Remote Pair Collaborations of CS Students: Leaving Women Behind?

Caroline Lott
University of Tulsa
Tulsa, OK, United States
ctl2149@utulsa.edu

Alexander McAuliffe
University of Tulsa
Tulsa, OK, United States
ajm8334@utulsa.edu

Sandeep Kaur Kuttal
University of Tulsa
Tulsa, OK, United States
sandeep-kuttal@utulsa.edu

***Abstract*—Remote pair programming research indicates benefits for CS students, increasing productivity, code quality, teamwork, knowledge management, and morale. The COVID-19 pandemic increased the prevalence of remote pair programming. Gender gaps persist in CS classes and workplaces, which may negatively impact the way pairs coordinate, communicate, and collaborate. To understand these effects, we conducted a large-scale survey to investigate differences between men and women as well as same- and mixed-gender pairs. The survey questions were adapted from established literature on gender differences in the fields of education, communication, management, human-robotic interaction, and human-computer interaction. Quantitative analysis of the survey data using ANOVA and pairwise t-tests indicated that women participants reported their men partners made gender-based assumptions about them, and felt dominated and interrupted with men partners. Men participants felt their men partners were more rude and gave more negative feedback than women partners. Further, qualitative analysis of interviews gave insights to several challenges CS students face in same and mixed-gender pairs when programming remotely. Our findings have implications for researchers, practitioners, and educators to promote gender inclusivity in collaborative environments.**

I. Introduction

Software development is a complex problem-solving task that requires appropriate engineering practices such as pair programming, an agile framework that organizes people to produce high quality software more productively [8]–[11]. When two programmers collaborate by pair programming, one partner is the driver (writes the code) and other is the navigator (reviews the code), switching roles after a period of time [12], [13]. Global software development and the COVID-19 pandemic have forced CS students to collaborate remotely. Research has found remote pair programming to offer similar advantages as in person pair programming, such as improving code quality, productivity, teamwork, performance, knowledge management, and morale [14]–[24].

Pair collaboration, critical for a software project's success, is affected by programmers' time zones, cultures, races, and genders [25], [26]. This paper focuses on gender effects, where gender is an individual's self-identification [27], [28].

Gender gaps persist in Computer Science (CS) classes and workplaces [29]. Women represent only 18% of CS undergraduate students and 10% of CS jobholders [30]. In addition, some communication software has implicit gender biases in its interface design [31]–[34], which may influence remote pair programming dynamics. Further, established research indicates that genders differ in problem solving [35], [36], creativity [37]–[39], communication [6], [40]–[43] and leadership [6], [44]–[56] styles. These gender differences are relevant to remote pair collaborations.

To understand these gender differences, we examined how coordination, communication, and collaboration, the key enablers of software development [57]–[60], are affected by gender dynamics in pairs of remote programmers. Hence, we formulated three research questions:

- **RQ1: How do same- and mixed-gender pairs of CS students coordinate remotely?**
  Remote pair coordination can be influenced by the leadership and interruption style of CS students. Studies in management and psychology have reported gender differences in both leadership [44]–[56] and interruption styles [61]–[63]. Hence, we studied these two subtopics.
- **RQ2: How do same- and mixed-gender pairs of CS students communicate remotely?**
  Distributed software development projects take 2.5 times longer than co-located projects due to communication challenges [40]–[43], [64]. Gender differences in CS students' communication styles adversely affect remote collaboration. These differences are well studied in the fields of human-computer interaction, education and communication [65]–[69]. We studied the following four subtopics related to communication: the purpose of communication (to solve problems or build rapport), communication preference (remote or in person), primary communicator role (driver or navigator), and nonverbal communication style [6].
- **RQ3: How do same- and mixed-gender pairs of CS students collaborate remotely?**
  Research on teams of CS students collaborating remotely suggests that productive interaction depends on developing trust [70]–[74] and awareness [75]–[79]. We studied these two subtopics, as gender differences in these may affect remote pair programming.

Our survey adopted well-researched measures from the

978-1-6654-4592-4/21/$31.00 ©2021 IEEE

TABLE I
SURVEY QUESTIONS DESIGN

RQs	Survey Category	Sample Question (on a 5-point Likert Scale Ranging from Strongly Disagree to Strongly Agree)
Coordination	Leadership style [1]–[4]	*"My pair programming partner (Buddy) tended to dominate our conversations."*
	Interruption style	*"I frequently interrupted my partner (Buddy) while they were talking."*
Communication	Communication preference	*"I wished that my pair programming partner (Buddy) and I could have coded together in person."*
	Nonverbal cues [4]	*"I communicated my ideas using nonverbal cues more often than verbal expressions."*
	Primary communicator role (driver/navigator) [5]	*"When programming with my partner (Buddy), I usually typed the code while my partner directed."*
	Purpose of communication (to solve problems or build rapport) [1], [5]	*"Most often, I communicated with my partner (Buddy) in order to be friendly and get to know them."*
Collaboration	Gender awareness [6]	*"I think my pair programming partner (Buddy) assumed my skill level based of my gender."*
	Trust [1], [2], [5], [7]	*"My pair programming partner (Buddy) had productive and helpful interactions with me."*

fields of education, communication, management, human-robotic interaction, and human-computer interaction. We modified the questions based on expert feedback and pilot studies. The final survey garnered responses from 98 students and 9 professionals. We statistically analyzed survey data to investigate differences between same- and mixed-gender remote programming pairs. Further, we triangulated these results by interviewing 9 survey participants.

II. Related Work

Pair programming research has investigated gender differences in co-located pairs [80], [81]. The most comparable research to ours is a multiple-room lab study with six pairs of same- and mixed-gender CS students [6]. They identified coordination, communication, and collaboration hindrances to remote pair programming. Our study adds to their findings by conducting a large-scale survey followed by interviews.

Abramo et al. analyzed women's academic research collaborations in intramural, extramural, domestic and international settings, and found gaps between women's and men's capacity to collaborate internationally [82].

In remote environments, Miller, et al. reported that productivity may be affected by pair communication and satisfaction with social interactions [83], and Ralph et al. found that women, parents and people with disabilities may be disproportionately affected by remote programming challenges [84]. Schiller et al. reported that remote task performance affected only satisfaction in women pairs, but both trust and satisfaction in men pairs, and that impression management affected women pairs' satisfaction and trust more than men pairs' [85]. When communicating on social networks, Shen et al. found that men are more affected by attitude, positive anticipated emotions and group norms, whereas women are more affected by social identity and negative anticipated emotions [86].

III. Survey

A. Item Generation

The first step was to conduct a literature review, from which we established the eight categories (ref Table I, Column 2) to formulate the survey. The survey also included 10 background questions that asked participants for their gender and the gender they perceived their pair programming partner to be. In this study, we use "man" and "woman" as adjectives to describe the self-identified genders (as in [27], [28]) of the participants and the perceived genders of their programming partners. We do not use the adjectives "male" and "female," as the genders of participants and their partners are not necessarily the sexes they were assigned at birth.

Our initial survey draft contained 83 items inspired by measures from the areas of education, communication, management, human-robotic interaction, and human-computer interaction. We adapted survey items from Rovai et al. [1] that measured connectedness and student learning in online environments to formulate measures of trust, leadership style, and the purpose of communication. We adapted survey items related to the primary communicator role, purpose of communication, and trust from Goodboy et. al's [5] student communication satisfaction scale that investigated student-teacher relationships. To measure leadership style and trust, we adapted questions from Law et al. [2], who investigated the emotional intelligence of robots and human trust in robots. Leadership questions were also adapted from Alaloul et. al's study [3] that measured the coordination between construction workers on team projects. Some leadership and nonverbal communication style questions were adapted from Kyungsub Choi's study on interactions between different genders in pair programming collaborations [4]. We used Ku et. al [7] to develop measures of trust, as their survey measured the teamwork satisfaction and student attitude of graduate students collaborating in online environments. Similar to Ku et. al [7], we utilized a 5-point Likert scale.

To adapt the questions, we changed the language to reflect pair programming collaborations. For example, we modified Law et. al's [2] question, *"I would be comfortable giving the robot a task or problem which was critical to me, even if I could not monitor their actions"* by replacing *"the robot"* with *"my partner (Buddy)."* Using survey design framework [87], we formulated survey questions related to gender awareness, communication preferences, and interruption styles based on prior gender difference findings [6].

Table I, Column 3 shows example questions used in our survey. Survey items referred to each participant's pair programming partner as "Buddy" to remind them to share experiences with the same partner they reported in the background questionnaire.

B. Content Review

We used an iterative approach to design, develop, and refine the survey questions [88].

1) Expert Analysis: The survey was evaluated for completeness, depth, maturity, and organization by seven researchers and professionals with experience in gender research and collaboration. Four gender research experts reviewed our initial 83 survey items. The experts had 20 years (1), 5 years (1), and 1 year (2) of experience and had published gender research for programming domains. The three professionals with collaboration experience commented about the collaboration aspects of the survey. Overall, experts commented on the content and phrasing of survey items and on the survey's size and format. Their feedback led us to change our survey in several ways. For example, we reduced the survey from 83 to 46 items by combining similar and redundant questions, because our experts advised we shorten the survey. Additionally, in the leadership section, the items *"My pair programming and I both participated in making plans for the project"* and *"My pair programming partner and I shared the same goals for our project"* were very similar statements, so we decided to only keep the first one. Further, we clarified vague items per expert feedback, such as changing the wording of one item, *"Most often, I communicated with my partner in order to build rapport with them"* by replacing *"build rapport with them"* with *"be friendly and get to know them as a person."* Another expert suggested adding a short answer option that stated, *"In 1-2 sentences, please explain your reasoning for your answers."*

2) Pilot Studies: We conducted four pilot studies with students from our personal contacts. On average, participants took 12.5 minutes to complete the survey, and we made several changes to survey wording and coherence based on their feedback. For example, a participant suggested that the wording of a question (*"My pair programming partner was brusque or abrasive with me"*) was unclear. Hence, we changed the item to *"...was rude with me."*

C. Study Design

We recruited participants by sending emails to personal contacts, posting on social media (Facebook, Instagram, and LinkedIn), and using snowball sampling. We recruited student participants by sending emails to fifteen professors from personal contacts to share the survey with their CS students. We utilized snowball sampling, encouraging our contacts to recruit participants from universities and workplaces. We collected data for four weeks, resulting in 98 student and 9 professional participants. The aim of recruiting professional participants was to understand gender differences among CS jobholders. Due to low response rate among professional participants, we combined the data with student participants and do not report differences between the two groups.

Of the 107 survey respondents, 2 professional and 17 student participants indicated they had no pair programming experience on the background questionnaire, hence they were unable to access the full survey. The number of participants with remote pair programming experience was 88, comprised of 33 women, 54 men, and one non-binary person. The average participant was 24 years old, had 3.9 years of programming experience, and 3.2 years of pair programming experience. Participants were from the United States, Canada, Germany, Libya, China, India, and Norway.

Participants were given the option to be entered into a raffle of Amazon gift cards. 32 student participants were given \$10 and 3 professional participants were given \$15.

D. Data Analysis

The survey's multiple choice questions offered a 5-point Likert scale. We quantified participants' multiple choice responses by coding "Strongly Disagree" as -2, "Disagree" as -1, "Neutral" as 0, "Agree" as 1, and "Strongly Agree" as 2. Since 83 out of the 84 participants self-identified as men or women, we grouped survey data into four categories based on participants' self-identified gender and the perceived gender of their partner: Woman-Woman (WW), Woman-Man (WM), Man-Woman (MW), and Man-Man (MM). The gender listed first is the survey respondent's gender, and the gender listed second is the gender of their partner, "Buddy."

To analyze survey response data, we applied the Kruskal-Wallis ANOVA test, a non-parametric test that does not assume normal distribution of data or homogeneity of variances, to every question. The questions that rejected the null hypothesis for ANOVA were tested with a pairwise t-test. The six pairs tested were: (MM,MW), (MM,WW), (MM,WM), (MW,WW), (WW,WM), and (MW,WM). We use p=0 for values $p<0.0001$. We calculated each gender-pairing group's average response to the measure based on the -2 to 2 coding scale. Our scale's internal consistency was high (Cronbach's α = 0.84), indicating our items were closely related.

IV. Interviews

We conducted semi-structured interviews to understand survey data trends and to explore participants' opinions, behaviors, and experiences when remote programming in same- and mixed-gender pairs. The interviews helped triangulate (compare and attempt to refute) survey results.

A. Participants

Interview participants were recruited from the pool of survey participants. The survey allowed participants to indicate their interest in a 30-minute semi-structured online interview with a researcher. We reached out to 16 student and 2 professional participants. A total of 9 student participants completed interviews with a researcher in exchange for a \$15 Amazon gift card. The interviews were audio-recorded and conducted on the participant's digital platform of choice (Microsoft Teams (1) and Zoom (8)). We interviewed two participants from each gender pairing

(WW, WM, MW, MM), and one non-binary participant whose programming partner was a woman (NBW). So, four of our participants were men, four were women, and one was non-binary. We will refer to these participants using abbreviations, with MMP2 representing the second man participant who pair programmed with a man partner (Buddy).

B. Design

Generic interview questions were asked to all interview participants [89]. We formulated questions based on participants' survey responses, so questions were a reflection of their reported experiences. Interviews were scheduled to last 30 minutes.

C. Data Analysis

We transcribed 186 minutes of audio recordings from the 9 interviews. The transcripts were compared with trends found in the survey responses and prior research. Thematic analysis [90] was used to organize the interview responses into themes that relate back to the research questions.

V. Threats to Validity

External Validity: The results can not be generalized beyond university students due to the low response rate from professionals. Our results can not be generalized to other cultures, timezones, and individual differences. Though, the same and mixed gender pairs exist in cultures (Malaysia) where women are more represented in software industries.

Internal Validity: Due to the social desirability bias to show oneself in a favorable light, the survey and interview responses may have been biased. The responses may have also been affected by participant fatigue or inattention; however, we tried to mitigate these effects by limiting the length of the survey to 12.5 minutes.

Construct Validity: We restricted the responses to Likert-type questions, though we added short answer options. To avoid the problem of subjective interpretation, we received careful content review from experts, software developers, and students.

Conclusion Validity: Our survey questionnaire may not measure all the subcategories for coordination, collaboration and communication. We took special measures to include research questions as well as subcategories from literature and evaluated the question from other researchers.

VI. Results

Table II shows the survey responses with significant differences between groups using ANOVA for $\alpha = 0.05$. The averages of survey results indicate differences between same-gender (MM, WW) and mixed-gender (MW, WM) pairs.

A. RQ1: Coordinate

1) Leadership Style: Members of mixed-gender pairs seem likely to run into differences in leadership styles [44]–[56]. We compared leadership style among the pairs, finding:

Women participants felt dominated by men partners in conversation. Participants reported to what degree their partners dominated conversation while pair programming remotely. Analyzing the results with a pairwise t-test, we found significant differences between MM and WM ($p = 0.027$) and WW and WM ($p = 0.010$). Hence, women participants felt more dominated in conversations with men (WM avg = 0.381) than women partners (WW avg = -0.583). Men participants did not feel dominated by their programming partners (MW avg = -0.167 and MM avg = -0.233), irrespective of their gender.

WMP2 reported pairing with a man who *"perceived himself to have more experience than I did...he was dominating conversations...it seemed like he saw himself as the leader of the project."* Similarly, NBW related their experience: *"men, trans men, cis men...tend to be assertive most of the time. Women and feminine folks can tend to be a little more passive."* These results align with prior management research [56], [91] where statistical analysis indicated men tended to be more dominant (authoritative) than women (democratic).

Women participants received more constructive feedback from men partners. We asked participants in the driver role to what degree their partner gave them constructive feedback. A pairwise t-test showed significant differences between MM and WW ($p = 0.006$), MM and WM ($p = 0.009$), and MW and WW ($p = 0.019$). Hence, women perceived more constructive feedback from men (WM avg = 1.048) than women partners (WW avg = 1.333), while men reported less constructive feedback from both men (MM avg = 0.667) and women (MW avg = 0.708) partners.

WMP2 received sufficient help from her partner: *"He'd give me advice about...certain techniques to use, or different little tricks to use, and I remember we were programming in a new IDE that he had already used, so he...was really helpful."* alternatively, MMP1 sometimes received little support from his partner, explaining, *"if...the person has not really went to the lectures before, then sometimes they can't contribute anything because they weren't there, and then [I'm] doing all the work."* These findings may indicate that women benefit more from pair programming than men as found in [92].

Women participants asked their partners for help more often than men. We asked participants in the driver role how often they asked their partners for help. A pairwise t-test found significant differences between MM and WW ($p = 0.012$), MM and WM ($p = 0$), MW and WW ($p = 0.004$), and MW and WM ($p = 0$). This indicates that men participants asked for less help (MW avg = 0.292, MM avg = 0.433) while women participants, irrespective of their partner's gender, asked for more help (WW avg = 1.083, WM avg = 1.238).

MMP1 explained that he is not quick to ask for his partner's input. If his partner's code *"doesn't make sense at all and it has more syntax errors, that's more work for [me] to clean the code than to write it [myself...] In some courses, I did everything myself because of that."* On the other hand, WMP2 mentioned, *"It may even just be a subconscious bias I have, thinking that other people are more qualified than me...things, like, speaking up more in class,...being confident...makes me perceive them to be more experienced."*

TABLE II
SURVEY QUESTIONS WITH AVERAGES AND SIGNIFICANT DIFFERENCES (USING ANOVA)

RQs	Survey Category	Survey Question	Gender Composition (Averages)				ANOVA
			WW(12)	**WM**(21)	**MW**(24)	**MM**(30)	p-value
Coordination	*Leadership*	Partner dominated conversations	-0.583	0.381	-0.167	-0.233	0.016
		Partner gave constructive feedback	1.333	1.048	0.708	0.667	0.008
		Asked partner for help when driving	1.083	1.238	0.292	0.433	<0.001
	Interruption	Partner interrupted frequently	-0.917	0.381	-0.625	-0.633	<0.001
Communication	*Comm. Preference*	Experienced difficulty with online communication	-0.583	0.095	0.417	0.333	0.022
	Nonverbal Cues	Partner communicated with nonverbal cues	-0.750	-0.810	0.042	-0.467	0.019
	Primary Comm.	Generated more ideas than partner	0.000	-0.143	0.583	0.300	0.033
	Purpose of Comm.	Partner communicated primarily to work on code	0.667	1.143	0.042	0.567	<0.001
Collaboration	*Gender Awareness*	Partner made gender-based assumptions	-1.750	-0.143	-0.792	-0.600	0.005
		Made gender-based assumptions about partner	-1.750	-0.714	-1.042	-0.500	0.005
	Trust	Partner was rude	-1.833	-0.952	-1.083	-0.933	0.049
		Partner gave hurtful feedback	-1.750	-0.952	-0.750	-0.900	0.028

This indicates that women may have a lower self-efficacy [93]–[102] than men, resulting in them asking for more help.

These findings support a prior study [103] which found many men recognized a need to help women, but were not as open to receiving help as women.

2) Interruption Style: Prior research suggested that in mixed gender settings, men tend to interrupt people more often than women [61]–[63]. Our data supports this idea.

Women participants were most frequently interrupted by men partners. We asked participants how frequently they were interrupted by their programming partner while talking. A pairwise t-test showed significant differences between MM and WW (p = 0.049), MM and WM (p = 0.002), WM and WW (p = 0.001), and WM and MW (p = 0.006). Hence, women in the driver role did not feel interrupted by their women partners (WW avg = -0.917), whereas they felt interrupted by men partners (WM avg = 0.381). Irrespective of their partner's gender, men participants did not feel interrupted overall (MM avg = -0.633 and MW avg = -0.625).

WMP1 explained interruptions from her man partner, *"I start to say something and he would break in in the middle and start talking...and I realized this is not...gonna work well for us."* MMP2 narrated an experience programming in a group, *"The guy...I feel like...interrupt[ed] me a little more often. The girl...didn't really interrupt that much."*

These findings are similar to prior studies [61], [63]; interruptions can lead to frustration, misunderstanding and trust issues, especially among mixed-gender pairs [62].

B. RQ2: Communicate

1) Communication Preference: We investigated preferences for remote environments in the context of gender.

Women participants found the least difficulty communicating with women partners; men found the most difficulty communicating with women. We asked participants to what degree they experienced difficulties communicating remotely with their partners. A pairwise t-test showed significant differences between WW and MM (p = 0.006) and WW and MW (p = 0.007). Women participants reported more lapses in communication with men than with women partners (WW avg = -0.583, WM avg = 0.095). Similarly, men participants communicated slightly better with men (MM avg = 0.333) than women partners (MW avg = 0.417).

WMP1 preferred working online, expressing, *"I like having my own space to think without having someone breathing over my neck, because it gives me my own time to come up with the solutions."* Alternatively, most participants in their interviews elaborated their frustrations about remote programming environments (details in Section VII-C).

2) Nonverbal Communication: Communication depends on the words being exchanged and on the context, the sender of the nonverbal behavior, the receiver of that nonverbal behavior, the relationship between the sender and receiver, and the arrangement of other nonverbal cues [65].

Men participants perceived more nonverbal cues from women partners. We asked participants how often their partner used nonverbal cues. A pairwise t-test showed significant differences between MW and WW (p = 0.001) and MW and WM (p = 0.007). Hence, the data indicate that men participants perceived more (MW avg = 0.042) nonverbal cues from women partners than women participants perceived from men partners (WM avg = -0.810).

MMP2 explained, *"I feel like females will probably use more...body language, or maybe just emphasize different features of the body versus men."* Additionally, NBW found that *"men...don't use as many nonverbal cues, whereas women don't feel that they can be as direct."* WWP1 added, *"I like to communicate with verbal cues more with a man than I do with a woman."* Correspondingly, WWP2 explained, *"If I'm partnered with a man...I'm pretty direct because that seems to be very successful in getting my opinions heard."*

Women express more nonverbal behaviors, are more skilled at sending and receiving nonverbal messages [66], [67], and use body language differently than men [68]; hence, restricted nonverbal communication may contribute to remote communication lapses in mixed-gender pairs.

3) Primary Communicator Role: We investigated how participants perceived the roles of driver and navigator.

Both men and women participants reported that men partners generated more ideas. Participants were asked to report whether they felt they contributed more ideas for the programming project than their partners. A pairwise t-test showed significant differences between MW and WM ($p = 0.004$). Further, men participants reported generating more ideas for the project's code with a woman partner (MW avg = 0.583) than with a man partner (MM = 0.300). Women participants felt less involved in generating ideas with a man partner (WW avg = 0.000, WM avg = -0.143).

During participant interviews, we did not find insights on men generating more ideas but found that more experienced partners were usually designated as drivers. MWP1 reported it *"depended on if we were more comfortable with the assignment, then one person would take the lead [driver role]."* Regarding the distribution of work, he added, *"[on] weeks where I would have more of the previous knowledge or more insight,...I would do a lot more work, but there were also other weeks where my partner would do more work."* Further, the communicator role depended upon trust, as WMP2 mentioned that with her man partner, *"I knew they had some amount of experience just from talking with them in previous interactions, so I trusted them in both roles."*

4) Purpose of Communication: We analyzed what men and women tried to accomplish in their communication.

Women participants felt men partners communicated mostly to work on code, while men reported that women partners didn't. We asked participants if their pair programming partner primarily communicated with them in order to work on the code. A pairwise t-test showed significant differences between MM and MW ($p = 0.02208$), MM and WM ($p = 0.01536$), and MW and WM ($p = 0$). Men participants communicated with women partners about code (WM avg = 1.143) and perceived their women partners to communicate less about code (MW avg = 0.583).

In her interview, WWP1 corroborated this finding, reporting, *"I'd definitely say I talk a lot more with women than men."* These results are supported by a recent study [69] whose statistical analysis indicated that more men than women approached communication with a clear purpose, a point to be made, and a problem to be solved. In contrast, more women communicated as an act of sharing and an opportunity to build rapport.

C. RQ3: Collaborate

1) Gender Awareness: Research with remote software development teams shows that awareness of temporal, geographical, and socio-cultural differences is important to teams' trust, relationships, and efficiency [76]–[79]. We expect that awareness of gender differences may influence pairs' success, as seen in past research [6].

Women noticed less gender-based assumptions about their skill level from women compared to men partners. We asked participants how much they felt their partner made assumptions about their skill level based on their gender. A pairwise t-test showed significant differences between WW and MM ($p = 0$), WW and MW ($p = 0.002$), and WW and WM ($p=0.007$), indicating that the least skill assumptions occurred in WW pairs. Women participants felt less judged by women (WW avg = -1.750) than men partners (WM avg = -0.143). Men participants felt more judged by men (MM avg = -0.600) than women partners (MW avg = -0.792).

WMP1 expressed, *"I feel weird going into this field, because it's new to me as a girl, and I'm curious about how I'm going to be perceived...as a female in a heavily male-dominated industry."* One of WWP1's man programming partners *"assumed... that I hadn't had any coding knowledge coming into school at all because I was a girl."* WMP2 echoed this sentiment, describing an experience with a man partner: *"It seemed like he thought I didn't have as much as experience as him...I think that he perceived me as less qualified."*

Women participants made more assumptions about the skill level of men than women partners; men made more assumptions about women than men partners. We asked participants how much they assumed their partner's skill level based on their gender. A pairwise t-test showed significant differences between WW and MM ($p = 0$), WW and MW ($p = 0.009$), and WW and WM ($p = 0.002$) indicating that women were least likely to judge or be judged by women partners. Men participants judged men more heavily than women partners (MM avg = -0.500, MW avg = -1.042). Similarly, women participants also judged men more heavily than women partners (WW avg = -1.750, WM avg = -0.714).

WMP2 described patterns in conversations with partners of different genders, *"Working with females I'm at a more equal platform, whereas when I work with males...sometimes it feels like they're taking control over a project."* Additionally, WWP1 preferred to work with women partners, explaining, *"I feel like I interact more comfortably with a woman and I would be less intimidated."*

2) Trust: Global software development studies suggest that trust contributes to the success of distributed software teams [70]–[72]. Building trust between individuals is extensively studied in remote collaborations (e.g., [73], [74]). In interviews, participants mentioned that trust in their partner (1) decreased if their partner wrote messy code, (2) depended on the importance of the task, (3) was influenced by the pair's familiarity, (4) depended on their knowledge of their partner's skill level, (5) increased with partners who showed vulnerability, and (6) increased when a partner meticulously checked their own work.

Women participants felt that women were less rude than men partners, and men reported the same. We asked the participants to report the degree to which they felt their programming partner was rude to them. A pairwise t-test found significant differences between WW and MM ($p = 0$), WW and MW ($p = 0.002$), and WW and WM ($p = 0.007$), indicating that groups with men result in more perceived rudeness. Less rudeness was experienced in WW pairs (WW avg = -1.833) than MM pairs (MM avg = -0.933). Similarly, men were perceived as rude by women partners (WM avg = -0.952) while women were perceived as less rude by men partners (MW avg = -1.083).

MWP2 experienced more rudeness coding with a man partner than with a woman, explaining, *"I don't think he meant to be harsh, but he was definitely much harsher*

TABLE III
OVERALL DIFFERENCES BETWEEN WOMEN AND MEN

RQs	Survey Category	Women Participants Felt	Men Participants Felt
Coordination	*Leadership*	dominated by men partners	X
		they got constructive feedback (irrespective of gender)	they didn't get constructive feedback
		they asked for help (irrespective of gender)	they didn't ask for help
	Interruption	interrupted by men partners	X
Communication	*Comm. Preference*	X	lapses in communication with women partners
	Nonverbal Cues	X	more nonverbal cues with women partners
	Primary Comm.	they had less ideas than men partners	they had more ideas than women partners
	Purpose of Comm.	men partners communicated to work on code	women partners communicated less about code
Collaboration	*Gender Awareness*	judged by men partners	X
		men partners made assumptions about their skills	X
	Trust	women partners were not rude	men partners were rude
		less negative feedback from women vs. men partners	more negative feedback from men vs. women partners

than my woman friend [who] I programmed with." NBW preferred not to work with men due to perceived rudeness, expressing, *"it felt like they were being condescending...I would rather program with not cis-men."* This may be due to women's tendency to be more empathetic [104], which could lead women to perceive rude behaviour as more inappropriate than men. But, interview respondents of all genders perceived rudeness from men partners.

Women participants perceived less negative feedback from women than men partners. We asked participants to what degree they received negative feedback from their partners. A pairwise t-test showed significant differences between WW and MM ($p = 0$), WW and MW ($p = 0.0003$), and WW and WM ($p = 0.016$), indicating that less negative feedback was perceived in WW pairs (WW avg = -1.750) compared to other pairings (WM avg = -0.952). Men perceived less negative feedback from women (MW avg = -0.750) than men partners (MM avg = -0.900).

In her interview, WWP1 recalled coding with a man partner who was *"more unwilling to compromise because they felt like the way they were doing it was the correct way."* Correspondingly, WMP2 explained that her man partner *"seemed like he wanted to do it his way."* These results are similar to an in-person pair programming study of CS students, in which many women participants didn't feel "comfortable" working with men [81].

VII. DISCUSSIONS

A. Insights on Gender

1) Gender Differences: Table III displays gender differences based on our survey data. Highlights show where women (peach) and men (blue) participants faced difficulties pair programming. These may result in misinterpretations of how their partner coordinated, communicated, and collaborated.

2) Experience of Non-binary Gender: Discussions related to non-binary genders contribute to a growing base of literature that goes beyond two genders. In their interview, NBW appeared more comfortable talking about gender differences and sharing their experiences than other interview participants. While other interviewees elaborated little on responses about gender differences, NBW seemed eager to address *"that obviously gender roles in society play a big, big role"* in nonverbal communication. They brought up emotionally charged topics like *"toxic masculinity...instilled in [men]."* They specifically mentioned a sense of feeling excluded and unwelcome in their pair programming endeavors due to their gender, explaining, *"It's an implicit feeling...I get...the underlying sense of feeling 'othered'."*

B. Reducing Gender Biases

1) Gender-Inclusive Pedagogical Material: Despite efforts to fund the recruitment, retention, and advancement of women, gender biases in education systems and workplaces still exist [105]. Education research has found that gender stereotypes and men-focused curricula are often prevalent in schools [106]–[108]. WWP2 reported that the CS department at her university has *"a lot of female professors, and their opening statement on their syllabus and in the first class is very much about equity...[which] lays a lot of groundwork for our partnerships."* Based on interviewees' personal experiences in programming spheres with different gender ratios, it appears that CS programs emphasizing gender equality may result in less gender biases. Further research can be done on the way CS programs' emphasis on gender equality affects students' inter-gender judgement.

2) Keep Gender Balance in CS Classes and Workplaces: Both education and industry recognize the significance of gender balance and are working to address the gender gap, changing work and learning environments and curricular and pedagogical practices [109]. MWP1 reported that these efforts decrease bias: *"I [haven't] encountered that experience at all (gender bias). The computer science program,...it was 50% male, 50% female, so you're basically bound to have male and female partners...gender disparity was super equal."* At a different institution, however, NBW explained, *"It was never explicit...but...the way the professor treated me versus the five other men in that course...it was very clear who the professor thought was the smart one."* These interviews support the idea that investments in gender balance, such as manager and team member training, can be impactful.

3) Gender-Inclusive Interface Design: Research suggests that programmers interact differently with software features based on their gender; however, many software features favor men [31]–[34]. One way to address gender differences in software is through the application of GenderMag [110], [111], an inspection method using specialized cognitive-walkthroughs with research-based personas that reflect gender differences. GenderMag primarily supports single-user problem-solving activities involving technology, and therefore does not help two diverse people navigating collaborative activities, especially remotely. The complications caused by gender differences found in our study suggest a need to develop a method like GenderMag that supports two-user problem-solving activities involving software technology in order to identify gender biases in remote pairs.

4) Facilitator Agents to Support Gender-Inclusivity: Group facilitator agents (bots who listen to human conversations and provide feedback) [112]–[122] equalize participation, clarify processes, resolve conflict, and encourage idea development [113]. Similar agents designed for gender differences may reduce gender gaps by informing and educating users about gender biases in their voice, facial expressions, body language, and conversations, similar to results in Table III.

C. Remote Technologies Hindering Software Development

Remote programming tools are often used in global software development (GSD). The interviews revealed the following hindrances of remote collaborations, which may have implications for GSD research.

1) Limited Nonverbal Cue Exchange: Nonverbal cue expression is hindered in remote environments due to camera and interface limitations. In text-based interfaces, WMP1 found *"emojis help to convey the context,"* as *"[in] text-based...interface,...you don't really get the context cues."*

Although emojis and cameras attempt to compensate for limited nonverbal cues, they may increase the cognitive load on developers. NBW indicated, *"in person with someone, if they're getting annoyed,...impatient,...wandering off,...I feel that. But on Zoom,...you have to work a lot, lot harder to get those nonverbal cues, and that can cause a lot of fatigue."* They also added that on Zoom, they speak *"a lot louder...my inflections are a little more exaggerated...it's...more fatigue-inducing to get across what [I'm] feeling."*

Remote communication has significant effects on women [66], [67] and people from cultures that use more nonverbal cues, as MWP2 said, *"I am Italian, and I feel like I rely on that [nonverbal cues] a bit more than some people."*

Other remote methods besides video calling such as virtual [123], [124] or augmented reality should be explored further as alternatives for remote pair programming.

2) Inconvenience with Multiple Softwares: Complexity is added when programmers use remote software to solve coding problems. As MWP1 said, *"it's a lot more inconvenient online, especially when you're using one platform to video call, and another platform to code, and another platform to have a virtual whiteboard or blackboard."* In-person, he explained he *"can draw things out, or have a whiteboard or a chalkboard in the room"* and hence *"ideas just flow more quickly."* In GSD, sketch-based whiteboard apps like Explain Everything [125], RealtimeBoard [126], and LiveBoard [127] as well as software design tools supporting informal (eg., Calico [128], [129], Flexisketch [130]), formal (e.g., designing UML class diagrams on touch-screens [131]), or both (e.g., OctoUML [132]) designs exist, but these findings suggest a need for integration in IDEs to use the apps' full potential.

3) Trouble Resolving Hardware Issues: Remote development is further complicated by hardware difficulties. NBW expressed frustration with this limitation, *"When you're remote, you can only see the code side of things, you can't do the troubleshooting for the hardware side of things, which made it 30 billion times more difficult."* This issue may apply to GSD, as remote programmers may experience hardware issues difficult for their partner to diagnose or replicate.

4) Lag in Text-Based Interfaces: Some interviewees experienced difficulty communicating with their partner using a text-based interface. WMP1 expressed, *"on Slack [app]...text-based interface, that's a little challenging because...you can start typing something and someone else starts typing something in the middle of it, and there's that delay between... when it's received. And so sometimes we'd have two or three threads going at the same time."* Inherent lags in text-based interfaces can hinder communication for all genders.

5) Difficulty Role Switching with Screen Share Apps: Sharing ideas using screen share apps can be frustrating when switching roles, as MMP1 mentioned, *"you can't really work together that well...If someone has an idea and he just wants to [share]...but...looking at your screen over the screen share,...it's kind of annoying that you can't do it online but in person you can switch very easily."* This can hinder communication for all genders, and for those in navigator roles who do not want to interrupt their partner [6].

VIII. Conclusions

We identified hindrances women and men CS students face while working remotely in pairs, drawing from prior interdisciplinary research on gender differences and our empirical results. This paper makes the following contributions: (1) Our survey questionnaire and interview questions are available online for reproducibility by researchers and practitioners [133]. These can help statistically measure gender differences in collaborations of CS students or end-user programmers; (2) We report statistical differences between women and men participants as well as same- and mixed-gender pairs. This may assist in improving pedagogical techniques and developing gender-inclusive software; (3) Our focused design, evaluation, and refinement cycles to formulate survey questions can be utilized by researchers for collaborative problem-solving tasks in other domains such as requirement gathering, designing, and testing; and (4) We report limitations of text-based, video-based, and remote communication software. This research aims to bridge the gender gap among CS student pairs and contribute to literature on pair programming. The awareness of gender differences has the potential to reduce stereotypes and increase productivity in pair collaborations.

REFERENCES

[1] A. P. Rovai, "Development of an instrument to measure classroom community," *Internet and Higher Education*, vol. 5, pp. 197–211, 2002.

[2] T. Law, M. Chita-Tegmark, and M. Scheutz, "The interplay between emotional intelligence, trust, and gender in human–robot interaction: A vignette-based study," *International Journal of Social Robotics*, vol. 13, 04 2021.

[3] W. Alaloul, M. Liew, and N. A. Wan Abdullah Zawawi, "Identification of coordination factors affecting building projects performance," *Alexandria Engineering Journal*, vol. 55, 06 2016.

[4] K. Choi, "A comparative analysis of different gender pair combinations in pair programming," *Behaviour Information Technology*, vol. 34, pp. 1–13, 12 2014.

[5] A. Goodboy, M. Martin, and S. Bolkan, "The development and validation of the student communication satisfaction scale," *Communication Education - COMMUN EDUC*, vol. 58, pp. 372–396, 07 2009.

[6] S. Kaur Kuttal, K. Gerstner, and A. Bejarano, "Remote Pair Programming in Online CS Education: Investigating through a Gender Lens," in *2019 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, Oct 2019, pp. 75–85.

[7] H.-Y. Ku, H. W. Tseng, and C. Akarasriworn, "Collaboration factors, teamwork satisfaction, and student attitudes toward online collaborative learning," *Computers in Human Behavior*, vol. 29, no. 3, pp. 922–929, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563212003597

[8] E. Arisholm, H. Gallis, T. Dyba, and D. I. K. Sjoberg, "Evaluating pair programming with respect to system complexity and programmer expertise," *IEEE Transactions on Software Engineering*, vol. 33, no. 2, pp. 65–86, Feb 2007.

[9] M. Celepkolu and K. E. Boyer, "Thematic analysis of students' reflections on pair programming in cs1," in *Proceedings of the 49th ACM Technical Symposium on Computer Science Education.* ACM, 2018, pp. 771–776.

[10] J. Chong and T. Hurlbutt, "The social dynamics of pair programming," in *29th International Conference on Software Engineering (ICSE'07)*, 2007, pp. 354–363.

[11] I. Sommerville, *Software Engineering*, 9th ed. Harlow, England: Addison-Wesley, 2010.

[12] A. Cockburn and L. Williams, "Extreme programming examined," G. Succi and M. Marchesi, Eds. Addison-Wesley Longman Publishing Co., Inc., 2001, ch. The Costs and Benefits of Pair Programming, pp. 223–243.

[13] L. Williams, E. Wiebe, K. Yang, M. Ferzli, and C. Miller, "In support of pair programming in the introductory computer science course," *Computer Science Education*, vol. 12, no. 3, pp. 197–212, 2002.

[14] P. Baheti, E. F. Gehringer, and P. D. Stotts, "Exploring the efficacy of distributed pair programming," in *Proceedings of the Second XP Universe and First Agile Universe Conference on Extreme Programming and Agile Methods - XP/Agile Universe 2002.* Springer-Verlag, 2002, pp. 208–220.

[15] R. Duque and C. Bravo, "Analyzing work productivity and program quality in collaborative programming," in *2008 The Third International Conference on Software Engineering Advances*, 2008, pp. 270–276.

[16] B. Hanks, "Empirical evaluation of distributed pair programming," *Int. J. Hum.-Comput. Stud.*, vol. 66, no. 7, pp. 530–544, Jul 2008.

[17] H. Gallis, E. Arisholm, and T. Dyba, "A transition from partner programming to pair programming- an industrial case study," in *Proceedings of Pair programming workshop in 17th annual ACM conference on Object-Oriented Programming, Systems, Languages, and Applications*, 2002, pp. 1–8.

[18] K. Han, E. Lee, and Y. Lee, "The impact of a peer-learning agent based on pair programming in a programming course," *IEEE Transactions on Education*, vol. 53, no. 2, pp. 318–327, May 2010.

[19] D. L. Jones and S. D. Fleming, "What use is a backseat driver? a qualitative investigation of pair programming," Sep 2013, pp. 103–110.

[20] Z. Li and E. Kraemer, "Social effects of pair programming mitigate impact of bounded rationality," in *Proceedings of the 45th ACM Technical Symposium on Computer Science Education.* ACM, 2014, pp. 385–390.

[21] N. Nagappan, L. Williams, M. Ferzli, E. Wiebe, K. Yang, C. Miller, and S. Balik, "Improving the cs1 experience with pair programming," in *Proceedings of the 34th SIGCSE Technical Symposium on Computer Science Education.* ACM, 2003, pp. 359–362.

[22] C. O'Donnell, J. Buckley, A. Mahdi, J. Nelson, and M. English, "Evaluating pair-programming for non-computer science major students," in *Proceedings of the 46th ACM Technical Symposium on Computer Science Education.* ACM, 2015, pp. 569–574.

[23] L. Williams, D. S. McCrickard, L. Layman, and K. Hussein, "Eleven guidelines for implementing pair programming in the classroom," in *Agile 2008 Conference*, Aug 2008, pp. 445–452.

[24] L. Williams and R. Kessler, *Pair Programming Illuminated.* Addison-Wesley Longman Publishing Co., Inc., 2002.

[25] D. Dzvonyar, L. Alperowitz, D. Henze, and B. Bruegge, "Team composition in software engineering project courses," in *Proceedings of the 2nd International Workshop on Software Engineering Education for Millennials*, ser. SEEM '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 16–23. [Online]. Available: https://doi.org/10.1145/3194779.3194782

[26] V. Arnaoudova, S. Haiduc, A. Marcus, and G. Antoniol, "The use of text retrieval and natural language processing in software engineering," in *ICSE*, 2015, p. 949–950.

[27] J. Butler, "Revisiting bodies and pleasures," *Theory Culture & Society - THEOR CULT SOC*, vol. 16, pp. 11–20, 04 1999.

[28] C. West and D. H. Zimmerman, "Doing gender," *Gender and Society*, vol. 1, no. 2, pp. 125–151, 1987.

[29] "Diversity in computer science." [Online]. Available: https://code.org/diversity

[30] R. Strachan, A. Peixoto, I. Emembolu, and M. T. Restivo, "Women in engineering: Addressing the gender gap, exploring trust and our unconscious bias," in *IEEE Global Engineering Education Conference*, 2018, pp. 2088–2093.

[31] A. Shekhar and N. Marsden, "Cognitive walkthrough of a learning management system with gendered personas," in *Proceedings of the 4th Conference on Gender & IT.* ACM, 2018, pp. 191–198.

[32] C. Mendez, H. S. Padala, Z. Steine-Hanson, C. Hilderbrand, A. Horvath, C. Hill, L. Simpson, N. Patil, A. Sarma, and M. Burnett, "Open source barriers to entry, revisited: A sociotechnical perspective," in *Proceedings of the 40th International Conference on Software Engineering.* ACM, 2018, pp. 1004–1015.

[33] G. Charness and U. Gneezy, "Strong evidence for gender differences in risk taking," *Journal of Economic Behavior & Organization*, vol. 83, no. 1, pp. 50–58, 2012.

[34] M. Burnett, A. Peters, C. Hill, and N. Elarief, "Finding gender-inclusiveness software issues with gendermag: A field investigation," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, 2016, pp. 2586–2598.

[35] E. Arisholm, H. Gallis, T. Dybå, and D. Sjøberg, "Evaluating pair programming with respect to system complexity and programmer expertise," *IEEE TSE*, vol. 33, pp. 65–86, 03 2007.

[36] K. Falkner, N. J. Falkner, and R. Vivian, "Collaborative learning and anxiety: A phenomenographic study of collaborative learning activities," in *Technical Symposium on Computer Science Education*, 2013, pp. 227–232.

[37] W.-j. He and W.-c. Wong, "Gender differences in the distribution of creativity scores: Domain-specific patterns in divergent thinking and creative problem solving," *Frontiers in Psychology*, vol. 12, p. 571, 2021. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2021.626911

[38] W.-L. Lin, K.-Y. Hsu, H.-C. Chen, and J.-W. Wang, "The relations of gender and personality traits on different creativities: A dual-process theory account," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 6, pp. 112–123, 05 2012.

[39] A. Abraham, "Gender and creativity: an overview of psychological and neuroscientific literature," *Brain Imaging and Behavior*, vol. 10, no. 2, pp. 609–618, Jun 2016. [Online]. Available: https://doi.org/10.1007/s11682-015-9410-8

[40] B. Curtis, H. Krasner, and N. Iscoe, "A field study of the software design process for large systems," *Commun. ACM*, vol. 31, no. 11, pp. 1268–1287, Nov 1988.

[41] R. E. Grinter, "Recomposition: Putting it all back together again," in *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work.* ACM, 1998, pp. 393–402.

[42] R. E. Kraut and L. A. Streeter, “Coordination in software development,” *Commun. ACM*, vol. 38, no. 3, pp. 69–81, Mar 1995.

[43] D. Perry, N. Staudenmayer, and L. Votta, “People, organizations, and process improvement,” *IEEE Software*, vol. 11, no. 4, pp. 36–45, 1994.

[44] F. L. Denmark, “Women, leadership, and empowerment,” *Psychology of Women Quarterly*, vol. 17, no. 3, pp. 343–356, 1993.

[45] A. Eagly and S. Karau, “Gender and the emergence of leaders: A meta-analysis,” *Journal of Personality and Social Psychology*, vol. 60, no. 5, pp. 685–710, Jan 1991.

[46] ——, “Role congruity theory of prejudice toward female leaders,” *Psychological review*, vol. 109, pp. 573–98, Aug 2002.

[47] A. Eagly and B. Johnson, “Gender and leadership style: A meta-analysis,” *CHIP Documents*, vol. 108, Sep 1990.

[48] A. Eagly and S. Sczesny, *Stereotypes about women, men, and leaders: Have times changed?* American Psychological Association, 2009.

[49] C. L. Hoyt and J. L. Burnette, “Gender bias in leader evaluations: Merging implicit theories and role congruity perspectives,” *Personality and Social Psychology Bulletin*, vol. 39, no. 10, pp. 1306–1319, 2013.

[50] H. Ibarra and J. L. Petriglieri, “Impossible selves : Image strategies and identity threat in professional women’s career transitions by,” 2007.

[51] C. Kubu, “Who does she think she is? women, leadership and the ëbí(ias) word,” *The Clinical Neuropsychologist*, vol. 32, pp. 1–17, Dec 2017.

[52] K. Lanaj and J. R. Hollenbeck, “Leadership over-emergence in self-managing teams: The role of gender and countervailing biases,” *Academy of Management Journal*, vol. 58, no. 5, pp. 1476–1494, 2015.

[53] L. H. Liang, D. Brown, D. Lance Ferris, S. Hanig, H. Lian, and L. Keeping, “The dimensions and mechanisms of mindfulness in regulating aggressive behaviors,” *Journal of Applied Psychology*, vol. 103, Nov 2017.

[54] D. E. Perry, N. A. Staudenmayer, and L. G. Votta, “People, organizations, and process improvement,” *IEEE Software*, vol. 11, no. 4, pp. 36–45, Jul 1994.

[55] B. J. Tepper, “Consequences of abusive supervision,” *Academy of Management Journal*, vol. 43, no. 2, pp. 178–190, 2000.

[56] T. Yang and H. E. Aldrich, “Whoís the boss? explaining gender inequality in entrepreneurial teams,” *American Sociological Review*, vol. 79, no. 2, pp. 303–327, 2014.

[57] D. Damian, “Global software development: growing opportunities, ongoing challenges,” *Software Process: Improvement and Practice*, vol. 8, no. 4, pp. 179–182, 2003.

[58] J. D. Herbsleb and D. Moitra, “Global software development,” 2001.

[59] J. D. Herbsleb, D. J. Paulish, and M. Bass, “Global software development at siemens: experience from nine projects,” in *Proceedings. 27th International Conference on Software Engineering, 2005. ICSE 2005.*, 2005, pp. 524–533.

[60] S. D. Teasley, L. A. Covi, M. S. Krishnan, and J. S. Olson, “Rapid software development through team collocation,” *IEEE Trans. Softw. Eng.*, vol. 28, no. 7, pp. 671–683, Jul 2002.

[61] K. J. Anderson and C. Leaper, “Meta-analyses of gender effects on conversational interruption: Who, what, when, where, and how,” 1998.

[62] D. James and S. Clarke, “Women, men, and interruptions: A critical review,” in *D. Tannen (Ed.), Oxford studies in sociolinguistics. Gender and conversational interaction*, 1998, pp. 231–280.

[63] C. W. [160] Don H Zimmerman, “West, sex roles, interruptions and silences in conversation,” in *Communications of the ACM*, 1988, pp. 1268–1287.

[64] J. D. Herbsleb and A. Mockus, “An empirical study of speed and communication in globally distributed software development,” *IEEE Transactions on Software Engineering*, vol. 29, no. 6, pp. 481–494, Jun 2003.

[65] M. L. Knapp and J. A. Hall, “Non-verbal communication in human interaction,” 2010.

[66] J. A. Briton, Nancy J.and Hall, “Beliefs about female and male nonverbal communication,” *Sex Roles*, vol. 32, no. 1, pp. 79–90, Jan 1995.

[67] E. Leigh, “Men & women communicating in the workplace,” in *The Center for Healthcare Communications*, 2010.

[68] D. Tannen, “You just don’t understand: women and men in conversation,” 1990.

[69] M. Ubando, “Gender differences in intimacy, emotional expressivity, and relationship satisfaction,” *Pepperdine Journal of Communication Research*, vol. 4, no. 13, 2016.

[70] S. L. Jarvenpaa, K. Knoll, and D. E. Leidner, “Is anybody out there?: Antecedents of trust in global virtual teams,” *J. Manage. Inf. Syst.*, vol. 14, no. 4, pp. 29–64, Mar 1998.

[71] S. L. Jarvenpaa and D. E. Leidner, “Communication and trust in global virtual teams,” *Organization Science*, vol. 10, no. 6, pp. 791–815, Jun 1999.

[72] P. Kanawattanachai and Y. Yoo, “Dynamic nature of trust in virtual teams,” *Journal of Strategic Information Systems*, vol. 11, pp. 187–213, 2002.

[73] S. L. Jarvenpaa, T. R. Shaw, and D. S. Staples, “Toward contextualized theories of trust: The role of trust in global virtual teams,” *Information Systems Research*, vol. 15, no. 3, pp. 215—-310, Sep 2004.

[74] N. B. Moe and D. Šmite, “Understanding a lack of trust in global software teams: A multiple-case study,” *Softw. Process*, vol. 13, no. 3, pp. 217–231, May 2008.

[75] J. D. Herbsleb, A. Mockus, T. A. Finholt, and R. E. Grinter, “An empirical study of global software development: distance and speed,” in *Proceedings of the 23rd International Conference on Software Engineering. ICSE 2001*, 2001, pp. 81–90.

[76] P. Dourish and V. Bellotti, “Awareness and coordination in shared workspaces,” in *Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work.* ACM, 1992, pp. 107–114.

[77] D. Nutter and C. Boldyreff, “Historical awareness support and its evaluation in collaborative software engineering,” 07 2003, pp. 171–176.

[78] H. H. Olsson, E. Ó Conchúir, P. J. Agerfalk, and B. Fitzgerald, “Global software development challenges: A case study on temporal, geographical and socio-cultural distance,” Oct 2006, pp. 3–11.

[79] I. Steinmacher, C. Steinmacher, A. Paula, and M. Aurelio, “Awareness support in global software development: A systematic review based on the 3c collaboration model,” *Collaboration and Technology*, vol. 6257.

[80] B. Zhong, Q. Wang, and J. Chen, “The impact of social factors on pair programming in a primary school,” *Comput. Hum. Behav.*, vol. 64, pp. 423–431, Nov 2016.

[81] K. S. Choi, “A comparative analysis of different gender pair combinations in pair programming,” *Behaviour & Information Technology*, vol. 34, no. 8, pp. 825–837, 2015.

[82] G. Abramo, C. D’Angelo, and G. Murgia, “Gender differences in research collaboration,” *Journal of Informetrics*, vol. 7, p. 811–822, 10 2013.

[83] C. Miller, P. Rodeghero, M.-A. Storey, D. Ford, and T. Zimmermann, “"how was your weekend?" software development teams working from home during covid-19,” 2021.

[84] P. Ralph, S. Baltes, G. Adisaputri, R. Torkar, V. Kovalenko, M. Kalinowski, N. Novielli, S. Yoo, X. Devroey, X. Tan, and et al., “Pandemic programming,” *Empirical Software Engineering*, vol. 25, no. 6, p. 4927–4961, Sep 2020. [Online]. Available: http://dx.doi.org/10.1007/s10664-020-09875-y

[85] S. Schiller, F. Nah, B. Mennecke, and K. Siau, “Gender differences in virtual collaboration on a creative design task.” *International Conference on Information Systems 2011, ICIS 2011*, vol. 3, 01 2011.

[86] X.-L. Shen, M. Lee, C. Cheung, and H. Chen, “Gender differences in intentional social action: We-intention to engage in social network-facilitated team collaboration,” *Journal of Information Technology*, vol. 25, pp. 152–169, 06 2010.

[87] L. M. Rea and R. A. Parke, “Designing and Conducting Survey Research: A Comprehensive Guide,” 2015.

[88] “Remote pair collaboration survey questions.” [Online]. Available: https://drive.google.com/file/d/1sCeex80QbeVcpOwDLBft6LzsolsYxmH8/view?usp=sharing

[89] “Remote pair collaboration interviews.” [Online]. Available: https://drive.google.com/file/d/1XyNHR7OykMaEaupdfCt1ptsQd5ZpTnos/view?usp=sharing

[90] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qualitative research in psychology*, vol. 3, pp. 77–101, Jan 2006.

[91] I. Cuadrado, M. Navas, F. Molero, E. Ferrer, and J. F. Morales, “Gender differences in leadership styles as a function of leader and subordinates’ sex and type of organization,” 2012.

[92] K. K. P. R. Sandeep Kaur Kuttal, Bali Ong, “Trade-offs for substituting a human with an agent in a pair programming

context: The good, the bad, and the ugly," 2021. [Online]. Available: https://drive.google.com/drive/folders/1kIkt_d-q5F_DMsshhctv_Mbrm2YhtDdW?usp=sharing

[93] L. Beckwith, M. Burnett, S. Wiedenbeck, C. Cook, S. Sorte, and M. Hastings, "Effectiveness of end-user debugging software features: Are there gender issues?" in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2005, pp. 869–878.

[94] L. Beckwith, C. Kissinger, M. Burnett, S. Wiedenbeck, J. Lawrance, A. Blackwell, and C. Cook, "Tinkering and gender in end-user programmers' debugging," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2006, pp. 231–240.

[95] M. Burnett, S. D. Fleming, S. Iqbal, G. Venolia, V. Rajaram, U. Farooq, V. Grigoreanu, and M. Czerwinski, "Gender differences and programming environments: Across programming populations," in *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*. ACM, 2010, pp. 28:1–28:10.

[96] M. M. Burnett, L. Beckwith, S. Wiedenbeck, S. D. Fleming, J. Cao, T. H. Park, V. Grigoreanu, and K. Rector, "Gender pluralism in problem-solving software," *Interact. Comput.*, vol. 23, no. 5, pp. 450–460, Sep 2011.

[97] A. Durndell and Z. Haag, "Computer self efficacy, computer anxiety, attitudes towards the internet and reported experience with the internet, by gender, in an east european sample," *Computers in Human Behavior*, vol. 18, pp. 521–535, Sep 2002.

[98] K. Hartzel, "How self-efficacy and gender issues affect software adoption and use," *Commun. ACM*, vol. 46, no. 9, pp. 167–171, Sep 2003.

[99] A. Fisher and J. Margolis, "Unlocking the clubhouse: Women in computing," *SIGCSE Bull.*, vol. 35, no. 1, pp. .23–, Jan 2003.

[100] P. NSF ITWF & ITR/EWF Principal Investigator Conference, Philadelphia, "The influence of professional identification on the retention of women and racial minorities in the it workforce," 2004, pp. 65–69.

[101] "Piazza, blog. stem confidence gap," 2005. [Online]. Available: http://blog.piazza.com/stem-confidence-gap/

[102] A. Singh, V. Bhadauria, A. Jain, and A. Gurung, "Role of gender, self-efficacy, anxiety and testing formats in learning spreadsheets," *Comput. Hum. Behav.*, vol. 29, no. 3, pp. 739–746, May 2013.

[103] S. Kendall and D. Tannen, "Gender and language in the workplace, gender and discourse, ed. by ruth wodak," *Gender and Discourse*, pp. 81–105, 1997.

[104] T. Hentschel, M. E. Heilman, and C. V. Peus, "The multiple dimensions of gender stereotypes: A current look at men's and women's characterizations of others and themselves," *Frontiers in Psychology*, vol. 10, p. 11, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2019.00011

[105] E. National Academies of Sciences and Medicine, *Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine*, P. A. Johnson, S. E. Widnall, and F. F. Benya, Eds. Washington, DC: The National Academies Press, 2018. [Online]. Available: https://www.nap.edu/catalog/24994/sexual-harassment-of-women-climate-culture-and-consequences-in-academic

[106] M. P. Sadker and D. M. Sadker, *Failing at fairness : how America's schools cheat girls / Myra and David Sadker*. C. Scribner's Sons ; Maxwell Macmillan Canada ; Maxwell Macmillan International New York : Toronto : New York, 1994.

[107] C. Marshall and J. Reinhartz, "Gender issues in the classroom," *The Clearing House*, vol. 70, pp. 333–337, 1997.

[108] K. Weiler, "Feminist analyses of gender and schooling," *The critical pedagogy reader*, pp. 269–295, 2003.

[109] "National Center for Women & Information Technology." [Online]. Available: https://www.ncwit.org/

[110] M. Burnett, S. Stumpf, J. Macbeth, S. Makri, L. Beckwith, I. Kwan, A. Peters, and W. Jernigan, "Gendermag: A method for evaluating software's gender inclusiveness," *Interacting with Computers*, vol. 28, no. 6, pp. 760–787, Oct 2016.

[111] "Gendermag." [Online]. Available: http://gendermag.org/

[112] G. DeSanctis and R. B. Gallupe, "A Foundation for the Study of Group Decision Support Systems," *Management Science*, vol. 33, no. 5, pp. 589–609, May 1987.

[113] M. S. Poole, M. Homes, and G. DeSanctis, "Conflict Management and Group Decision Support Systems," in *Proceedings of the 1988 ACM conference on Computer-supported cooperative work - CSCW '88*. ACM Press, 1988.

[114] M. S. Poole, M. Holmes, R. Watson, and G. DeSanctis, "Group Decision Support Systems and Group Communication: A Comparison of Decision Making in Computer-Supported and Nonsupported Groups," *Communication Research*, vol. 20, no. 2, pp. 176–213, 1993.

[115] R. T. Watson, G. DeSanctis, and M. S. Poole, "Using a GDSS to Facilitate Group Consensus: Some Intended and Unintended Consequences," *MIS Quarterly*, vol. 12, no. 3, pp. 463–478, 1988.

[116] V. Sambamurthy and W. W. Chin, "The Effects of Group Attitudes Toward Alternative GDSS Designs on the Decision-making Performance of Computer-Supported Groups*," *Decision Sciences*, vol. 25, no. 2, pp. 215–241, 1994.

[117] M. Limayem, P. Banerjee, and L. Ma, "Impact of GDSS: Opening the Black Box," *Decision Support Systems*, vol. 42, pp. 945–957, 11 2006.

[118] T. X. Bui, *Co-oP: A Group Decision Support System for Cooperative Multiple Criteria Group Decision Making*. Springer Berlin Heidelberg, 1987.

[119] M. Aiken, M. Vanjami, and J. Krosp, "Group Decision Support Systems," vol. 16, no. 3, pp. 38+, 2020/3/2/ 1995.

[120] D. Vogel and J. Nunamaker, "Group Decision Support System impact: Multi-Methodological Exploration," *Information Management*, vol. 18, no. 1, pp. 15 – 28, 1990.

[121] P. Gray, "Group decision support systems," *Decision Support Systems*, vol. 3, no. 3, pp. 233–242, Sep. 1987.

[122] P. R. et al., "Can machine learning facilitate remote pair programming? challenges, insights implications," in *VL/HCC*, 2020, pp. 1–11.

[123] J. Dominic, B. Tubre, C. Ritter, J. Houser, C. Smith, and P. Rodeghero, "Remote pair programming in virtual reality," in *2020 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, 2020, pp. 406–417.

[124] "Moot microsoft vr system." [Online]. Available: https://mootup.com/microsoft-teams-in-vr/

[125] "Interactive whiteboard with real-time collaboration." [Online]. Available: https://explaineverything.com/

[126] "team collaboration software & online whiteboard for teams | miro." [Online]. Available: https://miro.com/

[127] "Online whiteboard application for tutoring and team collaboration." [Online]. Available: https://liveboard.online/

[128] N. Mangano, T. D. LaToza, M. Petre, and A. van der Hoek, "Supporting informal design with interactive whiteboards," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 331–340. [Online]. Available: https://doi.org/10.1145/2556288.2557411

[129] ——, "How software designers interact with sketches at the whiteboard," *IEEE Transactions on Software Engineering*, vol. 41, no. 2, pp. 135–156, 2015.

[130] D. Wüest, N. Seyff, and M. Glinz, "Flexisketch: A mobile sketching tool for software modeling," in *Mobile Computing, Applications, and Services*, D. Uhler, K. Mehta, and J. L. Wong, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 225–244.

[131] M. Magin and S. Kopf, "A collaborative multi-touch uml design tool," 02 2013.

[132] R. Jolak, B. Vesin, M. Isaksson, and M. R. V. Chaudron, "Towards a new generation of software design environments: Supporting the use of informal and formal notations with octouml," in *Proceedings of the Second International Workshop on Human Factors in Modeling co-located with ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems (MODELS 2016), Saint Malo, France, October 4, 2016*, ser. CEUR Workshop Proceedings, H. Störrle, V. Amaral, M. Chaudron, and M. Goulão, Eds., vol. 1805. CEUR-WS.org, 2016, pp. 3–10. [Online]. Available: http://ceur-ws.org/Vol-1805/Jolak2016HuFaMo.pdf

[133] "Survey to measure the gender differences." [Online]. Available: https://docs.google.com/spreadsheets/d/1vGGJRxKO98Fk_aGGyeBmX_KIHOSmsjhePO9Tgo1tgZU/edit#gid=1017942176