

Gender-dependent Contribution, Code and Creativity in a Virtual Programming Course

Isabella Graßl
isabella.grassl@uni-passau.de
University of Passau
Passau, Germany

Gordon Fraser
gordon.fraser@uni-passau.de
University of Passau
Passau, Germany

ABSTRACT

Since computer science is still mainly male dominated, academia, industry and education jointly seek ways to motivate and inspire girls, for example by introducing them to programming at an early age. The recent COVID-19 pandemic has forced many such endeavours to move to an online setting. While the gender-dependent differences in programming courses have been studied previously, for example revealing that girls may feel safer in same-sex groups, much less is known about gender-specific differences in *online* programming courses. In order to investigate whether gender-specific differences can be observed in online courses, we conducted an online introductory programming course for SCRATCH, in which we observed the gender-specific characteristics of participants with respect to how they interact, their enjoyment, the code they produce, and the creativity exposed by their programs. Overall, we observed no significant differences between how girls participated in all-female vs. mixed groups, and girls generally engaged with the course more actively than boys. This suggests that online courses can be a useful means to avoid gender-dependent group dynamics. However, when encouraging creative freedom in programming, girls and boys seem to fall back to socially inherited stereotypical behavior also in an online setting, influencing the choice of programming concepts applied. This may inhibit learning and is a challenge that needs to be addressed independently of whether courses are held online.

CCS CONCEPTS

• **Social and professional topics** → **Computing education; K-12 education; Gender.**

KEYWORDS

Scratch, gender, diversity, online course, primary education.

ACM Reference Format:

Isabella Graßl and Gordon Fraser. 2022. Gender-dependent Contribution, Code and Creativity in a Virtual Programming Course. In *Proceedings of the 17th Workshop in Primary and Secondary Computing Education (WiPSCE '22)*, October 31–November 2, 2022, Morschach, Switzerland. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3556787.3556861>

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

WiPSCE '22, October 31–November 2, 2022, Morschach, Switzerland

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9853-4/22/10...\$15.00

<https://doi.org/10.1145/3556787.3556861>

1 INTRODUCTION

The gender gap in computer science (CS) remains despite dedicated programs that specifically empower girls [40], especially in the field of software engineering [1]. Programming education plays a crucial role in addressing this gap, as girls have different expectations, perceptions, and learning outcomes from programming education than boys [36, 40]. This has led to two important conclusions: First, it is important to teach programming early, as girls' enthusiasm and self-efficacy [2] may be better addressed at a younger age than during or after puberty [37]. Second, it has been suggested by many studies [8, 9, 20, 22, 24, 28, 43] that same-sex courses support the engagement of girls in CS.

In the recent past the COVID-19 pandemic has thrown a spanner in the works of many such endeavours, forcing programming courses to move to online settings and thus causing new challenges: Active participation is a prerequisite for good results [14, 44], but it is difficult to control whether students are actually following what they are supposed to in an online course [18]. Similarly, it is unclear if the virtual setting has a negative impact on enjoyment, which is considered to be an important catalyst for engagement [7, 27, 38], creativity [34], or the resulting motivation [11]. Finally, the virtual setting influences how learners interact with each other and with teachers. All of these are important factors that may influence how girls engage with CS, thus calling for further research on gender-specific characteristics in online programming courses.

In order to investigate possible gender-specific characteristics, we conducted and observed an online programming course for young learners. We designed a six-week introductory programming course in SCRATCH [31]. In order to avoid biasing the behaviour of the observed learners, we put a particular focus on creating only gender-neutral tasks and exercises. We conducted this course online using a popular video conferencing system with 33 girls and 38 boys aged 10 to 14 years, divided into four groups: all-female, all-male and two mixed groups¹. Using this setup, we investigate whether girls' and boys' interest and interactions during the course, as well as the code and creativity of the programs they produced as part of this course, differ between girls and boys and between same-sex and mixed-sex groups.

Our results suggest that the virtual course setting does not reinforce gender-dependent characteristics as a result of specific group constellations: We observed no significant differences between how girls participated in all-female vs. mixed groups. This suggests that online courses may be a useful means to avoid gender-dependent group dynamics. Comparing genders across all groups, it appears that the online setting benefits girls: Overall, we observed that

¹We explicitly do not support binary gender thinking, however, the children categorized themselves as female and male.

girls performed slightly better than boys in terms of contributions, correctness of programs, creativity and additional tasks. However, despite our emphasis on a gender-neutral course design, both girls and boys resort to stereotypical topics when given creative freedom, also in an online course. This, in turn, has an effect on which programming concepts are applied in the projects created by boys and girls, thus potentially inhibiting further learning. Consequently, this is a challenge that needs to be addressed independently of whether courses are held online or in-person.

Our experiments are only the first step towards a deeper understanding of how different factors influence learners depending on gender in an online setting. In particular, further replications will be necessary in order to validate our initial findings. To support such replications and future research we provide all course materials and evaluation scripts online.²

2 BACKGROUND AND RELATED WORK

SCRATCH is a block-based programming language that makes basic programming concepts accessible for beginners using a simple visual format, in which they can engage creatively with programming [32]. This is relevant in motivating children to learn programming, as there are gender-dependent topics that girls and boys implement differently in SCRATCH projects [15, 19, 35]. Similarly, the implementations and complexities of these programs have been observed to differ between genders [13, 15].

It is important to investigate these possible gender differences because the ratio of women in CS is still comparatively low [1, 6], and a deeper understanding is a prerequisite to changing this. The possible reasons for gender differences are manifold: Extrinsic factors may be the persistent stereotyping of programming as a male discipline and the few female role models [23, 38]. Intrinsic reasons may be that girls often suffer from socially inherited low self-efficacy [2]. Especially in the field of CS, this might lead to a tendency to seek less demanding and difficult challenges as well as to more frustration and disinterest in the subject [3–5, 10, 30, 39].

One approach to address these problems is to divide students into same-sex groups as they may have a gender-specific positive effect on the level of experience and positive attitude towards computers [20]. However, this is mainly observed in first-time introductions to the subject, hence, an all-female group makes sense especially in early years. This experience can be the first major step in making girls aware of CS and their own abilities [27, 38, 41].

In the context of software engineering, grades in same-sex classes have been observed to be slightly better for both all-female and all-male groups than in mixed classes [30, 43]. Although the quality of the programs was similar between genders in prior studies, girls in all-female classes followed the instructions more attentively, were more likely to stick to the requirements, and submitted more programs [14, 26, 37, 44].

3 ONLINE COURSE DESIGN

To enable our investigation, and to continue encouraging young learners to engage with programming throughout the pandemic, we created an introductory SCRATCH course consisting of six one-hour lessons, with two objectives in mind: First, the course needs to be



Figure 1: Design of the task for each session with the introduced programming concept and project type.

suitable for teaching in an online, video-based scenario; second, to prevent any bias in the teaching material, all example programs and figures have to be gender-neutral as much as possible and avoid stereotypes. We provide all course material online.²

3.1 Lesson Structure

In order to provide the students with a consistent learning experience, all sessions of our course follow the same procedure: They start with a retrospective of the previous session, which ensures that all students are up to speed, and provides an opportunity for discussing open questions. Then, the programming concept of the session is introduced with a sample project, which also represents the objective of this session. The first steps of the example program are demonstrated using the screen sharing function of the video conferencing software, actively involving the students to work on the concrete problem solution. Then, the students are asked to implement the same steps by themselves. This alternation of showing, working together and independently continues until the programming concept of the lesson is realized, and students have co-constructed their insights. At the end of each session, students have the opportunity to present their programs so that their work is valued and their self-confidence is strengthened. After each lesson, a PDF summary of the lesson with possible solutions is sent to the students.

3.2 Online Supervision

Each session is presented by one dedicated instructor, who is supported by at least two additional supervisors. Students are supposed to unmute their microphones and ask their questions directly rather

²<https://github.com/se2p/scratch-online-course>

than typing them, allowing other students to contribute by answering questions. If a student has a more severe problem (e.g., technical difficulties), the supervisors can switch to a breakout room for one-on-one supervision. For basic lessons, all students are kept in a single online room. For larger, creative programming tasks after the introduction of the concept to be learned in the session (sessions 4 to 6), children are split into smaller groups with approx. 5 children and one individual supervisor.

3.3 Course Contents

The course consists of six lessons (Fig. 1), each structured into sub-tasks. The order and selection of programming concepts is based on the principle that sequences represent the simplest structures, followed by loops and conditional statements, which form the basis for conditional repetitions, which in turn implicitly introduces variables [13, 17]. For each lesson, we provide an optional additional task as homework to further practice the concept learned. We made sure to use gender-neutral characters and narratives, and to find a balance between stories and games. In particular, many animals were chosen as showcase characters because they have no gender in SCRATCH and are universally popular. By doing so, we aim to avoid possible gender or cognitive biases and allow the children the best possible level of creativity.

Session 1. The first lesson starts with questions about programming and the children’s experience with it as an icebreaker. The most basic functions of SCRATCH are explained using a wizard in the forest doing magic tricks as most children like those (Fig. 1a). Based on this, the standard blocks in the *looks*, *sound* and *motion* categories are introduced. We also make use of the blocks of the *text to speech* extension, which allows a more unique media experience [11]. Only sequences are used which are perceived as being easy for beginners [17]. The transformation of the wizard into a rabbit is intended as additional task.

Session 2. The programming concept *loops* (forever, repeat-times) is introduced using a ‘dancing’ milk carton, where the dancing is simulated by switching costumes, while standing on a stage with music from a music box (Fig. 1b). This example clearly demonstrates the use for loops as it quickly leads to a feeling of success, and additionally the choice of music provides a degree of creative freedom. The milk sprite is one of the few characters in the standard catalogue that make the appearance of dancing when changing costumes, but does not have a gender and does not fall into stereotypes. In a second task (Fig. 1c), the loop concept is practiced with a polar bear drawing a number of steps up to a penguin and saying *hi*. In an additional task, the polar bear should do a ‘flip’ after each step, using a nested loop inside the outer loop.

Session 3. *Conditional statements* (if, if-else) are introduced with a dog that asks the students what their name is (Fig. 1d). The students are immediately involved by choosing their own figure to ask for their names, thus establishing a lifeworld connection [11]. As a second task, a simple game in which a flying cat is controlled using the cursor keys, and has to fly to the sun without touching any clouds, is created. As an extra task, this program can be extended with multiple clouds, balloons that increase the speed, or a background with a winning-message.

Session 4. (Fig. 1f) In the fourth lesson, conditional repetition (repeat-until) is introduced because it presupposes the basic concepts of loops and conditional statements. We designed a game in which the cat has to catch the mouse. The cat is controlled by the cursor keys and moves continuously until it catches (i.e., touches) the mouse, and then says “gotcha”. As an additional task, different levels can be included using backdrop changes, for example, a chick asks the cat if it can turn it into a butterfly. Implicitly, variables are already present in the conditional repetition, which the next lesson explicitly addresses with active settings and changes [17].

Session 5. For the introduction of variables, we created two alternative versions of the same game for which students can vote which version they want to see. The game either consists of flying hippos that can be zapped with lightning (Fig. 1g), or trees that can be watered with raindrops (Fig. 1h). In both cases, variables are used to track the player score. Both games have the same code, but we explicitly wanted to show a positive event handling with the enlivening of a tree and one event with the destruction of something, but in a child-friendly way. As an additional task, the game can be extended with different levels, or by asking the name of the player and storing it in a variable before starting the game.

Session 6. (Fig. 1i) The final lesson starts with a recap of all concepts learned using a program containing all the figures from the previous sessions. The task for students is then to apply all these concepts in their own final program, in which they create different backdrops representing rooms or levels (variables) into which their chosen character is moved through doors controlled by the cursor keys. As additional task, they should creatively expand their world in some way, for example, adding signs as hints or figures that they should avoid. The demonstration project consists of a ball moving through colored doors on a gray background, dodging a beetle and receiving hints from signs. This choice aims to avoid that the children are influenced by the example project.

4 METHOD

In this paper we aim to empirically answer the following research questions in the context of an online programming course:

RQ1: *Do girls and boys show different levels of contribution?*

RQ2: *Do girls and boys show different levels of interest?*

RQ3: *Do programs of girls and boys show differences in code?*

RQ4: *Do programs of girls and boys show differences in creativity?*

4.1 Data Collection

We conducted the synchronous online course in April/May 2021 over six weeks, targeting children aged 10 to 14 without prior programming experience. An open invitation to the course was posted on the university and department websites, a local newspaper, and it was sent to all schools in the area by mail. We explicitly encouraged girls to get involved with programming in the invitation by asking them to simply give programming a try. The course was conducted online using the video conferencing software Zoom³. Although each session was scheduled for 60 minutes, this time was usually exceeded due to ongoing questions. We used Zoom’s survey feature to collect survey data at the end of each session. Researchers filled

³<https://zoom.us>

Table 1: Attendance of the students per session and in total.

Group	Sex	Session						Total
		1	2	3	4	5	6	
A (f)	f	13	14	15	15	14	12	16
B (m)	m	16	17	16	17	15	15	18
C (f/m)	f	7	9	9	8	8	7	18
	m	8	7	7	8	7	7	
D (f/m)	f	7	8	8	7	8	6	19
	m	10	10	9	8	9	11	

different roles within the sessions to ensure a clear course flow and data logging: an instructor explained the tasks and led through the course, an observer logged all interactions in a protocol, and at least two supervisors were available for questions, of which they also kept a log. The leadership roles were gender-balanced so that three sessions were led by a woman and three sessions were led by a man. We used the SCRATCH classroom feature and created a classroom and user accounts for participants, such that we had access to the projects created by the participants during and outside the lessons.

4.2 Dataset

A total of 88 students enrolled in the course, of which 8 did not show up and 80 participated. Out of these 80, we isolated 9 students who revealed prior experience up front into a separate “advanced” group, which is not included in the analysis. We consider in the analysis 71 German students (33 female, 38 male) randomly distributed by their sex in four groups, aged 10 to 14 years (age/sex: A=all-female: 10.68/16f, B=all-male: 11.35/18m, C=mixed: 11.00/10f/8m, D=mixed: 11.73/8f/11m). The majority of students attended high school (71,62 %), although students from junior high (14,86 %) and elementary schools participated (13,51 %). All sessions were consistently very well attended by all students (Table 1). Besides the protocols written by the researchers, the dataset consists of demographic and sex information for the participants (elicited as part of the course registration), survey responses collected at the end of each session via Zoom, and the participants’ SCRATCH projects. We accessed and downloaded all projects created by the participant users in the SCRATCH classroom six weeks after the end of the course.

4.3 Data Analysis

RQ1: Contribution. In order to determine the contributions of individual participants throughout the sessions, we noted all interactions in written protocols for the main Zoom session as well as the breakout groups. Each observation was assigned an ID, the student anonymized, and sex marked. A coding scheme was designed and each observation was labeled using these different categories and subcategories. Participation focused specifically on the categories of question and answer and whether they were SCRATCH-related, organizational, technical, or private. In addition, reactions and contributions that are considered smalltalk were categorized. To answer RQ1, we compare the questions and answers provided dependent on gender and group constellation.

RQ2: Interest. In order to determine the subjective perceptions of the students, we conducted a short survey at the end of each

session, and evaluate the responses for RQ2. Each survey included three 3-point Likert scale questions about whether (1) the lesson was fun and comprehensible (2) how difficult the lesson was, and (3) whether the students engaged with SCRATCH at home. The results of the surveys were aggregated, anonymized, and gender-tagged.

RQ3: Code. We evaluate the code of the students’ programs using LITTERBOX [12] to extract statistics about size, complexity, and other code attributes. In addition, we manually verified for all sessions and participants whether the concepts learned were implemented correctly and whether the additional tasks were completed. We reviewed to what extent the students published their projects, and which projects they created at all. We compare these statistics in terms of gender and group context, and measure statistical differences using a Wilcoxon Rank Sum test with $\alpha = 0.05$.

RQ4: Creativity. We evaluate the content of the students’ programs implemented in the course. We define creativity by manually checking whether any creative deviations exist from the sample projects, such as custom sprites and backgrounds, or additional figures and colors. Each program was reviewed by two independent researchers, resulting in an inter-rater agreement of $K = 0.92$. We compare the number of creative deviations and qualitatively study the topics chosen.

4.4 Threats to Validity

Due to the virtual setting, we cannot always ensure that children worked alone. Parents consented to use the data for research purposes and no image or audio files of the children were recorded. The observations in the course are subjective in nature, but these come from three independent supervisors per session. The evaluation of the protocols and projects can lead to misinterpretations, so these were also evaluated by several independent researchers. The design as well as the realization of the introductory course were evaluated by several independent didactic experts. Students were all from Germany with no further data available on social background. While the number of participants was large relative to the desired target group size for our course, more datapoints will be needed for generalizable results. In particular, the number of supervisors as well as observers limits the number of courses that can be reasonably conducted. Although we use the common value of $\alpha = 0.05$ for statistical tests, observations are of interest independently of this arbitrary boundary, so we provide the raw p -values to allow the readers to better interpret the data. We therefore provide all material and hope that this enables future replications of our online study.

5 RESULTS

Across all research questions, we observed almost no significant differences between the individual same-sex and mixed groups, i.e., comparing data between girls in the same-sex and mixed groups, and comparing data between boys in the same-sex and mixed groups. The p -values are summarized in Table 2: There is only a single significant difference, comparing the number of correct tasks (RQ3) between girls in the same-sex and mixed group D. We conjecture that this p -value is influenced by the relatively small number of girls

Table 2: Comparison of metrics between same-sex and mixed groups by their statistical significance ($\alpha = 0.05$).

RQ	Metric	A-B	A-C(f)	A-D(f)	B-C(m)	B-D(m)
RQ1	contribution	.06	.43	.68	.31	.09
RQ2	fun	.44	.38	.44	.27	.41
	difficulty	.24	.49	.40	.29	.74
	engagement	.08	.49	.18	.46	.19
RQ3	correct tasks	.10	.94	.01	.25	.46
	add. task	.15	.43	.68	.43	.84
RQ4	add. sprite	.06	.49	.99	.64	.07

Table 3: Categories of the observations for each group.

Category	Subcategory	A (f)	B (m)	C (f/m)	D (f/m)
answer	correct	51	56	64	47
	incorrect	5	7	18	4
	suggestion	4	1	1	6
chitchat	general	2	9	10	0
	task-related	8	8	11	25
question	creativity	9	9	6	13
	organizational	6	7	7	7
	scratch general	16	31	13	10
	task-related	83	87	109	108
	technical	4	3	15	6
reaction	status	14	3	10	9
	showing project	15	17	14	18
	other	7	4	7	5
Σ		230	242	285	258

in group D, who happened to provide a lower number of submissions in the last two tasks. Thus, overall we conclude that same-sex groups seem to provide no benefits in our online setting.

RQ1–4 Same-sex vs. Mixed Groups Summary. We observed almost no significant differences between how girls participate and perform in same-sex and mixed groups in an online setting, and there are also no significant differences how boys participate in same-sex and mixed groups.

In the following we will therefore only discuss the overall differences between girls and boys across all groups, and omit further comparisons of groups.

5.1 RQ1: Contribution

Overall, almost across all groups and the whole course, over 50 % of all students actively contributed—with the highest active participation of all groups in the fourth session with over 80 %. For a virtual setting in particular, this is significant as it shows that students were involved [14, 18]. Table 3 breaks down the observations into the categories and subcategories for each group. Overall, the two mixed groups (C, D) have the most active participation, while the two same-sex groups (A, B) have comparable values in the frequency of participation.

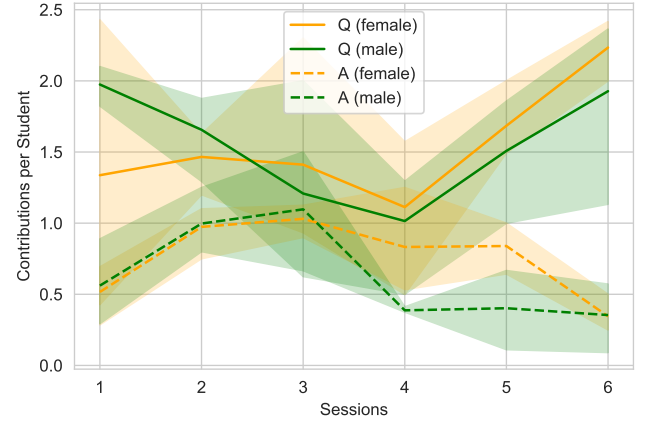


Figure 2: Questions and answers of students.

Figure 2 illustrates the contributions classified in questions (solid lines) and answers (dashed lines) in total, summed up and divided by the participants per session. The plot further shows the 95 % confidence intervals; intuitively, if the confidence intervals overlap, there is no statistically significant difference. The total number of questions and answers over all sessions is comparable between female and male students (f: 407, m: 435). In terms of a virtual setting, the number of contributions on average per session regarding questions and answers is striking (f: 67.83, m: 72.50).

According to Figure 2, more questions were asked than answers given regardless of gender, which is due to the nature of an introductory course. For both genders, the highest number of questions as well as the highest ratio of questions to answers can be seen in the first and last sessions: In the first session mostly organizational matters and technical problems were discussed and there was not much room for students providing their own answers. In the last session, no new programming concepts were introduced, so there was intuitively less need for answers, but because students were encouraged to integrate their own ideas many questions were raised. The ratio of questions and answers is most balanced in the second and third sessions. This may be due to the course design, as in these two sessions an introductory task was implemented before moving on to the main task, where it consequently may have been perceived as easier for the students to participate. In the fourth session participants of both genders showed the least interactions, which may be related to the complexity of the programming concepts introduced in this session (conditional loops).

Comparing the behavior of girls and boys, especially in the first session, boys asked about one third more questions than girls (Fig. 2). However, girls participated more often than boys from the third session onwards with questions as well as answers, which may be due to their initial lower self-efficacy or self-confidence [23, 41]. Increased confidence after positive experiences apparently caused them to participate more actively overall [3, 7]. In addition, the high response rate in sessions two and three may suggest that girls mainly respond to quick positive feedback [40] that they received through the introductory tasks in these sessions. Notably, girls

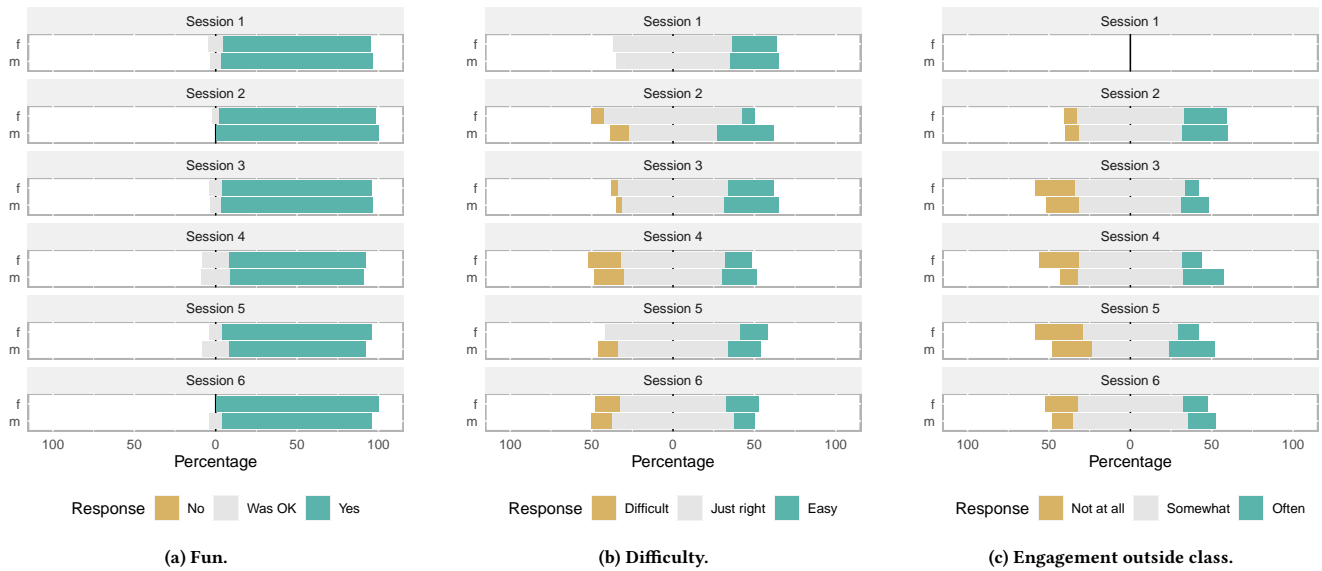


Figure 3: Distribution of fun, difficulty, and engagement.

provided significantly more answers in the fourth session, which introduced more advanced programming concepts (Fig. 2).

RQ1 Summary. All students contributed very actively. Girls asked fewer questions in the first sessions, but contributed considerably more overall than boys after the third session.

The virtual setting has the potential to avoid gender-dependent group dynamics. Quick positive feedback in tasks and first positive experiences seems also relevant for a higher contribution in an online course, especially for girls.

5.2 RQ2: Interest

Figure 3 summarizes the results of the surveys conducted at the end of each session. Both genders overwhelmingly enjoyed the course (f: 92 %, m: 91 %; Fig. 3a). The majority of students also felt the difficulty of the sessions (Fig. 3b) was appropriate (f: 73 %, m: 65 %), with boys finding the sessions slightly easier (f: 19 %, m: 26 %). Overall, this shows that the course was appealingly designed and realized. The fourth session stands out slightly with somewhat lower reported fun and higher reported difficulty. This may be related both to the greater independence of the increasingly free tasks and to the flattening learning curve, so that the advanced programming concepts are more challenging to acquire.

According to their self-reported engagement (Fig. 3c), the majority of participants of both genders engaged *somewhat* with SCRATCH individually after the sessions (f: 64 %, m: 62 %). The claimed engagement following the second session stands out with both genders engaging *often* with SCRATCH almost similarly, which may be due to their initial motivation and curiosity. From sessions 3 onwards, boys claimed to engage outside the classroom more than girls. This engagement, however, is not reflected in completion rate of the

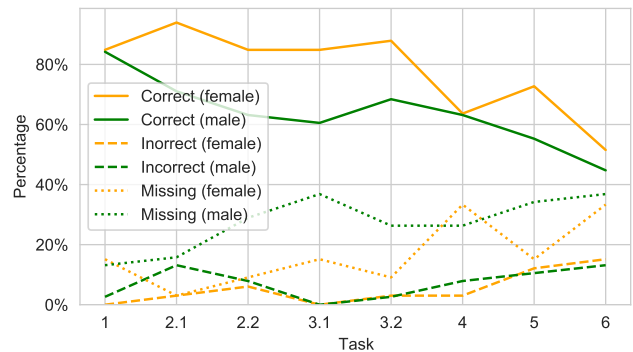


Figure 4: Distribution of implementing the tasks correctly.

additional task of session two, which will be discussed as part of RQ3 (cf. Fig. 5 in Section 5.3.2).

RQ2 Summary. Both genders enjoyed the course and engaged equally. The majority found the difficulty to be adequate.

Since the virtual setting did not seem to have any negative effects on the students' interest of either gender, it is suitable for further courses. The small differences observed suggest that teachers should maybe try to strengthen the girls' self-concept due to the difference between it and their actually performed tasks.

5.3 RQ3: Code

In order to answer RQ3, we consider the code of the SCRATCH projects created by the learners.

5.3.1 Correctness. Figure 4 summarizes to what degree the individual tasks were correctly solved by the participants throughout

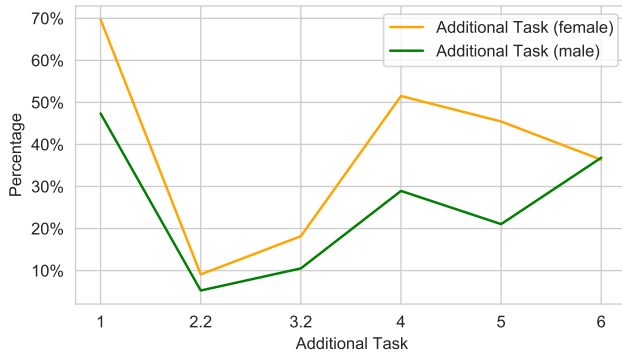


Figure 5: Distribution of implementing additional tasks.

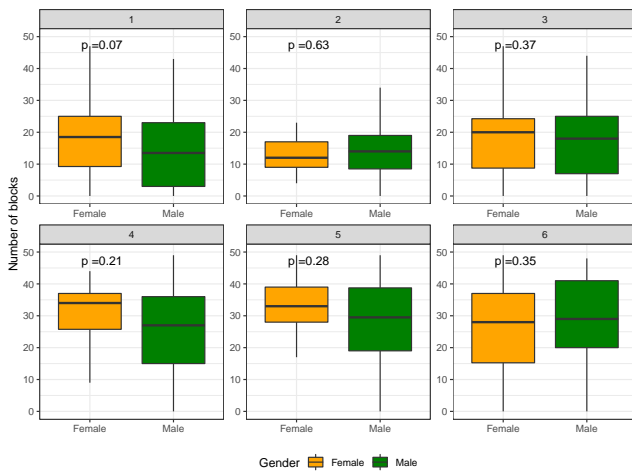


Figure 6: Distribution of the number of blocks among sessions/genders.

the course, separated by gender. Girls provided correct implementations of the tasks considerably more often than the boys over the entire course ($p = 0.18$), with particularly pronounced differences in the second and third session. Both genders also show a similar pattern of incorrect implementations, but the percentage is higher for boys ($p = 0.16$). Boys' projects are also completely missing more often than girls' projects ($p = 0.12$), in total almost twice as many are missing from boys (83) than from girls (44). This generally suggests that girls appear to follow the tasks more attentively, are less distracted and more into the subject because their projects are less absent, which is in line with the findings of recent studies [25].

5.3.2 Engagement. Figure 5 summarizes the completion of the additional (optional) tasks throughout the course per gender. Girls completed the additional task (Section 3.3) more often than boys throughout the entire course ($p = 0.23$). The largest difference between the genders can be seen in the additional task of sessions four and five. Beyond the optional homework tasks, 35 female students and 30 male students also created projects on their own independently of any tasks set. Especially after the first (f: 9, m:

7) and the last session (f: 11, m: 8) the creation of self-initiative projects was highest.

When contrasting the students' self-concept with the actual results, it is interesting to note that the same number of children of both genders claimed to engaged *a little bit* with SCRATCH and girls indicated less to engage *often* with it (Fig. 3c), when the implementation of the additional tasks and additional projects shows us (Fig. 5) that girls actually engaged *more* with SCRATCH using their SCRATCH account across the board!

5.3.3 Code Metrics. Figure 6 illustrates that the number of blocks used by the students for the defined tasks increases slightly with each task, and is relatively balanced, with girls tending to use slightly more blocks. Up to the third session, the block types used by both genders (Fig. 7) are very similar. However, for the more complex and less constrained successive tasks, there are differences: Especially in the fourth and fifth sessions, girls used many *looks* blocks such as changing the costume of the sprites, which generally represents the type of block they use preferably [13, 15, 33]. This is particularly interesting since explicit attention was paid to a gender-neutral course design in terms of topics and project types. The use of *data* blocks is also interesting because the concept of variables was not introduced until the fifth session, but is applied by both girls and boys in prior lessons. Girls in particular seem to have internalized variables in the last lesson, as they use them most often for implementing their own world.

Considering the code complexity of the implemented tasks (Fig. 8), the difference is small for the first three tasks, which do not allow for much variety. However, the girls' programs are considerably more complex for the fourth session. This is likely because more than half of these projects include code resulting from the additional task (Fig. 5), which involved incorporating different stage changes to represent different levels, and incorporating different event blocks that have different ways of starting a script when switching to a new level. This also explains the high proportion of looks, motion, control and event blocks (Fig. 7). However, the girls' projects in the fifth and sixth session show a distinctly lower complexity compared to the boys' projects. This seems to be because the girls implemented only the most necessary code elements and then focused more on adding other elements such as decorations or signs without further functionality.

RQ3 Summary. Both genders use slightly different types of blocks and have programs of varying complexity. Girls provided correct and additional implementations more often than the boys.

As girls follow tasks more attentively their extra work should be acknowledged, whereas boys should be encouraged to also follow tasks attentively. When setting open tasks it is important to include playful minimum requirements, as otherwise the girls' knowledge level may stagnate, causing problems with respect to long-term interest.

5.4 RQ4: Creativity

The sprites and backgrounds used in the programs provide insights into the creative ideas implemented by the children, which is relevant as it is important to allow learners to express their own creativity [15, 34] in order to achieve the best possible learning and

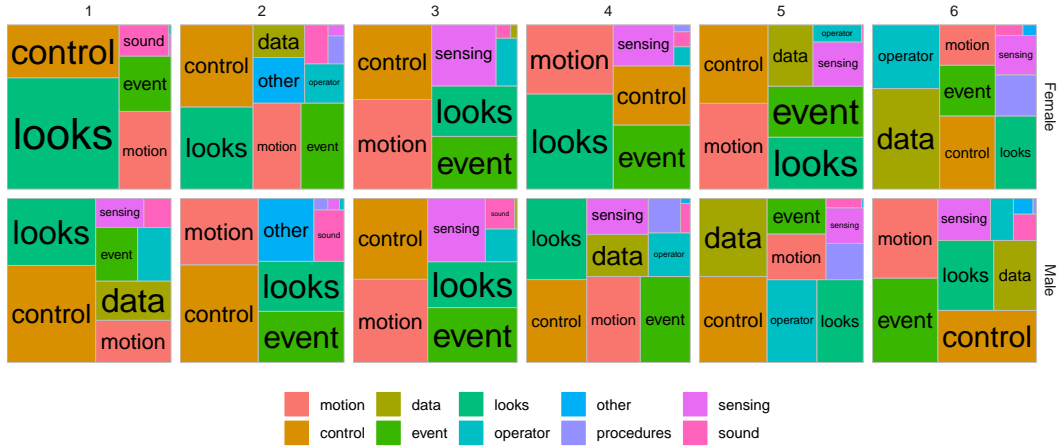


Figure 7: Distribution of block types among session/gender.

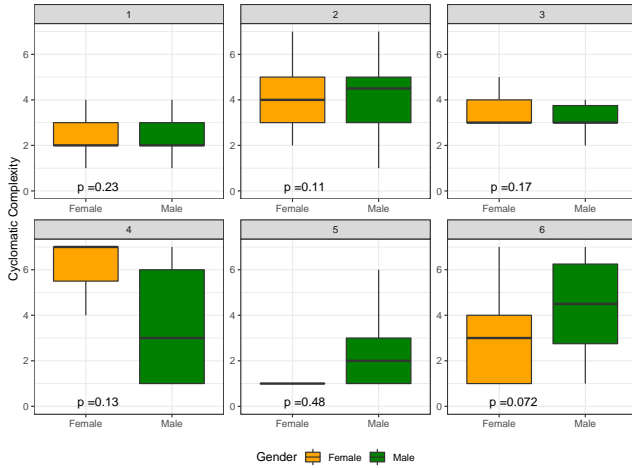


Figure 8: Distribution of code complexity measured by ICC.

motivational success. Girls use more of their own sprites than boys ($p = 0.21$) in all but the second session's introduction task (Fig. 9). This suggests that the girls have adapted and tweaked the projects according to their own preferences and therefore have richer programs [33]. In particular for the last three sessions, which allowed for more creative freedom, the choice of sprites provide insights into the topics chosen by the students: The fourth task is a catching game, demonstrated with a cat and mouse. Both genders remained committed to animal sprites and use chicks, penguin, butterfly and monkey (Fig. 10a) as well as bear, crab, fish, owl, fox, polar bear and squirrels (Fig. 10b). Both genders also use fantasy creatures, with girls using unicorns and centaur and boys using dinosaurs. The fifth task is a simple arcade game introduced in two versions, space or greenland scenario. Girls create programs with the setting in a castle featuring princesses, crystals, griffins and ghosts (Fig. 10c) while boys preferred the space theme and added robots, dogs, bats and explosion (Fig. 10d). In the last session, the children

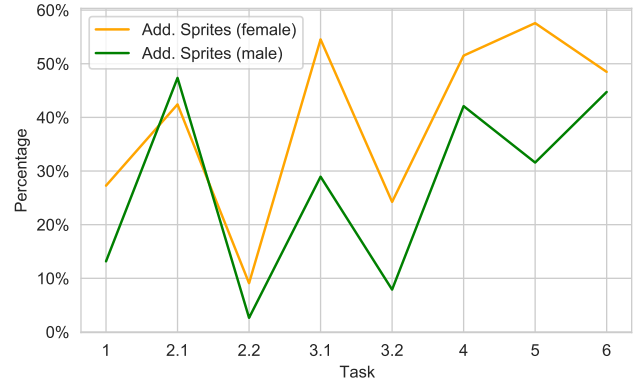


Figure 9: Distribution of creating an own sprite.

were asked to create their own world. The world of girls seems to be characterized by the color pink as well as fantasy creatures like fairies and mermaids, with bats in it as well (Fig. 10e). The boys use ball as sprite as well as soccer as topic (Fig. 10f). There are many different animals in both worlds.

Clearly, both girls and boys resort to stereotypical topics when given creative freedom. This is particularly interesting since we placed explicit emphasis on using gender-neutral topics in the course design and depicting neither implicit nor explicit female or male sprites or stereotypical topics.

RQ4 Summary. Girls use their own sprites more than boys. Both genders like to use animals as well as stereotypical preferences.

Our observations suggest that gender-specific interests are already internalized at a young age and students express them—analogueous to other media—in programming. It is therefore desirable for teachers to ensure that tasks are not gender-specific, so that as many students as possible feel addressed and can apply their creativity without bias.



Figure 10: Distribution of the names of the self-chosen sprites, where the size of the name indicates its frequency.

6 CONCLUSIONS AND FUTURE WORK

We designed and implemented an online introductory programming course in SCRATCH with gender-neutral tasks covering the main programming concepts. Since both genders equally enjoyed the course and found the level of difficulty appropriate, this suggests that the course design is successful even with the challenges of a virtual setting.

Our data shows no significant differences in how girls participate in all-female vs. mixed groups. In line with previous studies, girls were reserved in the first sessions, but contributed more than boys from halfway through the course—both in questions and answers. Similarly, the rate of completed tasks was comparable, although girls slightly had the edge while boys’ projects are more often missing. Furthermore, girls, regardless of the group constellation, considerably more often implemented the additional task and also additional projects during the sessions. In terms of creativity, girls used their own sprites more often. Even in a gender-neutral course setting both genders chose topics in their projects that reflect stereotypical established norms from society.

An important question for future research is whether the observation that girls contributed more often and performed more additional tasks in our study is influenced more by the course design or the online format. We are aware of the limited generalizability of our observations, and since the implementation of extra-curricular courses and the participation of sufficient girls is challenging, we encourage replication studies using our replication package online.⁴

In addition, the effects of working in small groups could be investigated with the concept of pair programming, which has already been shown in some small studies [21, 29, 42, 45].

Our experience has taught us that we have to reach out not only to the girls, but also to their environment, especially their parents: Many parents have misconceptions about programming and computer science [16], so it could be helpful to convince them as well, as a very motivated student (12f) from the all-female group said in the last session about the subsequent advanced course we offered: “There is a continuation? My father signed me up for the course here because he said I should do that, but my mother said that will definitely not interest me”—but it definitely did.

ACKNOWLEDGMENTS

We thank Verena Ewald and Florian Obermüller for their contribution to the data collection. This work is supported by the Federal Ministry of Education and Research through project “primary::programming” (01JA2021) as part of the “Qualitätsoffensive Lehrerbildung”, a joint initiative of the Federal Government and the Länder. The authors are responsible for the content of this publication.

REFERENCES

- [1] K. Albusays, P. Bjorn, L. Dabbish, D. Ford, E. Murphy-Hill, A. Serebrenik, and M.-A. Storey. 2021. The Diversity Crisis in Software Development. *IEEE Software* 38, 2 (March 2021), 19–25.
- [2] A. Bandura. 1977. Self-Efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review* 84 (1977), 191–215.
- [3] L. Beckwith, M. Burnett, S. Wiedenbeck, C. Cook, S. Sorte, and M. Hastings. 2005. Effectiveness of end-user debugging software features: Are there gender issues?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 869–878.
- [4] N. E. Betz and G. Hackett. 1986. Applications of self-efficacy theory to understanding career choice behavior. *Journal of social and clinical psychology* 4, 3 (1986), 279–289.
- [5] S. Beyer. 2014. Why Are Women Underrepresented in Computer Science? Gender Differences in Stereotypes, Self-Efficacy, Values, and Interests and Predictors of Future CS Course-Taking and Grades. *Computer Science Education* 24, 2-3 (July 2014), 153–192.
- [6] A. Bosu and K. Z. Sultana. 2019. Diversity and Inclusion in Open Source Software (OSS) Projects: Where Do We Stand?. In *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. 1–11.
- [7] N. A. Çakır, A. Gass, A. Foster, and F. J. Lee. 2017. Development of a game-design workshop to promote young girls’ interest towards computing through identity exploration. *Computers & Education* 108 (2017), 115–130.
- [8] L. Cen, D. Ruta, L. Powell, and J. Ng. 2014. Does Gender Matter for Collaborative Learning?. In *2014 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*. 433–440.
- [9] G. Crombie, T. Abarbanel, and C. Anderson. 2000. All-Female Computer Science. *Science Teacher* 67, 3 (2000), 40–43.
- [10] D. A. Fields, Y. B. Kafai, and M. T. Giang. [n.d.]. Youth Computational Participation in the Wild: Understanding Experience and Equity in Participating and Programming in the Online Scratch Community. *ACM Transactions on Computing Education* 17, 3 ([n.d.]), 22.
- [11] D. A. Fields, Y. B. Kafai, A. Strommer, E. Wolf, and B. Seiner. 2014. INTERACTIVE STORYTELLING FOR PROMOTING CREATIVE EXPRESSION IN MEDIA AND CODING IN YOUTH ONLINE COLLABORATIVES IN SCRATCH. (2014), 11.
- [12] G. Fraser, U. Heuer, N. Körber, E. Wasmeier, et al. 2021. LitterBox: A Linter for Scratch Programs. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*. IEEE, 183–188.
- [13] A. Funke and K. 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.
- [14] F. González-Gómez, J. Guardiola, Ó. M. Rodríguez, and M. Á. M. Alonso. 2012. Gender differences in e-learning satisfaction. *Computers & Education* 58, 1 (2012), 283–290.
- [15] I. Graßl, K. Geldreich, and G. Fraser. 2021. Data-driven Analysis of Gender Differences and Similarities in Scratch Programs. In *Proceedings of the 16th Workshop*

⁴<https://github.com/se2p/scratch-online-course>

- in *Primary and Secondary Computing Education (WiPSCE)*. ACM. To appear.
- [16] L. Greifenstein, I. Graßl, and G. Fraser. 2021. Challenging but Full of Opportunities: Teachers' Perspectives on Programming in Primary Schools. In *21st Koli Calling International Conference on Computing Education Research*. 1–10.
 - [17] S. Grover, R. Pea, and S. Cooper. 2015. Designing for deeper learning in a blended computer science course for middle school students. *Computer science education* 25, 2 (2015), 199–237.
 - [18] F. Hermans and E. Aivaloglou. 2017. Teaching Software Engineering Principles to K-12 Students: A MOOC on Scratch. In *2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering Education and Training Track (ICSE-SEET)*. 13–22.
 - [19] P. Hubwieser, E. Hubwieser, and D. 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 (Lecture Notes in Computer Science)*, A. Brodnik and F. Tort (Eds.). Springer International Publishing, Cham, 40–52.
 - [20] T. Jones and V. A. Clarke. 1995. Diversity as a Determinant of Attitudes: A Possible Explanation of the Apparent Advantage of Single-Sex Settings. *Journal of Educational Computing Research* 12, 1 (Jan. 1995), 51–64.
 - [21] C. M. Lewis and N. Shah. 2015. How Equity and Inequity Can Emerge in Pair Programming. In *Proceedings of the Eleventh Annual International Conference on International Computing Education Research*. ACM, Omaha Nebraska USA, 41–50.
 - [22] S. M. Lindberg, J. S. Hyde, J. L. Petersen, and M. C. Linn. 2010. New Trends in Gender and Mathematics Performance: A Meta-Analysis. *Psychological bulletin* 136, 6 (Nov. 2010), 1123–1135.
 - [23] A. Lishinski, A. Yadav, J. Good, and R. Enbody. 2016. Learning to Program: Gender Differences and Interactive Effects of Students' Motivation, Goals, and Self-Efficacy on Performance. In *Proceedings of the 2016 ACM Conference on International Computing Education Research*. ACM, Melbourne VIC Australia, 211–220.
 - [24] J. Margolis and A. Fisher. 2002. *Unlocking the Clubhouse: Women in Computing*. MIT Press.
 - [25] J. McBroom, I. Koprinska, and K. Yacef. 2020. Understanding Gender Differences to Improve Equity in Computer Programming Education. In *Proceedings of the Twenty-Second Australasian Computing Education Conference*. ACM, Melbourne VIC Australia, 185–194.
 - [26] C. McDowell, L. Werner, H. E. Bullock, and J. Fernald. 2006. Pair programming improves student retention, confidence, and program quality. *Commun. ACM* 49, 8 (2006), 90–95.
 - [27] M. Mitchell. 1993. Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of educational psychology* 85, 3 (1993), 424.
 - [28] P. Moorman and E. Johnson. 2003. Still A Stranger Here: Attitudes Among Secondary School Students Towards Computer Science. *ACM SIGCSE Bulletin* 35, 3 (2003), 193–197.
 - [29] S. Papadakis. 2018. Is Pair Programming More Effective than Solo Programming for Secondary Education Novice Programmers?: A Case Study. *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)* 13, 1 (Jan. 2018), 1–16.
 - [30] REDINE. 2019. *Conference Proceedings EDUNOVATIC 2018: 3rd Virtual International Conference on Education, Innovation and ICT*. Adaya Press.
 - [31] M. Resnick, J. Maloney, A. Monroy-Hernández, N. Rusk, E. Eastmond, K. Brennan, A. Millner, E. Rosenbaum, J. Silver, B. Silverman, and Y. Kafai. 2009. Scratch: Programming for All. *Commun. ACM* 52, 11 (Nov. 2009), 60–67.
 - [32] G. T. Richard and Y. B. Kafai. 2016. Blind Spots in Youth DIY Programming: Examining Diversity in Creators, Content, and Comments within the Scratch Online Community. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, Santa Clara, California, USA, 1473–1485.
 - [33] J. Robertson. 2012. Making games in the classroom: Benefits and gender concerns. *Computers & Education* 59, 2 (2012), 385–398.
 - [34] R. Roque, N. Rusk, and M. Resnick. 2016. Supporting Diverse and Creative Collaboration in the Scratch Online Community. In *Mass Collaboration and Education*, U. Cress, J. Moskaliuk, and H. Jeong (Eds.). Springer International Publishing, Cham, 241–256.
 - [35] E. Rubegni, M. Landoni, and L. Jaccheri. 2020. Design for Change With and for Children: How to Design Digital StoryTelling Tool to Raise Stereotypes Awareness. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. ACM, Eindhoven Netherlands, 505–518.
 - [36] M. A. Rubio, R. Romero-Zalaz, C. Mañoso, and A. P. de Madrid. 2015. Closing the Gender Gap in an Introductory Programming Course. *Computers & Education* 82 (March 2015), 409–420.
 - [37] A. Sullivan and M. 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.
 - [38] J. Teague. 2002. Women in computing: What brings them to it, what keeps them in it? *ACM SIGCSE Bulletin* 34, 2 (2002), 147–158.
 - [39] E. L. Usher and F. Pajares. 2008. Sources of Self-Efficacy in School: Critical Review of the Literature and Future Directions. *Review of Educational Research* 78, 4 (Dec. 2008), 751–796.
 - [40] T. Vrieler, A. Nylén, and Å. Cajander. 2020. Computer Science Club for Girls and Boys – a Survey Study on Gender Differences. *Computer Science Education* (Oct. 2020), 1–31.
 - [41] J. Weese, R. Feldhausen, and N. Bean. 2016. The Impact of STEM Experiences on Student Self-Efficacy in Computational Thinking. In *2016 ASEE Annual Conference & Exposition Proceedings*. ASEE Conferences, New Orleans, Louisiana, 26179.
 - [42] K. M. Ying, L. G. Pezzullo, M. Ahmed, K. Crompton, J. Blanchard, and K. E. Boyer. 2019. In Their Own Words: Gender Differences in Student Perceptions of Pair Programming. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. ACM, Minneapolis MN USA, 1053–1059.
 - [43] A. Zeid and R. El-Bahey. 2011. Impact of Introducing Single-Gender Classrooms in Higher Education on Student Achievement Levels: A Case Study in Software Engineering Courses in the GCC Region. In *2011 Frontiers in Education Conference (FIE)*. T2H-1–T2H-6.
 - [44] Z. Zhan, P. Fong, H. Mei, and T. Liang. 2015. Effects of Gender Grouping on Students' Group Performance, Individual Achievements and Attitudes in Computer-Supported Collaborative Learning. *Computers in Human Behavior* 48 (July 2015), 587–596.
 - [45] B. Zhong, Q. Wang, and J. Chen. 2016. The Impact of Social Factors on Pair Programming in a Primary School. *Computers in Human Behavior* 64 (Nov. 2016), 423–431.