Gender Differences in Story, Game and Visual Adventures in Hedy

Shirley de Wit
s.dewit@tudelft.nl
Delft University of Technology
The Netherlands

Felienne Hermans
f.f.j.hermans@vu.nl
Vrije Universiteit Amsterdam
The Netherlands

Marcus Specht
m.m.specht@tudelft.nl
Delft University of Technology
The Netherlands

Efthimia Aivaloglou
e.aivaloglou@tudelft.nl
Delft University of Technology
The Netherlands

Abstract

Even though the field of Computer Science (CS) affects different aspects of society, several groups of society are underrepresented, including women and nonbinary people. Children might have different learning opportunities in CS due to their project preferences. Girls are likelier to work on stories and simple programs in Scratch, whereas boys tend to create games and more complex programs. We explore whether preferences and program implementation differ between genders within a story, game and visual adventure in Hedy, a gradual textual programming language. We analysed 14,233 programs within five Hedy levels created by 2,819 users who turned 10 to 14 in 2023. We found that boys, girls and nonbinary children worked most on the game adventure. Within the individual adventures, gender differences occur in all three adventures in the most elaborate Hedy level analysed. However, for some levels, no gender differences were found. Thus, programming assignments can be created in which children of different genders work on similar programs in terms of size and number of (unique) commands used.

CCS Concepts: • **Social and professional topics** → **Computer science education**; **Gender**.

Keywords: Hedy, programming, education, gender

ACM Reference Format:
Shirley de Wit, Felienne Hermans, Marcus Specht, and Efthimia Aivaloglou. 2024. Gender Differences in Story, Game and Visual Adventures in Hedy. In *Proceedings of the 2024 ACM SIGPLAN International Symposium on SPLASH-E (SPLASH-E '24), October 24, 2024, Pasadena, CA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3689493.3689983

SPLASH-E '24, October 24, 2024, Pasadena, CA, USA

ACM ISBN 979-8-4007-1216-6/24/10
https://doi.org/10.1145/3689493.3689983

1 Introduction

The field of Computer Science (CS) affects different aspects of society, including education [43], healthcare [23, 24], and transport [1]. However, not everyone has equitable access to contribute to CS [36], including women and nonbinary people [11, 32, 38]. This implies that they are disproportionately excluded from a field with fast-growing employment [40]. Due to this lack of diversity, we miss out on innovations [19]. Moreover, it results in biased technology and systems [25, 45].

The lack of women in CS starts at an early age with girls opting out of CS due to, among others, stereotypes [34], low self-efficacy [3, 48] and lack of interest development [22, 48]. These reasons might differ for nonbinary people [32]. They are likely confident in their CS abilities but can have a low sense of belonging [41]. One way to attract more girls and nonbinary children to the field is to ensure that educational CS activities appeal to them. For example, female students are likelier to prefer educational activities involving people over things [8, 33]. Moreover, girls often prefer to create stories, while boys work more often on games [13, 15].

The type of programs children work on seems to influence their program implementation. Children who work on stories in Scratch, a programming language with a block-based editor, implement fewer programming concepts and create less complex programs than children who work on games [15]. Findings from different studies [5, 13, 15] imply that (gender) preferences in various types of programs or assignments can result in different learning opportunities.

Studies related to gender and program type are mainly done in Scratch. However, less is known about other programming languages in which children can learn to program. One of these languages is Hedy, a gradual textual programming language designed to teach novices syntax [18]. Users work on various assignments, called adventures, while syntax and commands build up throughout the different levels.

In this study, we explore whether children of different genders choose to work on different types of adventures in Hedy. Moreover, we are interested in gender differences in

program size and commands used and to what extent these findings occur within different adventures. Additionally, we extend current work by including nonbinary children. With our study, we contribute to the knowledge of how to create equitable programming experiences for all children, independent of their gender. We do so by answering the following research questions:

RQ1. What are the gender differences in the type of adventures children work on in Hedy?

RQ2. How do program length and the number of commands used in Hedy differ between genders?

RQ3. To what extent do the differences in program length and the number of commands used between genders occur within different types of adventure in Hedy?

To answer our research questions, we analyse 14,233 programs of 2,819 users active in Hedy between May 2023 and May 2024. These programs are saved within a selected story, game and visual adventure. We explore whether children worked on an adventure and how many programs they saved. We also analyse the programs created by looking at their length and the number of (unique) commands used. For each analysis, we compare the findings between genders.

2 Background

2.1 Programming Languages for Young Novices

A variety of languages is used to teach young novices to program; these include (but are not limited to) Scratch, Java and Python [35]. Since previous research on gender differences in programming for children is often done in Scratch, and we will do so in Hedy, we elaborate on these two.

Scratch[1] is a block-based programming environment which aims to introduce children with no prior experience to programming [31]. Scratch is designed for 8 to 16-year-olds [31] from economically disadvantaged and culturally diverse communities [30]. Although Scratch is designed for informal education, it is also used in schools [31]. Scratch builds upon constructionism [31], which is "an educational theory developed by Seymour Papert and emphasizes students' hands-on, experiential learning through building, creating, and sharing artefacts with peers" [7]. This is reflected in Scratch's emphasis on exploring and sharing projects with other users [31]. Scratch is used in many studies [35, 39] covering a variety of topics including learning computer science and coding [29, 35, 37], self-efficacy [3, 6], and computational thinking [12, 26].

Hedy[2] is a gradual textual programming language designed to teach syntax to novices (age 10 and up) in the classroom [17, 18]. The design of Hedy is "based on how punctuation is taught to novice readers in natural language education" [18]. Currently, learners can work through 18 levels. They start programming with a limited number of commands and no syntactic element, such as brackets, colons or indentations. The rules and available commands "slowly and gradually change until the novices are programming in Python" [18]. The users go through the levels at their own pace while working on different adventures, enabling them to create simple yet fun and meaningful programs [18]. An adventure includes one or more cohesive assignments, such as creating a rock, paper, scissors game or a story. Some of these adventures reoccur at multiple levels. Since Hedy is a relatively new programming language, a small number of studies researched the (use of) the language, covering learner experiences [14], learning to program [27], and behavioural intention to use Hedy [47].

One of the differences between Scratch and Hedy is that Scratch is a programming language with a block-based editor, while Hedy is a textual programming language. A meta-analysis of 13 studies that compared these types of languages did not find a significant overall effect size on the cognitive and affective learning outcomes [46]. A recent study by Zdawczyk and Varma [48] investigated the effects of Scratch and Python on the beliefs and attitudes of upper elementary and middle school children. Their findings include children having a higher self-efficacy in Scratch than in Python. Moreover, more girls than boys preferred using Scratch over Python. On the other hand, boys had a higher sense of belonging, more interest, and a preference for Python over Scratch.

2.2 Gender Differences in Type of Programs

Previous work found (binary) gender differences in students' preferences for CS assignments and projects. For example, female high school and university students are likelier to prefer CS assignments with people over assignments with things [8, 33]. Male students, on the other hand, have either no preference or a slight preference towards assignments with people.

By analysing 127 Scratch projects, Funke et al. [13] found that students (ages 9-10) mostly create stories, animations and games. The majority of the stories were created by girls, while most games were made by boys [13]. Animations were made equally by boys and girls. Graßl et al. [15] expanded on this work and analysed 317 Scratch programs by children aged 8-10. The programs created by girls contained more blocks typically used to create stories and animations, while the programs created by boys contained more blocks common for making games.

2.3 Gender Differences in Program Implementation

According to the literature review by Bati [4], girls in their early childhood can perform well in computational thinking and programming given an appropriate educational setting. However, some studies do show gender differences in programming performance. Although there were no gender

[1] https://scratch.mit.edu
[2] https://hedy.org

differences between 4 to 7-year-old students when doing robotics and simple programming tasks [42], boys did perform better in the advanced tasks. Similarly, girls aged 10 and up tend to perform better in relatively easier tasks [21].

Within Scratch, boys create more complex programs [15]. Moreover, boys apply more programming concepts such as conditional statements or iterations [15] and use more different blocks than girls [5, 13]. There is also a difference in the type of blocks used in Scratch: girls use more Look blocks, while boys use more Motion blocks [13]. Moreover, programs using keyboard controls are made mainly by boys [13].

The differences in the type of blocks and concepts used can relate to the type of program children work on [13, 15]. Graßl et al. [15] explored the topics children work on in combination with the type of program and program implementation. Programs on topics popular among girls, such as dancing or unicorns, were mostly animations and stories. Within these topics, boys and girls created programs of similar size and complexity. Moreover, girls were likelier to work on animation and story-typed projects independent of the topic of the program. Within these types of projects, the programs implemented were less complex and contained fewer programming concepts. In line with this, Bentz and Standl [5] explored topics and program implementation in Scratch projects. Next to analysing gender differences, they surveyed children (aged 12-14) on their interests in people and human surroundings and in things and spatial arrangements. Students with an interest in people focused on dialogues and sequences in their programs. In contrast, children with an interest in things focused more on movements and controls. This implies that personal preferences influence what programming concepts children use. However, we would like all children to work on projects they prefer while still practising similar programming concepts [15].

3 Method

3.1 Participants

Our participants are users registered in Hedy who saved at least one program between May 1st 2023 and April 30th 2024 in one of the adventures described in Section 3.2. We focused on users in childhood (which is till the age of 14 [44]). Moreover, Hedy is designed for users aged 10 and up. Therefore, we focused on users between the ages of 10 and 14. However, users self-report only their year of birth, meaning that we don't know the exact age of the user when they save a program. We decided to include all users who turned 10 to 14 in 2023, so users born between 2009 and 2013. Moreover, we only included users who answered the optional question about their gender identity. They could self-report their gender as female, male, or other. We refer to these groups as girls, boys and nonbinary children. These criteria resulted in 2,819 users. The distribution of self-reported birth year and gender is shown in Figure 1.

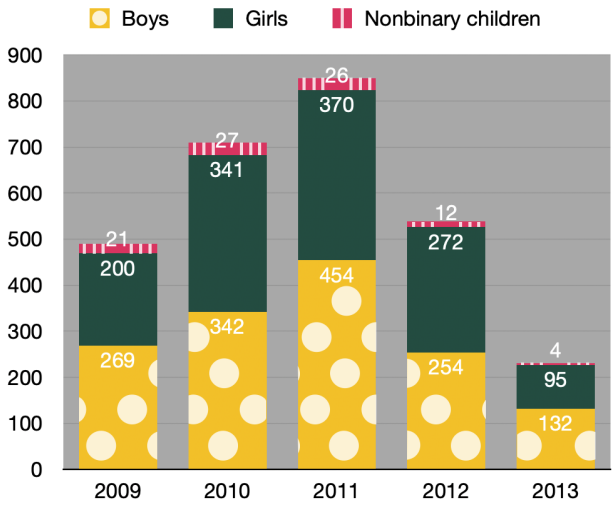


Figure 1. Distribution of birth year of users, including the number of boys (n = 1,451), girls (n = 1,278) and nonbinary children (n=90)

3.2 Materials

We selected Hedy adventures to represent the story, games and visual program type. We decided on these categories since previous work in Scratch showed that children mostly work on stories, animations and games [13]. Animations can not be created in Hedy in a similar way as in Scratch. However, it is possible to work on adventures in Hedy with a visual output instead of a textual output. We looked for representative adventures with story, game and visual aspects, which all re-occur in multiple Hedy levels. We selected the following adventures for our analysis:

- Story: Story adventure
  In this adventure, users create stories. They are encouraged to write a story about a main character and any topic. In levels 2 to 5, users are asked to copy their story from the previous level and expand on it.
- Game: Rock adventure
  In this adventure, users build a rock, paper, scissors game. They ask for user input, make the computer pick rock, paper or scissors, and determine whether there is a tie. In level 2, there are two assignments related to making the game. We included the one assignment that has rock as adventure name.
- Visual: Turtle adventure
  In this adventure, users create drawings. They are encouraged to draw various figures (square, staircase, triangle, circle) and random figures. Users also learn to use different colours.

Since the selected adventures occur in each of the first five levels of Hedy, we decided to focus on programs within these five levels and their commands as described in Table 1. Moreover, Hedy is an open-source project that is still under

Table 1. Hedy commands in level 1 to 5 [16]

Level	Available commands
1	print, ask, forward, turn, echo
2	print, ask, forward, turn, is, sleep
3	print, ask, forward, turn, is, sleep, at random, add to, remove from
4	print, ask, forward, turn, is, sleep, at random, add to, remove from
5	print, ask, forward, turn, is, sleep, at random, add to, remove from, if, else, in

development. We tried to scope our research so that the number of changes made (based on the git history) to the selected adventures was as minimal as possible. Focusing on the first five levels helped in this. It also motivated our decision to analyse programs saved between May 1st 2023 and April 30th 2024.

Besides the adventures, we also needed to decide which languages to include. Hedy is currently translated in 47 languages [17]. However, some translations are more complete than others. We only looked at programs created in the four languages that have more than 85% of the text translated: Chinese (simplified), Dutch, English and Spanish.

Thus, all programs saved between May 1st 2023 and April 30th 2024 in levels 1 to 5, in either Chinese (simplified), Dutch, English or Spanish, for the story, game or visual adventure are included. These criteria resulted in 14,233 programs, of which 7,155 were created by boys, 6,581 by girls and 497 by nonbinary children.

3.3 Measures and Analysis

When analysing the programs, we observed that 61 programs in the dataset are empty. We excluded these 61 programs from our analysis. Of the users in our dataset, 8 only saved one or more of these empty programs. We excluded these users from the analysis.

We analysed the programs using Python scripts[3] and the `pandas`, `SciPy` and `NumPy` libraries. For the statistical tests, we report on p-values smaller than 0.05. P-values smaller than 0.001 are reported as p<.001.

3.3.1 Adventures Children Work On. To answer our first research question, we analysed which adventures users worked on and the number of programs saved per adventure.

First, we analysed which adventures children *worked on at least once*. When a user saved at least one program in an adventure, we say this user worked on the adventure at least once. We reported the percentage of users who worked at least once on an adventure per gender. Since both the independent and dependent variables are categorical, we used the chi-square test of independence to determine whether the variables relate. Since the chi-square test works with frequencies, we counted (per gender) how many users worked and how many did not work on a specific adventure. We tested within each gender whether there is a relation between the type of adventure and (not) working on it. We also tested within each adventure whether gender relates to (not) working on that adventure. Moreover, we calculated Cramer's V to determine the effect size. Since the degrees of freedom is 2, a Cramer's V of 0.07 indicates a small, 0.21 a medium, and 0.35 a large effect size [28].

Secondly, we analysed how many *programs each user saved per adventure*. We reported the mean number of programs saved and its standard deviation per gender per adventure. We compared the number of saved programs within and between genders by using independent t-tests. Moreover, we used Cohen's d to calculate the effect size. A value of 0.2 indicates a small effect, 0.5 indicates a medium effect, and 0.8 indicates a large effect [9].

3.3.2 Program Implementation. To answer our second and third research questions, we analysed the length of the programs and the number of commands used. Since the available commands differ per level, we analysed the program implementations per level. We analysed the differences between genders per level (RQ2) and between genders per level per adventure (RQ3).

We analysed the *length of the programs*, as done by others [2, 15]. We reported the mean and standard deviation for the lines of code (LOC). We compared the results between genders using independent t-tests and calculated the effect size using Cohen's d.

Similar to previous work analysing elements in the programs of novices in Scratch [2, 5, 13, 15], we analysed the usage of commands described in Table 1. More specifically, we counted the *number of unique commands used* and the *number of total commands used*. However, commands can also occur in natural language. For example, 'print the dog is in the house' contains not only `print` but also `is` and `in` while only `print` is intended as a command. To eliminate these and similar issues, we used the semantics described within the Hedy GitHub [16]. Moreover, we tested our understanding of the semantics within the individual Hedy levels. We verified the outcomes of our Python scripts by manually counting the commands in 20 randomly selected programs, which covered all three adventures as well as each level. The Python scripts resulted in the same counts as the manual counting. We reported on the mean number of unique commands and the number of total commands used, including

[3]Python scripts can be found at https://shirleydewit.com/splashe2024

standard deviation. We compared these numbers between genders in each level and between genders within the individual adventures in each level using independent t-tests and calculated the effect size using Cohen's d.

4 Results

4.1 Adventures Children Work On

For each adventure, the majority of the users worked on it at least once, as shown in Table 2. The most popular adventure is the game adventure. We found a relation with medium effect size between adventures and working on them for boys (χ^2(2)=235.9, p<.001, V=0.23) and girls (χ^2(2)=190.9, p<.001, V=0.22). For nonbinary children, this relation has a small effect size (χ^2(2)=11.3, p=.003, V=0.20). We did not find a relation between genders in any of the adventures.

Table 2. Number of users per gender and percentage of those users that worked on the adventure at least once

Gender	n	Story	Game	Visual
All	2,811	61.1%	85.1%	76.5%
Boys	1,448	59.5%	83.9%	77.0%
Girls	1,273	62.8%	86.4%	76.4%
Nonbinary	90	62.2%	84.4%	71.1%

For the number of saved programs, the mean and standard deviation per gender for each adventure are shown in Table 3. We found that boys, girls and nonbinary children saved more games than stories. The significant difference between the number of saved games and stories has a small effect size for boys (t=10.62, p<.001, d=0.39), girls (t=8.77, p<.001, d=0.35) and nonbinary children (t=2.22, p=.028, d=0.33). Boys and girls also saved more games than visuals. The significant difference between the number of saved games and visuals has a negligible effect size for boys (t=3.66. p<.001, d=0.14), while this difference has a small effect size for girls (t= 5.25, p<.001, d=0.21). Furthermore, both boys and girls saved more visuals than stories. The significant difference between saved visuals and stories has a small effect size for boys (t=7.07, p=<.001, d=0.26) and a negligible effect size for girls (t=3.65, p<.001, d=0.14). When comparing between genders within each of the adventures, we only found that girls saved more stories than boys. This significant difference has a negligible effect size (t=2.50, p=.012, d=0.10).

Table 3. The mean ± standard deviation of saved programs per gender per adventure

Gender	Story	Game	Visual
All	1.4 ± 1.6	2.0 ± 1.6	1.7 ± 1.6
Boys	1.3 ± 1.5	1.9 ± 1.6	1.7 ± 1.6
Girls	1.5 ± 1.6	2.0 ± 1.7	1.7 ± 1.6
Nonbinary	1.5 ± 1.7	2.1 ± 1.7	1.9 ± 1.8

Thus, the game adventure is most popular among boys, girls and nonbinary children in our dataset. There are no significant differences with at least a small effect size between gender and working on an adventure or between gender and the number of programs saved.

4.2 Program Implementation

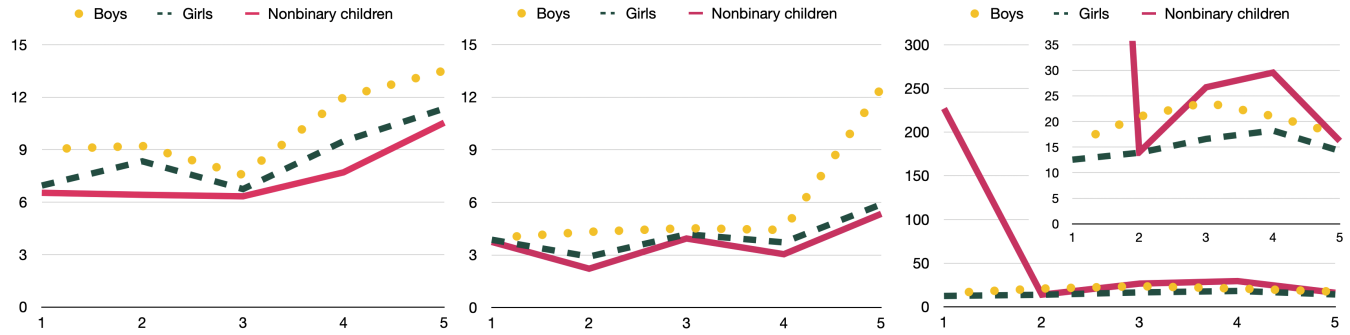
For each level, we analysed whether the LOC, unique number of commands and total number of commands used differ between genders. Moreover, we analysed whether there are differences between genders within each adventure in the LOC, unique commands and total commands (see also Figures 2, 3, and 4).

4.2.1 Level 1. The mean and standard deviation of the LOC, unique commands and total commands used for level 1 are displayed in Table 4. On average, nonbinary children create programs with a higher LOC than boys and girls. The significant difference between programs created by nonbinary children and boys has a small effect size (t=3.96, p<.001, d=0.32), as well as between nonbinary children and girls (t=3.89, p<.001, d=0.32). Moreover, boys created programs with more LOC than girls, but this difference has a negligible effect size (t=2.55, p=.011, d=0.07). Both boys and nonbinary children used more commands in total than girls. The significant difference between the number of commands used by boys and girls has a negligible effect size (t=3.88, p<.001, d=0.11), while the difference between the number of commands used by nonbinary children and girls has a small effect size (t=2.86, p=.004, d=0.23).

Table 4. The number of programs saved and their mean ± standard deviation of LOC, unique commands and total commands in level 1

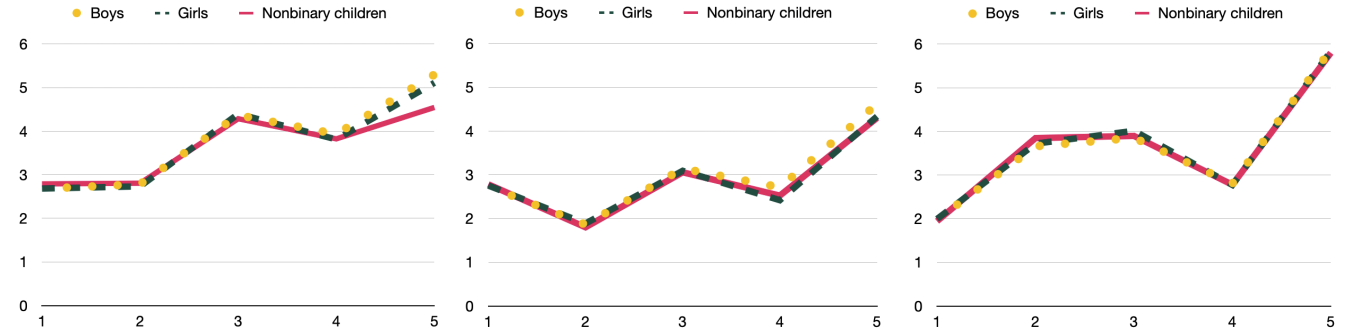
Gender	n	LOC	Unique	Total
Boys	2,537	9.4 ± 27.2	2.5 ± 0.6	8.3 ± 12.5
Girls	2,305	7.7 ± 19.4	2.5 ± 0.6	7.1 ± 8.0
Nonbinary	160	78.4 ± 873.0	2.5 ± 0.7	9.2 ± 15.6

Within both the story and game adventure, there are no significant differences between genders. For the visual adventure, programs created by boys and nonbinary children have more LOC than programs by girls. This significant difference between boys and girls has a negligible effect size (t=2.44, p=.015, d=0.12), while the difference has a medium effect size between nonbinary children and girls (t=3.99, p<.001, d=0.57). Moreover, the visual programs of nonbinary children have more LOC than those of boys, with the differences having a medium effect size (t=4.19, p<.001, d=0.60). For the total number of commands used, both boys and nonbinary children used more commands than girls. This difference between boys and girls has a small effect size (t=4.82, p<.001, d=0.24) and between nonbinary children and girls a medium effect size (t=3.77, p<.001, d=0.54).



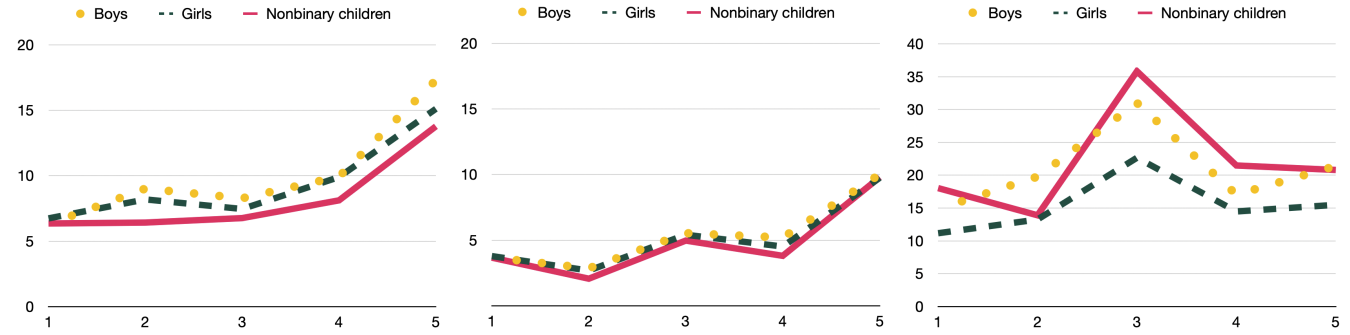
(a) Story adventure **(b)** Game adventure **(c)** Visual adventure

Figure 2. Mean LOC per level per gender for each adventure. Note the differences in the y-axis.



(a) Story adventure **(b)** Game adventure **(c)** Visual adventure

Figure 3. Mean number of unique commands used per level per gender for each adventure.



(a) Story adventure **(b)** Game adventure **(c)** Visual adventure

Figure 4. Mean number of total commands used per level per gender for each adventure. Note the differences in the y-axis.

4.2.2 Level 2. The mean and standard deviation of the LOC, unique commands and total commands used for level 2 are displayed in Table 5. We found that boys created programs with more LOC and that they used more commands in total than girls. However, both the significant difference in the number of LOC (t=2.47, p=.013, d=0.09) and command used (t=2.14, p=.033, d=0.08) have a negligible effect size.

Within all three adventures, there are no significant differences between genders for LOC, unique commands, and total number of commands in level 2.

Table 5. The number of programs saved and their mean ± standard deviation of LOC, unique commands and total commands in level 2

Gender	n	LOC	Unique	Total
Boys	1,637	11.4 ± 52.2	2.7 ± 1.0	10.3 ± 47.7
Girls	1,455	7.9 ± 10.9	2.7 ± 1.0	7.6 ± 10.4
Nonbinary	115	7.3 ± 13.9	2.8 ± 1.0	7.3 ± 14.0

4.2.3 Level 3. The mean and standard deviation of the LOC, unique commands and total commands used for level 3 are displayed in Table 6. There are no significant differences between genders for LOC, unique commands and total number of commands in level 3.

Table 6. The number of programs saved and their mean ± standard deviation of LOC, unique commands and total commands in level 3

Gender	n	LOC	Unique	Total
Boys	1,213	11.9 ± 35.8	3.7 ± 1.1	15.0 ± 51.7
Girls	1,071	9.0 ± 42.0	3.7 ± 1.1	11.7 ± 62.5
Nonbinary	85	12.3 ± 24.4	3.6 ± 1.2	15.9 ± 35.0

For the story and game adventures, there are no significant differences. For the visual adventure, girls used more unique commands than boys. This difference has a small effect size (t=3.26, p=.001, d=0.23).

4.2.4 Level 4. The mean and standard deviation of the LOC, unique commands and total commands used for level 4 are displayed in Table 7. Boys used more unique commands than girls. This significant difference has a negligible effect size (t=2.29, p=.022, d=0.10).

Table 7. The number of programs saved and their mean ± standard deviation of LOC, unique commands and total commands in level 4

Gender	n	LOC	Unique	Total
Boys	990	11.9 ± 29.2	3.1 ± 1.4	10.3 ± 17.1
Girls	966	9.9 ± 26.1	3.0 ± 1.4	9.2 ± 22.7
Nonbinary	65	15.3 ± 26.9	3.0 ± 1.4	12.3 ± 19.8

We did not find any significant differences within the story and visual adventures between genders. Within the game adventure, boys worked on programs with more LOC, and they used more unique commands than girls. The significant difference between the LOC in programs created by boys and girls has a negligible effect size (t=1.98, p=.048, d=0.14). The significant difference between the unique commands used between boys and girls has a small effect size (t=3.23, p=.001, d=0.24).

4.2.5 Level 5. The mean and standard deviation of the LOC, unique commands and total commands used for level 5 are displayed in Table 8. Boys used more unique commands and more commands in total than girls. Both the significant differences between number of unique commands (t=2.48, p=.013, d=0.13) and total commands (t=3.57, p<.001, d=0.18) have a negligible effect size.

Within the story adventure, boys use more unique commands than nonbinary children. This difference has a small effect size (t=2.07, p=.039, d=0.46). Within the game adventure, boys used more unique commands than girls. This difference has a small effect size (t=2.20, p=.028, d=0.20). Within the visual adventure, boys used more commands in total than girls. This difference has a small effect size (t=2.84, p=.005, d=0.25).

Table 8. The number of programs saved and their mean ± standard deviation of LOC, unique commands and total commands in level 5

Gender	n	LOC	Unique	Total
Boys	746	14.6 ± 59.4	5.3 ± 1.7	16.4 ± 19.6
Girls	765	10.4 ± 19.6	5.1 ± 1.6	13.4 ± 12.6
Nonbinary	62	10.7 ± 14.4	4.9 ± 2.0	14.8 ± 16.9

Thus, we found some significant differences between genders in the program implementation with at least a small effect size. We found that in level 1 nonbinary children create programs with more LOC than both boys and girls, and nonbinary children use more commands in total than girls. For the story adventure, there are no gender differences in levels 1 to 4. In level 5, boys use more unique commands than nonbinary children. For the game adventure, there are no gender differences in levels 1, 2 and 3. In levels 4 and 5, boys use more unique commands than girls. For the visual adventure, there are no gender differences in levels 2 and 4. In level 1, nonbinary children create longer programs than both boys and girls. Moreover, both boys and nonbinary children use more commands in total than girls. In level 3, girls use more unique commands than boys. In level 5, boys use more commands in total than girls.

5 Discussion

We were motivated to do this study because previous work in Scratch found (binary) gender differences in program implementation [13, 15]. These differences might be caused by differences in project preferences. Therefore, we explored gender differences in Hedy, focusing on different types of adventures and program implementation.

5.1 Reflection on the Results

Our results show that the game adventure is popular among boys, girls, and nonbinary children. This is in contrast with studies where stories were more popular among girls than games [13, 15]. Our findings not aligning with these studies might be explained by the fast-changing interests of children and our study including slightly older children, as also mentioned by Bentz and Standl [5]. Moreover, the type of game in Hedy (rock, paper, scissors) could also play a role in the popularity of the game adventure. So, educators who try to make educational CS activities that appeal to gender

minorities in CS should be careful about choosing stereotypical interests since they might not be valid for their group of students.

Throughout the levels in Hedy, we found some gender differences in the size of the programs and commands used. Within the first level, nonbinary children created much longer programs than both boys and girls. Since the difference in the LOC is large, we suspect that there is a user who created very large programs. Since fewer programs are created by nonbinary children, such an outlier has a relatively big impact on the results. Nevertheless, they also used more commands than girls. This might be an indication that although nonbinary people are also a gender minority in CS, they face different obstacles than women, which is in line with previous work [32, 41]. The findings in the other levels have negligible effect sizes. This might be because of the large sample size, especially for boys and girls, resulting in lower p-values even when the actual difference is not that big. Another aspect is that, in general, users created relatively small programs, resulting in small differences. We observed that for most of the gender differences between boys and girls, boys created larger programs and used more (different) commands. This is in line with previous work in which boys used more (unique) commands in Scratch [5, 13, 15]. These gender differences could be related to boys having a higher self-efficacy in programming [48] and being more willing to take risks [10].

For the individual adventures, we found significant differences in some but not all levels. For the story adventure, there are no significant differences between boys and girls in all of the levels. This aligns with findings from Graßl et al. [15], where boys who worked on girl-dominated topics created projects similar in size and complexity to the projects made by girls. One possible explanation is that girls have a higher self-efficacy in languages [20]. Girls feeling more comfortable in creating stories might compensate for the often lower self-efficacy in programming. We did find that boys use more unique commands than nonbinary children in the last level of the story adventure. For the game adventure, there are no differences in levels 1, 2 and 3. Within level 4, boys created longer programs. In levels 4 and 5, boys used more unique commands than girls. The occurrence of gender differences, when the tasks become more difficult, aligns with others [21, 42]. For the visual adventure, we found gender differences for all levels but levels 2 and 4. In level 1, nonbinary children create longer programs than both boys and girls. They also used more commands than girls. Girls used more unique commands in level 3, while boys used more commands than girls in levels 1 and 5. We are unsure why there are inconsistent differences in this particular adventure. We did observe that, overall, the visual adventure seems to have larger programs with more commands than the other two adventures.

Since most related work we found report on binary gender, it is more difficult to compare results related to nonbinary children with previous work. However, finding gender differences even with the smaller sample size indicates that we should continue studying this group of children. Especially since nonbinary children might have their own preferences and little is known about the effectiveness of CS education for this group.

Our results imply that it is possible to create programming assignments where children of different genders work on similar programs in terms of size and number of commands. In our study, this was the case for the story adventure levels 1-4, game adventure levels 1-3 and visual adventure levels 2 and 4. At least for the story and game adventure, differences between genders emerge when Hedy as a programming language becomes more elaborate. This might relate to girls having lower self-efficacy in programming [48] or having less programming experience [22]. Educators should be aware that when CS materials become more difficult, gender differences are more likely to occur. Therefore, it is important to continue researching how to develop CS education that enables equitable experiences for all children.

5.2 Limitations

There are several limitations to this study. We obtained the data via Hedy so we rely on Hedy to ensure its quality. Moreover, Hedy is an active open-source project. Although we tried to minimise the risk, it is possible that assignments or commands changed between May 2023 and May 2024. Furthermore, we analysed a subset of levels, adventures and languages. Since we analysed programs per level, each level/adventure combination consist of one assignment. Including more adventures to represent story, game and visual program types or combining levels could improve the generalisation of the results.

Furthermore, the users self-reported both their gender and their age. Although self-reports come with reliability issues, we analysed a big group of users without interacting with them or assigning genders ourselves, which is needed in studies using Scratch. We think the self-reports mostly impacted the results related to those who selected the 'other' option in the survey. We expect nonbinary children to use the 'other' option, but it is also very likely that children who do not want to disclose their gender choose 'other'. However, we think it is important that CS education research, and especially those focusing on gender, reports more beyond the binary genders to make programming education inclusive to all. Moreover, the dataset does not distinguish trans boys and girls from their cisgender counterparts. Future collection of gender could include an extended list of options or provide open-ended items as suggested by Maloy et al. [32]. A comparison between age groups would also be interesting. However, since only the year of birth is reported, it is more difficult to interpret results.

We think the programs can be analysed more extensively. When using the infrastructure of Hedy to interpret the programs, it becomes easier to evaluate the use of variables, learn more about incorrect use of commands, and apply metrics to measure complexity as done by others [5, 15].

Lastly, we want to mention that we did not take into account where or how children use Hedy. It is possible to see which children belong to a virtual classroom with an assigned teacher, but this was not within the scope of this study. So, in the current work, we don't know whether users programmed in an informal setting (such as at home or at a code club) or at school and whether they used Hedy alone, with peers and/or with the help of teachers.

6 Conclusion

Several groups in society, including women and nonbinary people, face higher barriers to contribute to CS even though CS impacts them. Their interest could be developed at an early age by doing CS activities in a context they like. However, the type of project children work on might influence their learning opportunities. We explore whether children of different genders choose to work on story, game or visual adventures in Hedy, a gradual textual programming language. Moreover, we explore program size and commands used.

We found that the majority of the users worked on each of the adventures. The game adventure seems to be most popular among boys, girls and nonbinary children. For the program length and commands used, we analysed programs created in the first five levels in Hedy per level since each level contains different syntax and commands. Results include nonbinary children creating longer programs than boys and girls, as well as using more commands than girls in the first level of Hedy. When looking at the individual adventures, we did not find any gender differences in story adventure levels 1-4, game adventure levels 1-3, and visual adventure levels 2 and 4. Gender differences do occur in the most elaborate Hedy level for all three adventures, with boys, on average, using more unique commands than nonbinary children in the story adventure, using more unique commands than girls in the game adventure, and using more commands in total than girls in the visual adventure.

Our work can be extended in multiple ways. Firstly, future work could research gender differences in the behaviours of users in Hedy, including users' persistence when facing an error and their navigation through the adventures and levels. Secondly, we suggest continuing to research the learning environments in CS education. This includes the physical environment but also the digital environment and its instructional techniques. Future work should also take prior programming experience into account. With this and future work, we can gain an understanding of how CS education can enable equitable learning opportunities for all children.

References

[1] Rusul Abduljabbar, Hussein Dia, Sohani Liyanage, and Saeed Asadi Bagloee. 2019. Applications of Artificial Intelligence in Transport: An Overview. *Sustainability* 11, 1 (Jan. 2019), 189. https://doi.org/10.3390/su11010189

[2] Joel C. Adams and Andrew R. Webster. 2012. What Do Students Learn About Programming From Game, Music Video, And Storytelling Projects?. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education*. ACM, Raleigh North Carolina USA, 643–648. https://doi.org/10.1145/2157136.2157319

[3] Efthimia Aivaloglou and Felienne Hermans. 2019. Early Programming Education and Career Orientation: The Effects of Gender, Self-Efficacy, Motivation and Stereotypes. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. ACM, Minneapolis MN USA, 679–685. https://doi.org/10.1145/3287324.3287358

[4] Kaan Bati. 2022. A systematic literature review regarding computational thinking and programming in early childhood education. *Education and Information Technologies* 27, 2 (March 2022), 2059–2082. https://doi.org/10.1007/s10639-021-10700-2

[5] Anette Bentz and Bernhard Standl. 2023. Beyond Gender Differences: Can Scratch Programs Indicate Students' Preferences?. In *Proceedings of the 18th WiPSCE Conference on Primary and Secondary Computing Education Research*. ACM, Cambridge United Kingdom, 1–10. https://doi.org/10.1145/3605468.3609771

[6] İbrahim Bozan and Erdal Taşlıdere. 2024. The Effect of Digital Game Design Supported Coding Education on Gifted Students' Scratch Achievement and Self-Efficacy. *International Journal of Contemporary Educational Research* 11, 1 (March 2024), 20–28. https://doi.org/10.52380/ijcer.2024.11.1.531

[7] Oladele O. Campbell, Oluwatoyin Adelakun-Adeyemo, Fatimah Yetunde Akinrinola, Patience Chewachong Akih, Ethel Tshukudu, and Brett A. Becker. 2023. The Impacts of a Constructionist Scratch Programming Pedagogy on Student Achievement with a Focus on Gender. In *Proceedings of the ACM Conference on Global Computing Education Vol 1*. ACM, Hyderabad India, 29–35. https://doi.org/10.1145/3576882.3617911

[8] Ingrid Maria Christensen, Melissa Høegh Marcher, Paweł Grabarczyk, Therese Graversen, and Claus Brabrand. 2021. Computing Educational Activities Involving People Rather Than Things Appeal More to Women (Recruitment Perspective). In *Proceedings of the 17th ACM Conference on International Computing Education Research*. ACM, Virtual Event USA, 127–144. https://doi.org/10.1145/3446871.3469758

[9] Jacob Cohen. 1992. A power primer. *Psychological Bulletin* 112, 1 (1992), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

[10] Caitlin Duncan and Tim Bell. 2015. A Pilot Computer Science and Programming Course for Primary School Students. In *Proceedings of the Workshop in Primary and Secondary Computing Education*. ACM, London United Kingdom, 39–48. https://doi.org/10.1145/2818314.2818328

[11] Eurostat. 2024. ICT specialists in employment. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=ICT_specialists_in_employment#ICT_specialists_by_se

[12] Janne Fagerlund, Päivi Häkkinen, Mikko Vesisenaho, and Jouni Viiri. 2021. Computational thinking in programming with Scratch in primary schools: A systematic review. *Computer Applications in Engineering Education* 29, 1 (Jan. 2021), 12–28. https://doi.org/10.1002/cae.22255

[13] Alexandra Funke and Katharina Geldreich. 2017. Gender Differences in Scratch Programs of Primary School Children. In *Proceedings of the 12th Workshop on Primary and Secondary Computing Education*. ACM, Nijmegen Netherlands, 57–64. https://doi.org/10.1145/3137065.3137067

[14] Marleen Gilsing, Jesús Pelay, and Felienne Hermans. 2022. Design, implementation and evaluation of the Hedy programming language. *Journal of Computer Languages* 73 (Dec. 2022), 101158. https://doi.org/10.1016/j.cola.2022.101158

[15] Isabella Graßl, Katharina Geldreich, and Gordon Fraser. 2021. Data-driven Analysis of Gender Differences and Similarities in Scratch Programs. In *The 16th Workshop in Primary and Secondary Computing Education*. ACM, Virtual Event Germany, 1–10. https://doi.org/10.1145/3481312.3481345

[16] Hedy. 2022. Semantics. https://github.com/hedyorg/hedy/blob/main/semantics/SEMANTICS.md

[17] Hedy. 2024. Homepage Hedy.org. https://www.hedy.org/?language=en&keyword_language=en

[18] Felienne Hermans. 2020. Hedy: A Gradual Language for Programming Education. In *Proceedings of the 2020 ACM Conference on International Computing Education Research*. ACM, Virtual Event New Zealand, 259–270. https://doi.org/10.1145/3372782.3406262

[19] Bas Hofstra, Vivek V. Kulkarni, Sebastian Munoz-Najar Galvez, Bryan He, Dan Jurafsky, and Daniel A. McFarland. 2020. The Diversity–Innovation Paradox in Science. *Proceedings of the National Academy of Sciences* 117, 17 (April 2020), 9284–9291. https://doi.org/10.1073/pnas.1915378117

[20] Chiungjung Huang. 2013. Gender differences in academic self-efficacy: a meta-analysis. *European Journal of Psychology of Education* 28, 1 (March 2013), 1–35. https://doi.org/10.1007/s10212-011-0097-y

[21] Peter Hubwieser, Elena Hubwieser, and Dorothee Graswald. 2016. How to Attract the Girls: Gender-Specific Performance and Motivation in the Bebras Challenge. In *Informatics in Schools: Improvement of Informatics Knowledge and Perception*, Andrej Brodnik and Françoise Tort (Eds.). Vol. 9973. Springer International Publishing, Cham, 40–52. https://doi.org/10.1007/978-3-319-46747-4_4 Series Title: Lecture Notes in Computer Science.

[22] Jung Won Hur, Carey E. Andrzejewski, and Daniela Marghitu. 2017. Girls and computer science: experiences, perceptions, and career aspirations. *Computer Science Education* 27, 2 (April 2017), 100–120. https://doi.org/10.1080/08993408.2017.1376385

[23] Hassan Mansur Hussien, Sharifah Md Yasin, Nur Izura Udzir, Mohd Izuan Hafez Ninggal, and Sadeq Salman. 2021. Blockchain technology in the healthcare industry: Trends and opportunities. *Journal of Industrial Information Integration* 22 (June 2021), 100217. https://doi.org/10.1016/j.jii.2021.100217

[24] Marilena Ianculescu, Adriana Alexandru, and Florin Pop. 2021. Critical Analysis and Evaluation of Current Digital Healthcare Solutions. In *2021 23rd International Conference on Control Systems and Computer Science (CSCS)*. IEEE, Bucharest, Romania, 482–488. https://doi.org/10.1109/CSCS52396.2021.00085

[25] Letizia Jaccheri, Cristina Pereira, and Swetlana Fast. 2021. Gender Issues in Computer Science: Lessons Learnt and Reflections for the Future. (2021). https://doi.org/10.48550/ARXIV.2102.00188 Publisher: arXiv Version Number: 1.

[26] Bo Jiang and Zhixuan Li. 2021. Effect of Scratch on computational thinking skills of Chinese primary school students. *Journal of Computers in Education* 8, 4 (Dec. 2021), 505–525. https://doi.org/10.1007/s40692-021-00190-z

[27] Oscar Karnalim, Mewati Ayub, Maresha Caroline Wijanto, and Felienne Hermans. 2024. Does Hedy, the Gradual Programming Language Help Computing Undergraduates to Learn Programming? In *Towards a Hybrid, Flexible and Socially Engaged Higher Education*, Michael E. Auer, Uriel R. Cukierman, Eduardo Vendrell Vidal, and Edmundo Tovar Caro (Eds.). Vol. 901. Springer Nature Switzerland, Cham, 187–198. https://doi.org/10.1007/978-3-031-53022-7_19 Series Title: Lecture Notes in Networks and Systems.

[28] Hae-Young Kim. 2017. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restorative Dentistry & Endodontics* 42, 2 (2017), 152. https://doi.org/10.5395/rde.2017.42.2.152

[29] Xingyu Long, Peeratham Techapalokul, and Eli Tilevich. 2021. The Common Coder's Scratch Programming Idioms and Their Impact on Project Remixing. In *Proceedings of the 2021 ACM SIGPLAN International Symposium on SPLASH-E*. ACM, Chicago IL USA, 1–12. https://doi.org/10.1145/3484272.3484959

[30] J. Maloney, L. Burd, Y. Kafai, N. Rusk, B. Silverman, and M. Resnick. 2004. Scratch: a sneak preview. In *Proceedings. Second International Conference on Creating, Connecting and Collaborating through Computing, 2004.* IEEE, Kyoto, Japan, 104–109. https://doi.org/10.1109/C5.2004.1314376

[31] John Maloney, Mitchel Resnick, Natalie Rusk, Brian Silverman, and Evelyn Eastmond. 2010. The Scratch Programming Language and Environment. *ACM Transactions on Computing Education* 10, 4 (Nov. 2010), 1–15. https://doi.org/10.1145/1868358.1868363

[32] Jeffrey Maloy, Monika B. Kwapisz, and Bryce E. Hughes. 2022. Factors Influencing Retention of Transgender and Gender Nonconforming Students in Undergraduate STEM Majors. *CBE—Life Sciences Education* 21, 1 (March 2022), ar13. https://doi.org/10.1187/cbe.21-05-0136

[33] Melissa Høegh Marcher, Ingrid Maria Christensen, Paweł Grabarczyk, Therese Graversen, and Claus Brabrand. 2021. Computing Educational Activities Involving People Rather Than Things Appeal More to Women (CS1 Appeal Perspective). In *Proceedings of the 17th ACM Conference on International Computing Education Research*. ACM, Virtual Event USA, 145–156. https://doi.org/10.1145/3446871.3469761

[34] Allison Master, Andrew N. Meltzoff, and Sapna Cheryan. 2021. Gender stereotypes about interests start early and cause gender disparities in computer science and engineering. *Proceedings of the National Academy of Sciences* 118, 48 (Nov. 2021), e2100030118. https://doi.org/10.1073/pnas.2100030118

[35] Monica M. McGill and Adrienne Decker. 2020. Tools, Languages, and Environments Used in Primary and Secondary Computing Education. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education*. ACM, Trondheim Norway, 103–109. https://doi.org/10.1145/3341525.3387365

[36] Connie L. McNeely and Lisa M. Frehill. 2024. A Differentiated Diversity: Demographic Patterns and Contextual Delineations in U.S. Computing. *Computing in Science & Engineering* (2024), 1–16. https://doi.org/10.1109/MCSE.2024.3380614

[37] Orni Meerbaum-Salant, Michal Armoni, and Mordechai (Moti) Ben-Ari. 2010. Learning Computer Science Concepts with Scratch. In *Proceedings of the Sixth international workshop on Computing education research*. ACM, Aarhus Denmark, 69–76. https://doi.org/10.1145/1839594.1839607

[38] Amanda Menier, Rebecca Zarch, and Stacey Sexton. 2021. Broadening Gender in Computing for Transgender and Nonbinary Learners. In *2021 Conference on Research in Equitable and Sustained Participation in Engineering, Computing, and Technology (RESPECT)*. IEEE, Philadelphia, PA, USA, 1–5. https://doi.org/10.1109/RESPECT51740.2021.9620612

[39] Mark Noone and Aidan Mooney. 2018. Visual and textual programming languages: a systematic review of the literature. *Journal of Computers in Education* 5, 2 (June 2018), 149–174. https://doi.org/10.1007/s40692-018-0101-5

[40] U.S. Bureau of Labor Statistics. 2024. Computer and Information Technology Occupations. https://www.bls.gov/ooh/computer-and-information-technology/home.htm

[41] Makseem Angel Skorodinsky. 2024. *More Than Binary, More Than Normative, More Than Quantities: Transgender and Gender Nonconforming Students in Postsecondary Computer Science Education.* Ph. D. Dissertation. University of Oregon. https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/29713/Skorodinsky_oregon_0171A_13802.pdf

[42] Amanda Sullivan and Marina Umashi Bers. 2016. Girls, Boys, and Bots: Gender Differences in Young Children's Performance on Robotics and Programming Tasks. *Journal of Information Technology Education: Innovations in Practice* 15 (2016), 145–165. https://doi.org/10.28945/3547

[43] Stella Timotheou, Ourania Miliou, Yiannis Dimitriadis, Sara Villagrá Sobrino, Nikoleta Giannoutsou, Romina Cachia, Alejandra Martínez

Monés, and Andri Ioannou. 2023. Impacts of digital technologies on education and factors influencing schools' digital capacity and transformation: A literature review. *Education and Information Technologies* 28, 6 (June 2023), 6695–6726. https://doi.org/10.1007/s10639-022-11431-8

[44] Mark Watson and Mary McMahon. 2016. *Career Exploration and Development in Childhood: Perspectives from theory, practice and research* (1 ed.). Routledge, Abingdon, Oxon ; New York, NY : Routledge, 2017. https://doi.org/10.4324/9781315683362

[45] Gayna Williams. 2014. Are You Sure Your Software Is Gender-Neutral? *Interactions* 21, 1 (Jan. 2014), 36–39. https://doi.org/10.1145/2524808

[46] Zhen Xu, Albert D. Ritzhaupt, Fengchun Tian, and Karthikeyan Umapathy. 2019. Block-based versus text-based programming environments on novice student learning outcomes: a meta-analysis study. *Computer Science Education* 29, 2-3 (July 2019), 177–204. https://doi.org/10.1080/08993408.2019.1565233

[47] Sabiha Yeni and Anna Van Der Meulen. 2022. Students' Behavioral Intention to Use Gradual Programming Language Hedy: A Technology Acceptance Model. In *Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 1*. ACM, Dublin Ireland, 331–336. https://doi.org/10.1145/3502718.3524797

[48] Christina Zdawczyk and Keisha Varma. 2023. Engaging girls in computer science: gender differences in attitudes and beliefs about learning scratch and python. *Computer Science Education* 33, 4 (Oct. 2023), 600–620. https://doi.org/10.1080/08993408.2022.2095593

Received 2024-07-09; accepted 2024-08-08