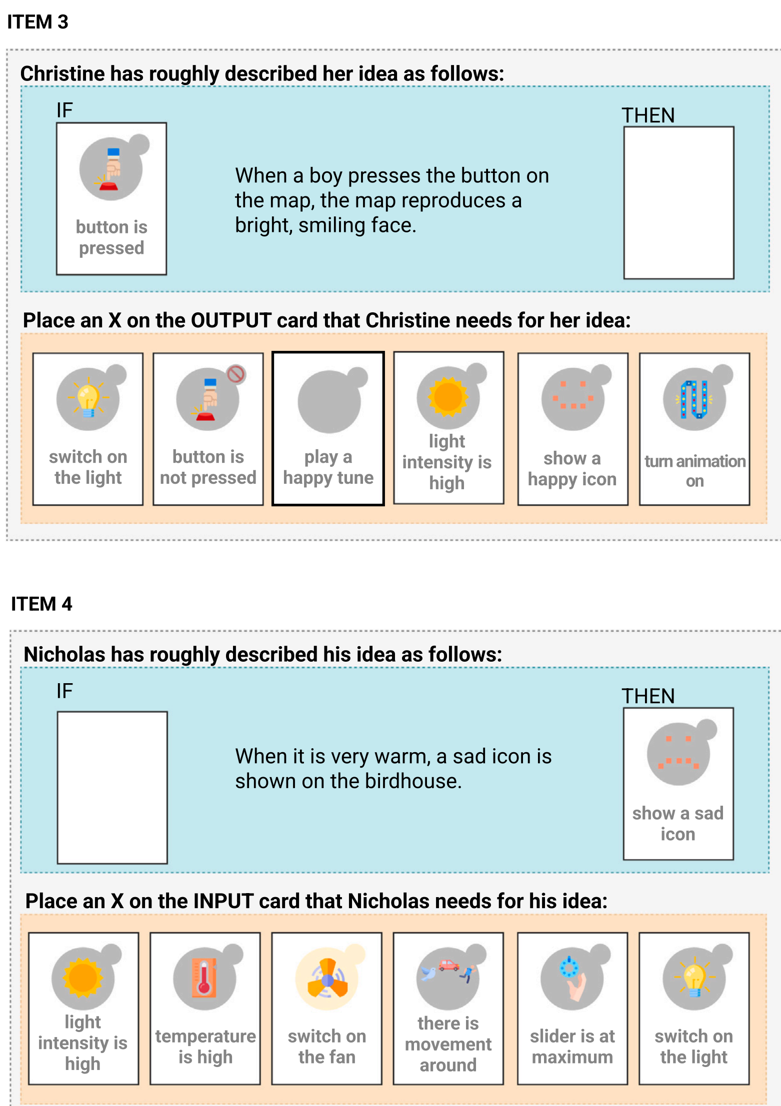

Fig. A.12. Item 3 and 4 of the final questionnaire.

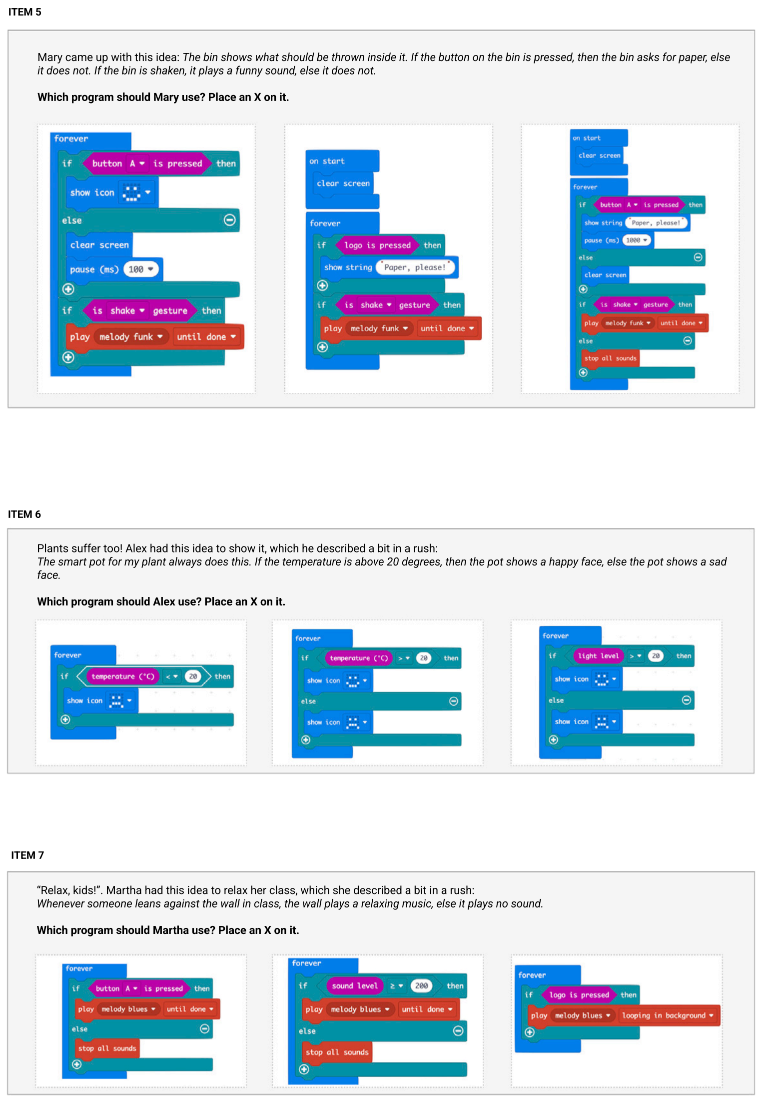

Fig. A.13. Item 5, 6 and 7 of the final questionnaire.

by teachers, with an external teacher, as part of their class curricula. Parents were informed by schools and their approval sought. Data were always anonymously collected by teachers and safely processed by researchers in line with national, European regulations and the guidelines of the two boards.

CRediT authorship contribution statement

Andrea Bonani: Recruitement, Data curation, Validation. **Rosella Gennari:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Data curation, Conceptualization. **Alessandra Melonio:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Pierpaolo Vittorini:** Writing – review & editing, Writing – original draft, Validation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially supported by the project "TORUS: Toolkits for Transdisciplinary Human-Centred Design", funded under the Interdisciplinary Research grant of the Free University of Bozen-Bolzano.

We sincerely thank all schools that participated in the work reported in the paper: the primary schools A. Langer, M.L. King, and Don Bosco from Bolzano; the middle schools L. Da Vinci, Archimede, Alfieri and Alpi from Bolzano, as well as Segantini from Merano.Their invaluable contributions and collaboration were essential to the success of this research.

Appendix. The final version of the questionnaire

Figs. A.11, A.12, A.13 shows the final version of the questionnaire. The final version of the questionnaire is made available online at ⟨https://vittorini.univaq.it/iotgoquest/⟩.

Data availability

The data that has been used is confidential.

References

Anderson, L., Krathwohl, D., & Bloom, B. (2001). A taxonomy for learning, teaching, and assessing : a revision of bloom's taxonomy of educational objectives. In *A taxonomy for learning, teaching, and assessing : A revision of bloom's taxonomy of educational objectives*.

Angeli, Charoula, & Giannakos, Michail (2020). Computational thinking education: Issues and challenges. *Computers in Human Behavior*, [ISSN: 0747-5632] *105*, Article 106185. http://dx.doi.org/10.1016/j.chb.2019.106185, URL https://www.sciencedirect.com/science/article/pii/S0747563219303978.

Austin, Jonny, Baker, Howard, Ball, Thomas, Devine, James, Finney, Joe, Halleux, Peli De, et al. (2020). The BBC Micro:bit: from the u.k. to the world. *Communications of the ACM*, [ISSN: 0001-0782] *63*(3), 62–69. http://dx.doi.org/10.1145/3368856.

Babazadeh, Masiar, & Negrini, Lucio (2022). How is computational thinking assessed in European k-12 education? a systematic review. *International Journal of Computer Science Education in Schools*, *5*, 3–19. http://dx.doi.org/10.21585/ijcses.v5i4.138.

Ball, Thomas, Chatra, Abhijith, Halleux, Peli de, Hodges, Steve, Moskal, Michał, & Russell, Jacqueline (2019). Microsoft makecode: Embedded programming for education, in blocks and typescript. In *Proceedings of the 2019 ACM SIGPLAN symposium on SPLASH-E* (pp. 7–12). New York, NY, USA: Association for Computing Machinery, ISBN: 9781450369893, http://dx.doi.org/10.1145/3358711.3361630.

Barr, David, Harrison, John, & Conery, Leslie (2011). Computational thinking: A digital age skill for everyone. *Learning and Leading with Technology*, *38*, 20–23, URL https://api.semanticscholar.org/CorpusID:53794701.

BBC (2024). Micro:bit. URL https://microbit.org/. (Retrieved January 2024).

Bieraugel, Mark, & Neill, Stern (2017). Ascending bloom's pyramid: Fostering student creativity and innovation in academic library spaces. *College & Research Libraries*, [ISSN: 2150-6701] *78*(1), 35. http://dx.doi.org/10.5860/crl.78.1.35, URL https://crl.acrl.org/index.php/crl/article/view/16566.

Bjørn, Pernille, Menendez-Blanco, Maria, & Borsotti, Valeria (2022). *Diversity in computer science*. New York, NY, USA: Springer, http://dx.doi.org/10.1007/978-3-031-13314-5, ISBN 9783031133145 9783031133145. URL https://library.oapen.org/handle/20.500.12657/59372.

Blikstein, Paulo (2013). Gears of our childhood: Constructionist toolkits, robotics, and physical computing, past and future. In *Proceedings of the 12th international conference on interaction design and children* (pp. 173–182). New York, NY, USA: Association for Computing Machinery, ISBN: 9781450319188, http://dx.doi.org/10.1145/2485760.2485786.

Bloom, B., Engelhart, M., Furst, E., Hill, W., & Krathwohl, D. (1964). *Taxonomy of educational objectives*. Green New York: Longmans.

Bonani, Andrea, Gennari, Rosella, Melonio, Alessandra, & Rizvi, Mehdi (2022). Design and computational thinking with IoTgo: What teachers think. In *Lecture notes in networks and systems*: *vol. 580*, *Methodologies and intelligent systems for technology enhanced learning, 12th international conference, MIS4TEL* (pp. 165–174). Springer, http://dx.doi.org/10.1007/978-3-031-20617-7_21.

Bowler, Leanne, & Champagne, Ryan (2016). Mindful makers: Question prompts to help guide Young peoples' critical technical practices in maker spaces in libraries, museums, and community-based youth organizations. *Library & Information Science Research*, [ISSN: 0740-8188] *38*(2), 117–124. http://dx.doi.org/10.1016/j.lisr.2016.04.006, URL https://www.sciencedirect.com/science/article/pii/S0740818815300840.

Brackmann, Christian P., Román-González, Marcos, Robles, Gregorio, Moreno-León, Jesús, Casali, Ana, & Barone, Dante (2017). Development of computational thinking skills through unplugged activities in primary school. In *Proceedings of the 12th workshop on primary and secondary computing education* (pp. 65–72). New York, NY, USA: Association for Computing Machinery, ISBN: 9781450354288, http://dx.doi.org/10.1145/3137065.3137069.

Cápay, Martin, & Klimová, Nika (2019). Engage your students via physical computing!. In *2019 IEEE global engineering education conference* (pp. 1216–1223). http://dx.doi.org/10.1109/EDUCON.2019.8725101.

Chiazzese, Giuseppe, Arrigo, Marco, Chifari, Antonella, Lonati, Violetta, & Tosto, Crispino (2019). Educational robotics in primary school: Measuring the development of computational thinking skills with the bebras tasks. *Informatics*, [ISSN: 2227-9709] *6*(4), http://dx.doi.org/10.3390/informatics6040043, URL https://www.mdpi.com/2227-9709/6/4/43.

Cronbach, Lee J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika.*, [ISSN: 0033-3123] *16*(3), 297–334. http://dx.doi.org/10.1007/BF02310555, URL http://link.springer.com/10.1007/BF02310555.

Denning, Peter J., & Tedre, Matti (2019). *Computational thinking*. The MIT Press, ISBN: 9780262536561.

di Mascio, T., Melonio, A., Tarantino, L., & Vittorini, P. (2014). Many children and short project timing: How TERENCE harmonized these conflicting requirements. *vol. 292*, In *Advances in intelligent systems and computing*. http://dx.doi.org/10.1007/978-3-319-07698-0_9.

Dindler, Christian, Smith, Rachel, & Iversen, Ole (2020). Computational empowerment: Participatory design in education. *CoDesign*, *16*, 1–15. http://dx.doi.org/10.1080/15710882.2020.1722173.

Fabrigar, Leandre R., MacCallum, Robert C., Wegener, Duane T., & Strahan, Erin J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, [ISSN: 1082989X] *4*(3), 272–299. http://dx.doi.org/10.1037/1082-989X.4.3.272.

Galadima, Ahmad Adamu (2014). Arduino as a learning tool. In *2014 11th international conference on electronics, computer and computation* (pp. 1–4). http://dx.doi.org/10.1109/ICECCO.2014.6997577.

Gennari, Rosella, Matera, Maristella, Melonio, Alessandra, & Rizvi, Mehdi (2024). How to enable Young teens to design responsibly. *Future Generation Computer Systems*, [ISSN: 0167-739X] *150*, 303–316. http://dx.doi.org/10.1016/j.future.2023.09.004, URL https://www.sciencedirect.com/science/article/pii/S0167739X23003321.

Gennari, Rosella, Matera, Maristella, Melonio, Alessandra, Rizvi, Mehdi, & Roumelioti, Eftychia (2020). Reflection and awareness in the design process children ideating, programming and prototyping smart objects. *Multimedia Tools and Applications*, *80*(26–27), 34909–34932. http://dx.doi.org/10.1007/s11042-020-09927-x.

Gennari, Rosella, Melonio, Alessandra, & Rizvi, Mehdi (2022). From children's ideas to prototypes for the internet of things: a case study of cross-generational end-user design. *Behaviour & Information Technology*, *41*(15), 3281–3300. http://dx.doi.org/10.1080/0144929X.2021.1979654.

Gennari, Rosella, Melonio, Alessandra, & Rizvi, Mehdi (2023). A tool for guiding teachers and their learners: the case study of an art class. In *Extended abstracts of the 2023 CHI conference on human factors in computing systems*. New York, NY, USA: ACM, ISBN: 9781450394222, http://dx.doi.org/10.1145/3544549.3573863.

Greenstein, L. (2012). Assessing 21st century skills: A guide to evaluating mastery and authentic learning. URL https://api.semanticscholar.org/CorpusID:142960718.

Grover, Shuchi (2020). Designing an assessment for introductory programming concepts in middle school computer science. In *Proceedings of the 51st ACM technical symposium on computer science education* (pp. 678–684). New York, NY, USA: Association for Computing Machinery, ISBN: 9781450367936, http://dx.doi.org/10.1145/3328778.3366896.

Hambleton, Ronald K., & Swaminathan, Hariharan (1985). *Item response theory: Principles and applications*. Springer Netherlands, http://dx.doi.org/10.1007/978-94-017-1988-9.

Harter, Susan, & Pike, Robin (1984). The pictorial scale of perceived competence and social acceptance for Young children. *Child Development*, *55*(6), 1969–1982, ISSN 00093920 14678624. URL http://www.jstor.org/stable/1129772.

Hayat Alajlan, Reem Alebaikan, & Almassaad, Ahmad (2023). Computational thinking in K–12 computer education: Appropriate pedagogy. *Technology, Pedagogy and Education*, *32*(3), 337–349. http://dx.doi.org/10.1080/1475939X.2023.2184857.

Hodges, Steve, Sentance, Sue, Finney, Joe, & Ball, Thomas (2020). Physical computing: A key element of modern computer science education. *Computer*, *53*(4), 20–30.

Iversen, Ole Sejer, Smith, Rachel Charlotte, & Dindler, Christian (2018). From computational thinking to computational empowerment: A 21st century pd agenda. In *Proceedings of the 15th participatory design conference: Full papers - volume 1*. New York, NY, USA: Association for Computing Machinery, ISBN: 9781450363716, http://dx.doi.org/10.1145/3210586.3210592.

Iwata, Megumi, Pitkänen, Kati, Laru, Jari, & Mäkitalo, Kati (2020). Exploring potentials and challenges to develop twenty-first century skills and computational thinking in k-12 maker education. *Frontiers in Education*, [ISSN: 2504-284X] http://dx.doi.org/10.3389/feduc.2020.00087, URL https://www.frontiersin.org/articles/10.3389/feduc.2020.00087.

Katterfeldt, Eva-Sophie, Dittert, Nadine, & Schelhowe, Heidi (2015). Designing digital fabrication learning environments for bildung: Implications from ten years of physical computing workshops. *International Journal of Child-Computer Interaction*, *5*, 3–10.

Kim, Min, & Min, Sun (2021). Developing children s computational thinking through physical computing lessons. *lnternational Electronic Journal of Elementary Education*, *13*, 183–198. http://dx.doi.org/10.26822/iejee.2021.183.

Kong, Siu-Cheung, Abelson, Harold, & Kwok, Wai-Ying (2022). Introduction to computational thinking education in K–12. In *Computational thinking education in k–12: Artificial intelligence literacy and physical computing*. The MIT Press, ISBN: 9780262368971, http://dx.doi.org/10.7551/mitpress/13375.003.0002.

Lechelt, Susan, Rogers, Yvonne, & Marquardt, Nicolai (2020). Coming to your senses: Promoting critical thinking about sensors through playful interaction in classrooms. In *Proceedings of the interaction design and children conference* (pp. 11–22). New York, NY, USA: Association for Computing Machinery, ISBN: 9781450379816, http://dx.doi.org/10.1145/3392063.3394401.

Lego (2024). Lego mindstorms rc. URL https://www.lego.com/en-it/themes/mindstorms. (Retrieved January 2024).

Liu, Zhichun (2023). Integrating computational thinking into k-12 education: Bridging the gap between theories and practices: Bridging the gap between theories and practices. *STEM Education Review*, *1*, http://dx.doi.org/10.54844/stemer.2023.0467.

Liu, Ruohan, Luo, Feiya, & Israel, Maya (2021). What do we know about assessing computational thinking? A new methodological perspective from the literature. In *Proceedings of the 26th ACM conference on innovation and technology in computer science education v. 1* (pp. 269–275). New York, NY, USA: Association for Computing Machinery, ISBN: 9781450382144, http://dx.doi.org/10.1145/3430665.3456380.

Lodi, Michael, & Martini, Simone (2021). Computational thinking, between papert and wing. *Science & Education*, *30*(4), 883–908. http://dx.doi.org/10.1007/s11191-021-00202-5.

McMillan, Sara, King, Michelle, & Tully, Mary (2016). How to use the nominal group and delphi techniques. *International Journal of Clinical Pharmacy*, *38*, http://dx.doi.org/10.1007/s11096-016-0257.

Microsoft (2024). URL https://studio.code.org/s/csd6-2022. (Retrieved January 2024).

Miliou, Ourania, Adamou, Maria, Mavri, Aekaterini, & Ioannou, Andri (2023). An exploratory case study of the use of a digital self-assessment tool of 21st-century skills in makerspace contexts. *Educational Technology Research and Development*, http://dx.doi.org/10.1007/s11423-023-10314-0.

Neutens, Tom, & wyffels, Francis (2016). Teacher professional development through a physical computing workshop. In *Proceedings of the 11th workshop in primary and secondary computing education* (pp. 108–109). New York, NY, USA: Association for Computing Machinery, ISBN: 9781450342230, http://dx.doi.org/10.1145/2978249.2978270.

Nunnally, Jum C., & Bernstein, Ira H. (1994). *Psychometric theory*. McGraw-Hill.

Papert, S., & Harel, I. Situating constructionism, chapter costructionism (chap. 1). Ablex Publishing Corporation.

Philip Chalmers, R. (2012). MIRT: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. http://dx.doi.org/10.18637/jss.v048.i06.

Przybylla, Mareen, & Romeike, Ralf (2014). Physical computing and its scope—Towards a constructionist computer science curriculum with physical computing. *Informatics in Education*, [ISSN: 1648-5831] *13*(2), 225–240. http://dx.doi.org/10.15388/infedu.2014.14.

Psycharis, Sarantos, Kalovrektis, Konstantinos, Sakellaridi, Eva, Konstantinos, Korres, & Mastorodimos, Dimitrios (2018). Unfolding the curriculum: Physical computing, computational thinking and computational experiment in STEM's transdisciplinary approach. *European Journal of Engineering and Technology Research*, 19–24. http://dx.doi.org/10.24018/ejeng.2018.0.CIE.639.

Raspberry Pi Foundation (2024). Raspberry pi projects. URL https://projects.raspberrypi.org/en/projects/physical-computing. (Retrieved January 2024).

Revelle, William (2022). Psych: Procedures for psychological, psychometric, and personality research. URL https://CRAN.R-project.org/package=psych.

Román-González, Marcos, Moreno-León, Jesús, & Robles, Gregorio (2017). Complementary tools for computational thinking assessment.

Rosseel, Yves (2012). Lavaan: An r package for structural equation modeling. *Journal of Statistical Software*, [ISSN: 1548-7660] *48*, 1–36. http://dx.doi.org/10.18637/JSS.V048.I02.

Sáez-López, José-Manuel, Román-González, Marcos, & Vázquez-Cano, Esteban (2016). Visual programming languages integrated across the curriculum in elementary school: A two year case study using scratch in five schools. *Computers & Education*, [ISSN: 0360-1315] *97*, 129–141. http://dx.doi.org/10.1016/j.compedu.2016.03.003, URL https://www.sciencedirect.com/science/article/pii/S0360131516300549.

Stager, Gary S., & Martinez, Sylvia (2013). Invent to learn: Making, tinkering, and engineering in the classroom. URL https://api.semanticscholar.org/CorpusID:114798877.

Thompson, Bruce (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association, http://dx.doi.org/10.1037/10694-000.

Thompson, Errol, Luxton-Reilly, Andrew, Whalley, Jacqueline L., Hu, Minjie, & Robbins, Phil (2008). Bloom's taxonomy for CS assessment. In *Proceedings of the tenth conference on australasian computing education–volume 78* (pp. 155–161). AUS, Australian Computer Society, Inc., ISBN: 9781920682590.

Van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2007). Fully conditional specification in multivariate imputation. [ISSN: 00949655] 76(12), 1049–1064. http://dx.doi.org/10.1080/10629360600810434. URL https://www.tandfonline.com/doi/abs/10.1080/10629360600810434.

Wang, Ge, Zhao, Jun, Kleek, Max Van, & Shadbolt, Nigel (2023). 12 ways to empower: Designing for children's digital autonomy. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery., ISBN: 9781450394215, http://dx.doi.org/10.1145/3544548.3580935.

Weintrop, David, Beheshti, Elham, Horn, Michael, Orton, Kai, Jona, Kemi, Trouille, Laura, et al. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology*, *25*(1), 127–147. http://dx.doi.org/10.1007/s10956-015-9581-5.

Weintrop, David, & Wilensky, Uri (2019). Transitioning from introductory block-based and text-based environments to professional programming languages in high school computer science classrooms. *Computers & Education*, [ISSN: 0360-1315] *142*, Article 103646. http://dx.doi.org/10.1016/j.compedu.2019.103646, URL https://www.sciencedirect.com/science/article/pii/S036013151930199X.

Wu, Sheng-Yi, & Su, Yu-Sheng (2021). Visual programming environments and computational thinking performance of fifth- and sixth-grade students. *Journal of Educational Computing Research*, *59*(6), 1075–1092.