Gender Differences in the Group Dynamics of Smaller CS1 Project Groups

Bjørn Hjorth Westh*, Nanna Inie†, Louise Barkhuus‡ and Claus Brabrand§
Center for Computing Education Research (CCER)
IT University of Copenhagen (ITU); Copenhagen, Denmark
Email: *b.a.westh@gmail.com, †nans@itu.dk, ‡barkhuus@itu.dk, §brabrand@itu.dk

Abstract—Abundant research has investigated gender differences from the perspective of the individual, but little research has explored gender differences from a group perspective. This paper presents a survey-based quantitative exploration of gender differences when \( N=138 \) students working in smaller (3 students) groups that are either men-majority (WMM) or men-only (MMM). The study was conducted during four weeks of CS1 project work, and the investigation focused on student struggle, task division, group satisfaction, and course outcomes. We observed four significant gender differences: (i) Many women students express a lack of confidence early in the project work, as do many men students that are part of mixed-gender (WMM) groups, but not men that are part of men-only (MMM) groups; (ii) In men-majority groups, women perform more ‘non-technical’ tasks than men (e.g., analysis, design, and report editing); and (iii) women students in men-majority groups are less satisfied with several aspects of the group work than men students in the same groups. These findings add to our understanding of group dynamics and collaboration (including what to avoid and where to pay extra attention) when assigning projects in smaller groups in computing education.

Index Terms—Diversity, Gender, Higher Education, Project-Based Learning, Groups, Group Dynamics, Struggle, CS1

I. INTRODUCTION

Computing fields suffer from a lack of gender diversity [1], [2], [3], [4], [5]. Much research has been dedicated to identifying ways to increase the recruitment and retention of women students in formal computer science education, including discovering factors that cause women students to struggle in computer science (CS). Some work has focused on identifying factors that are inherent to the individual, such as mathematical skills [6], [7], prior exposure to programming [8], and modes of thinking [9], [7]. A substantial body of work has also focused on the experiences, students have during their education, especially during the early semesters, for instance, sense of belonging, self-efficacy, and attitudes to computer science, e.g., [10], [11], [12], [13], [14], [15]. Despite the fact that they often perform as well or better in final exams, women students’ sense of belonging, self-efficacy, and attitudes towards computer science fall below averages, and it is clear that the problem of recruiting and retaining women students in computer science and related fields is complex and multifaceted [15], [16].

One dimension which is less understood is the interplay between gender and group work in computing education. Little is known about how knowledge is shared, developed, and distributed in group work specifically in CS1. While we know more about how groups collaborate in for example work contexts and research contexts, we know less about how teamwork is experienced by students – despite extensive recognition that computer science education is increasingly collaborative, reflecting how most students will be expected to work in teams in industry [17].

Gender-heterogeneous groups have generally been shown to perform better in most aspects [18], [19], [20]. On the other hand, some research has shown that all-women CS1 classes lead to greater social connections and comfort, greater feelings of support within their class, more confidence in their CS knowledge, and a more welcoming classroom environment compared to women in the traditional class [21]. Considering that women students constitute a minority of the total class body, is it more optimal to opt for as many heterogeneous groups as possible, even if the groups are comprised of more men than women? Given a proportion of 20% women students (which is common in computing education in many countries), there are simply too few women to compose many heterogeneous groups where women are not the minority.

In the study presented in this paper, we explored gender differences in how group work was distributed and perceived in small group work. Small groups usually consist of 3-4 people and are considered optimal for creativity, diversity, and participation [22]. The paper presents a study of 46 groups of three \( (N=138) \) first-year software development students in their Introductory Programming (CS1) course and explores gender differences along four different parameters: the students’ personal experience of struggle, the groups’ task division, the students’ satisfaction with group work, and the course outcome (exam data).

II. RELATED WORK

A. Self-Efficacy, Sense of Belonging, and Confidence

Within social cognitive theory, self-efficacy is a critical variable in that it affects learning, motivation, and self-development [23]. Successful performances raise self-efficacy, whereas failures may lower it. Quielle et al. 2017 [7] compared the profile of men and women students enrolled in introductory programming modules, and found that women students have significantly less programming self-efficacy than men students, women have higher test anxiety, and women students perform...
Weaker than men in early programming tests but as good if not better at the end of the year examinations.

Similarly, Krause-Levy et al. [13] found that women have a lower sense of belonging in early computer science courses across three consecutive quarters. A lower sense of belonging was correlated with negative course outcomes regarding pass rates and course performance. They also found that a sense of belonging was less tied to student performance as students got further into the CS curriculum [13].

Lishinski et al. [9] identified an interesting feedback loop between performance and self-efficacy in early programming courses: women students responded to performance feedback earlier, which led them to revise their self-efficacy. The reduced self-efficacy then caused them to perform worse in subsequent programming tests—although not in the final exam outcome. The authors suggested that women students may be more prone to internalize early failures, and thus disengage from programming.

Salguero et al. [15] identified four primary factors that identified how much an individual student struggles during early computer science courses. These factors were personal obligations, lacking sense of belonging, in-class confusion, and lack of confidence. The authors argue that studies that focus on only one factor of student struggle, such as self-efficacy or sense of belonging, overlook the bigger picture of how many different factors interweave and affect each other. Their thorough investigation of these factors in multiple early CS courses reveals that students from traditionally underrepresented groups (such as women and people of color) tend to report struggling slightly more than average. The authors also point out that an important future direction for research is understanding how these factors interact. In our research, we take a departure from the survey developed by Salguero et al. and investigate whether student struggle interacts with group dynamics in men-majority and all-men groups.

B. Group Work in Computing Education

Social factors, such as communication, have been shown to play a bigger role in developing confidence for women students than objective measures of ability [10]. Despite this, computing is usually almost exclusively focused on technical subjects, and very seldom on “human” topics, such as how to work as part of a team, despite the fact that team participation is a skill of crucial importance to employers and industry in general [24], [25]. Underrepresented groups in computing (women, people of color, first-generation college students) tend to hold strong communal goals [26], [27], [28], [11], but technical fields are believed to afford little opportunity to be communal compared to social or life sciences.

Some research has indicated minor to no differences in communication styles by women versus men in mixed-gender groups [29]. Bender et al. found that social sensitivity (the personal ability to perceive the mind and mood of others) was correlated with team effectiveness [30]. In fact, it was a primary predictor of the effectiveness of the team in completing short-term tasks. The study also revealed that computer scientists (and people in related fields) generally are less socially sensitive, highlighting the need for better understanding and supporting group dynamics and satisfaction in CS.

In the context of CS, women are a minority proportion of the total body of students, but little is known about whether this is replicated in group formation, and whether that is good or bad for group satisfaction and performance. Wong et al. [19] studied 3rd. year university students and showed that having a mix of genders improved “socio-emotional” communication, which in turn led to improved “social presence”, leading to “decision process satisfaction”, leading to better group performance. Houldsworth and Mathews [18] also found that mixtures of gender were beneficial. Mixed gender teams have been shown to be more satisfying for students working on long (8 months) projects than for students working on short duration (5 weeks) tasks [20], [19]. It was proposed that a longer-term project allows the team to become better acquainted and better able to capitalize on the communication skills of women [20].

Research investigating specifically how women students experience group work in CS has found seemingly mixed results. Some have indicated that women generally experience more stress in peer work [31], and some have found that women agree that collaborative work increases their confidence [32]. Most studies are, however, not clear about the group compositions, i.e., whether the groups are composed of mixed-gender students, and whether women are also a minority in the individual groups. As CS education is increasingly collaborative, it is clear that we need a better understanding of how students experience working in groups and whether group work correlates with the experience of struggling in computer science.

III. Methodology

The study is based on a quantitative exploratory research method intended to identify potentially relevant factors from which new hypotheses could be derived; which then, in turn, can be tested by future quantitative and qualitative studies. Exploratory studies are capable of dealing with unknowns and discovering new knowledge. In the words of John Wilder Tukey [33]: “Exploratory Data Analysis (EDA) is detective work – numerical detective work [...] it can never be the whole story, but nothing else can serve as the foundation stone - as the first step.” – John W. Tukey, 1977. For these reasons, we deliberately use open-ended research questions.

A. Objectives

In the context of computing education, our goal is to investigate gender differences in the group dynamics of small groups for which there are more men than women (aka, men-majority groups). The study is based on a four-week project at the end of a CS1 (Introductory Programming) course. Specifically, we focus our study on answering the following four research questions that are all intended to explore potential gender differences:
**RQ1 (Personal Struggle):** What are potential gender differences in men-majority student project groups in terms of personal struggle?

**RQ2 (Task Division):** What are potential gender differences in men-majority student project groups in terms of division of labor?

**RQ3 (Group Satisfaction):** What are potential gender differences in men-majority student project groups in terms of the satisfaction with the group itself and the perception of its cooperation and communication?

**RQ4 (Course Outcome):** What are potential gender differences in men-majority student project groups in terms of the outcome of the course?

### B. Educational Context

The Danish educational context is situated in a democratic society with the vast majority of primary and secondary education being public and tertiary education being public overall and as such state-funded.

**CS1 Course.** The study takes place in the context of the CS1 (Introductory Programming) course which is a 15 ECTS\(^1\) course on the first-semester of a Bachelor in Software Development at IT University of Copenhagen. The CS1 (Introductory Programming) course is meant to teach hands-on introductory (Java) programming using the object-oriented paradigm based on the “Objects First” approach [34].

The first 11 weeks of the course as a total of 14 weeks are organized in terms of lectures (2x2 hours per week) taught by the teacher followed by exercise classes (2x2 hours per week) supervised by teaching assistants (TAs). Also, live coding (2 hours/week) demonstrates the non-linear programming process from scratch while deliberately (sometimes not) introducing bugs that students are subsequently shown how to debug, soliciting suggestions from students.

**CS1 Exam.** In order to be eligible for the exam, the students must complete three one-hour individual programming tests, three mandatory hand-in assignments, and a series of smaller programming tasks during the course of the full semester. The CS1 course ends with a 30-minute oral exam based where students are asked to: (1) present a randomly selected topic (out of nine topics announced ahead of time); (2) reflect on their project, based on questions from a project examiner that arise from the project report; and (3) perform live programming tasks. The students receive one individual grade which reflects their performance of all three tasks while incorporating the assessed quality of their project and report.

**Other Courses.** In addition to the Introductory Programming course, all students had two other mandatory 7.5 ECTS courses entitled “Foundations of Computing – Discrete Mathematics” (which essentially teaches computing-relevant discrete mathematics) and “Project Work & Communication” (which teaches basic group dynamics along with classical rhetorical communication skills). The latter course also trains the students in conflict prevention and -resolution. The students do not have any other courses during this semester.

### C. Project

The last month (four weeks) of the CS1 course is dedicated to a student group project which is undertaken by students in groups of three, formed by the teachers. The group size has always been three in the CS1 course; the educational programme wants students to learn to work together in not-too-large groups. The students are given a project specification for which they have to perform object-oriented analysis, design, implementation, and testing. The groups have to explain their solution and argue for all choices made in a 15-20 page report. In the 2021 edition of the course, the project was a Netflix-inspired media-streaming service without the actual video. The solution had to support the browsing of media (films & series), thematic categories, seasons and episodes (for series), user profiles, watch lists, and media search facilities. Students were provided with the report of a so-called “good solution” of an earlier project (involving cinema booking).

### D. Groups

To achieve the maximal amount of gender heterogeneous groups possible, we compared men-majority\(^2\) groups versus men-only groups (see Figure 1 for details). The information about the gender of students was obtained from the University’s enrollment system which is based on information from the central person registry of Denmark; i.e., our definition of gender is the government-registered gender of the individual. In Denmark, the government-registered gender corresponds to biological gender (sex) and is binary. An official process exists for changing one’s registered gender in the national person registry status, so we have relied on the current national person registry status. The experimental cohort consists of \(N=138\) students who have been randomly assigned to 23 men-majority

---

\(^1\)One academic year is 60 ECTS (European Credit Transfer and Accumulation System).

\(^2\)After ample discussion and deliberation, the authors decided to call the group type men-majority groups. Initially, we used the term women-minority groups but abandoned this term to avoid unintended “othering.”
students are graded according to their fulfillment of explicit role of the censor (external examiner) is to make sure that gave the group project grade based on the group report. The used were the official grades which were given by a committee data from the exam (RQ4). The individual final exam grades contributed to individual tasks (such as project management, group member felt they, versus someone else in the group, how the reports would be divided from the perspective of the group, as well as general satisfaction with the group work and project. For the task part of the survey (RQ3), we asked the TAs to define a number of constituent tasks that the project is made up of. They identified and agreed on eight tasks: Project Management, Analysis & Design, Backend Development, Controller Development, GUI Programming, Systematic Testing, Writing Report, and Correcting Report. There is a risk that a group does not agree with this division of the project into smaller constituent non-overlapping tasks (e.g., they have created a solution program without a controller). Also, there may be elements of the project that are not captured by the eight tasks. However, the TAs of the course are former students of the course itself, so they have a good grasp of how the reports would be divided from the perspective of the students. The students were prompted to rate how much each group member felt they, versus someone else in the group, contributed to individual tasks (such as project management, GUI programming, testing, and writing the final report).

Exam data. In addition to the surveys, we also retrieved data from the exam (RQ4). The individual final exam grades used were the official grades which were given by a committee consisting of three examiners: the teacher, a project examiner, and a so-called censor (external examiner). The latter two also gave the group project grade based on the group report. The role of the censor (external examiner) is to make sure that students are graded according to their fulfillment of explicit learning goals, and nothing else. There are, of course, limitations in using summative assessment grades as a metric for individual and group performance, as everything is condensed into one number, but it was the most obvious metric, and probably one that has value to the students of the course.

In particular, individual student grades from the CS1 final (oral) exam, including whether or not the students attended the exam. Also, we asked the project examiner and censor (external examiner) to provide their grade assessment of the group project reports. (Note that students do not receive standalone grades from their project; they only receive one holistic grade for their entire exam which incorporates the quality assessment of the group project.)

F. Ethical Considerations

The experimental setup was granted ethics approval by the university in advance before the semester began. In addition, we consulted the national initiative on Diversity, Ethics, and Privacy of Digital Research Center Denmark before running the experiment. (Please note that teachers deliberately forming “diverse student groups” is common practice in Denmark (in particular, at our university) in order to train students in working with people who are different from themselves.)

Participation in the weekly and post-project surveys was voluntary and all participating students were explicitly asked to give informed consent for their data to be used for this study. After data collection and data linking (aggregating the different sources), all data has been anonymized such that data is identified only via a (consecutively enumerated) participant identification number (e.g., P123) as opposed to personal information. Further, we report only aggregated results for large cohorts of students from which it is not possible to derive any information about any particular student. Individual results are essentially diluted into 23 (WMM) groups or 23 (MMM) groups. Women students will appear only as “one out of 23.” No one involved in the grading (examiners, project examiners, and censors) has had access to any of the data prior to anonymization at the end of the course.

G. Execution

Survey data collection. The first weekly survey was conducted exactly one week after the project started; the second, after two weeks of the project; and the third, exactly three weeks into the project. Each survey was followed up with reminders two days after release. The first survey received 93/138 responses (19 W∈WMM, 30 M∈WMM, 44 M∈MMM); the second 104/138 responses (19 W∈WMM, 33 M∈WMM, 52 M∈MMM); and the third 95/138 responses.

3https://direc.dk/diversity-ethics-and-privacy-ws10/
4This project is partially funded by DIREC with the intention of uncovering knowledge important for the recruitment and retention of students in Computer Science. One recurring discussion is how to form student groups and whether to let students form groups autonomously or to have course instructors form groups (and on which basis). This study was part of an initiative to understand how these groups work in practice, in order to be able to make evidence-based recommendations and decisions about this in the future.
1. Potential Gender Differences in Personal Struggle (RQ1)

We identified two significant gender differences related to personal struggle [15]. The first involves the lack of confidence construct; the second involves workload pressure from other courses (from the personal obligations construct).

Figure 2 plots the average lack of confidence struggle scores for individual group members as a function of the number of project weeks (Weeks 1–3). Interestingly, during the first week of the project (Week 1), we see that women exhibit high lack of confidence struggle scores with an average of 3.4. Curiously, also the men, but only the ones in (mixed-gender) WMM groups with women, exhibit similar signs of confidence struggle with an almost equal average struggle score of 3.3. The men in men-only (MMM) groups, however, show virtually no signs of confidence struggle (average score of 1.6). We also see that after the initial week (Weeks 2 & 3), all struggle scores—for women and men alike—drop to a negligible struggle score level (on par with that of the men-only MMM group throughout the entire project).

In summary, both women and men in mixed-gender (WMM) groups appear to struggle significantly more than men in men-only (MMM) groups. The difference is statistically significant ($p=0.0056^{**}$). We summarize all of this as follows:

**Observation 1A:** Many women students struggle with a lack of confidence early in the project work; as do many men students that are part of men-majority groups, but not men that are part of men-only groups.

Figure 3 shows the distribution of personal obligation struggle scores involving other course requirements (including workload pressure from other courses) for W ∈ WMM vs M ∈ WMM and M ∈ MMM. We see that women have a high average struggle score of 3.8 whereas men have much lower scores (2.0 for WMM & 2.5 for MMM). The difference between the women and men, in general (M ∈ WMM U MMM), is significant ($p<0.0001^{***}$). We capture this by:

**Observation 1B:** During student group projects, women appear to be more susceptible than men, in general, to personal obligation struggle induced by other course requirements.

The remaining weekly survey questions on struggle involving illness, family obligations, work obligations, social/personal life issues, embarrassment, and lack of interest did not produce any significant gender-difference signals. The same was true for the remaining constructs, lacking sense of belonging and in-class confusion.
B. Potential Gender Differences in Task Division (RQ2)

Figure 4 visualizes the average response scores for task division according to each of the eight qualitatively different tasks that had been selected by the TAs. The results visualize the level of self-reported involvement in the task on a scale from 1 (I did nothing) to 5 (I did everything). Above the histogram bars, we give the gender ratio calculated as the average score of the women divided by that of the men. Hence, a gender ratio above 1x indicates that women were more involved in the task than men; in contrast, a ratio below 1x that men were more involved than women; whereas a ratio of exactly 1.0x indicates a gender equal task division. Below the histogram bars, we give information (using star notation) on whether or not the gender differences are significant.

We see that the women were significantly more involved in the *Analysis & Design* task with gender ratio of 1.2x and a *p*-value of 0.0088**. The same goes for *Correcting Report* with a gender ratio of 1.3x and a *p*-value of 0.013*. *Project Management*, *GUI Programming*, and *Writing Report* appear to be roughly equally distributed among the genders with gender ratios close to 1x (grayouted in the figure). *Systematic Testing* has a gender ratio of 0.8x and appears to be performed more by men than women, but the result is, in fact, not statistically significant (*p*=.104). The same goes for the *Controller Development* with a gender ratio of 0.7x and a *p*-value of .105. *Backend Development*, however, appears to be a statistically significantly more men-dominated task with a gender ratio of 0.7x and *p*-value of .020*. (We also analyzed the data using a *proportional* statistical analysis that compares the genders in terms of the number of students who reported that they did *more* than average in proportion to the number of students reporting that they did *less* than an average share of the work. According to this test, *Systematic Testing* also exhibits a significant work-division gender difference with a *p*-value of .03*.) We summarize our findings on task division:

**Observation 2:** In men-majority (WMM) groups, women appear to be more involved in “non-technical tasks” than men; men, on the other hand, appear to be more involved in “technical tasks” than women.

With men more involved in “technical tasks” such as *Backend* (and *Controller*) *Development*, one would also expect them to be more inclined to struggle related to “getting stuck on a bug.” Indeed, men indicate statistically significantly more struggle around this question than women (*p*=0.00043***).

C. Potential Gender Differences in Group Satisfaction (RQ3)

Figure 5 gives an overview of significant gender differences in terms of dis-satisfaction with various aspects of the project group work. The histogram bars show the percentage of (women vs men) students that were dis-satisfied. Above the histogram bars, we give the gender ratio of women over men percentages; below the histogram bars, we indicate (using star notation) the extent to which there was statistical significance. The results are obtained via a proportional statistical analysis that compares the genders in terms of the ratio of students who reported dis-satisfaction.

The first column (A) visualizes dissatisfaction with “Group & Project, in general”. We see that a third of the women (5/15 = 33%) were dissatisfied with the group & project. In stark contrast, none of the men (0/34) reported dissatisfaction with the group & project. The difference is statistically significant (*p*=0.0038***).

The second column (B) details dissatisfaction with “communication in the group.”. Two-fifths (6/15 = 40%) of the women were dissatisfied with the communication; whereas this was only 12% (4 out of 34) for the men in the WMM groups. The differences are statistically significant with a *p*-value of .024*. The results for men in mixed-gender (WMM) groups are the same as those for men-only (MMM) groups.

The third column (C) shows the number of students that were dissatisfied with the *attitudes* of group members. Close to a quarter of the women (4/15 = 27%) were dissatisfied with the attitudes. In contrast, only a single man out of 34 (2.9%) was dissatisfied with the attitudes. The difference is statistically significant (*p*=0.030*) using the Fisher test.

The final column (D) plots the number of students dissatisfied with the constructiveness of *discussions* during our group work. Two-fifths (6/15 = 40%) of the women were dissatisfied with the discussions. In contrast, this was only the case for 8.8% (3 out of 34) of the men in mixed-gender groups.
(WMM). The difference is statistically significant ($p=0.0439^*$). Men in WMM-groups were slightly more dissatisfied with the constructiveness of discussions in groups than men in MMM-groups. The difference is significant at $p = .027^*$; here, only one man out of 32 (3.1%) was dissatisfied with the constructiveness of discussions. In summary:

**Observation 3:** In men-majority (WMM) groups, women appear to be more dissatisfied, than men, with the group & project (in general), group communication, member attitudes, and the constructiveness of discussions.

D. Potential Gender Differences in Course Outcome (RQ4)

There appear to be no statistically significant gender differences in either the individual exam grades ($p=.28$) or in the group project grades for mixed-gender (WMM) vs men-only (MMM) groups ($p=.47$).

**Observation 4:** Despite various gender differences in struggle during the group project, there appear to be no significant gender differences in terms of either individual or group project grades.

V. Discussion

Despite Denmark (and Scandinavia) ranking in the top of the European countries regarding gender equality, computing education, in Denmark, suffers from a staggering lack of gender diversity [36]. Our exploratory study identified potential gender differences, and in this section, we will explore potential hypotheses that may explain these findings. Even if founded in prior research, hypothesizing carries a risk of over-simplifying or stereotyping, so we emphasize that our hypotheses are taken as speculations rather than conclusions.

Given the quantitative exploratory analysis methodology, it is important to consider the possibility of random correlations. This is an inevitable side-effect of wide quantitative exploratory studies tracking multiple independent variables. The remedial antidote is to not oversell and disguise observations as truths or proofs of correlation, but as testable hypotheses that may explain these findings. Even if founded in prior research, hypothesizing carries a risk of over-simplifying or stereotyping, so we emphasize that our hypotheses are taken as speculations rather than conclusions.

Hypothetical 1: While prior research has shown that women struggle more with a lack of confidence and self-efficacy early in computing education [7], [9], it is surprising that this early lack of confidence extends to men in the same groups. We speculate that this could be due to several reasons, and likely a mix of those. First, the general group discourse and atmosphere, where heterogeneous groups may speak more openly about experiencing difficulties and challenges. Heterogeneous groups have been shown to exercise more socio-emotional communication [19], and to promote an atmosphere of trust in the group [37], and this could mean that men in men-majority groups may be more ‘ok’ admitting (natural) early lack of confidence. Conversely, it is also possible that all-men groups establish an atmosphere of ‘hardcore’ competition [38], [10] which is left unchallenged by other group members, and thus men students in all-men groups are less prone to talk about, and therefore register any early lack of confidence. Secondly, prior research has indicated that women students tend to score higher in tests of planning capabilities [39], and it is possible that an increased focus early on planning the scope of the entire project may induce higher initial anxiousness in the group. A final hypothesis to explain this observation may be that women students have higher ambitions or a higher level of perfectionism and that this may increase the expectations of the group performance [40], [41], leading the whole group to feel less confident about its performance.

**Hypothesis 1b:** Women students report greater struggle with personal obligations than men students, but only on the parameter of workload pressure from other courses. The other courses at this period of the semester were a course in discrete mathematics and a course in project work and communication, both of about half the workload of the introductory programming course (cf. section III-B), and the semester was designed so that the majority of the courses’ workload would not overlap. One explanation for this finding is that women may generally experience a greater sense of obligation to perform well across all courses, or a greater sense of perfectionism leading to a greater sense of pressure [15]. Another hypothesis is that the course in discrete mathematics was experienced as more demanding for women, since some research has indicated that women in CS1 tend to have low self-efficacy in mathematics [42], although some research has, however, also shown that women students generally have a higher mathematical level and grade when enrolled in CS1 than their men peers [7].

**Hypothesis 2:** Regarding gender differences with respect to the involvement in “technical” vs “non-technical” tasks, we speculate that such a task division is a consequence of a mixture of personal confidence and interest, rather than skill. Prior research has shown that women students tend to be more interested in what they can achieve with a computer, rather than the computer itself [43], [44], and that women and men differ in preferences, problem-solving style, and willingness to tinker/explore [45]. Another possibility is that women either volunteer for or are asked to do low-promotability tasks as previously observed in mixed-gender groups [46]. Low-promotability tasks in relation to industry and in academia are defined as tasks that, while benefiting the organization, are less likely to affect the individual’s performance evaluation. What constitutes low-promotability tasks to students in a CS1 course may differ, but the notion that the task of “correcting the report” is comparably unattractive should not be dismissed.

**Hypothesis 3:** In men-majority groups, the women reported a lower degree of satisfaction with several dimensions of the group work. Although some previous studies have found similar patterns of gender-related satisfaction with group work [47], it is not clear where this gender gap stems from, and it is an important avenue for future research. It is interesting that some research has previously found that women tend to prefer to compete in teams, rather than individually [48], but women
students tend to be less satisfied with group work than men students. One possible explanation is a misalignment between the goals of the group; underrepresented groups in computing education appear to value communal goals to a higher degree than the majority of the student body [26], [27], [28], [11]. It is also possible that women generally are less satisfied in men-majority groups (but not all-women or women-majority groups) [49], [50], [21].

Reflections. Although we found possible gender differences in some of the measured metrics, it is worth noting that most of the surveyed topics did not reveal statistically significant differences based on gender. While this does not prove an absence of frustration or struggle for students in men-majority groups in computing education, at least few were evident by this survey metric. Therefore, the main objective of this study is to provide indicative evidence for future research in this area. The secondary objective is to inspire possible interventions to increase diversity in computing education. Based on the observations in this study, we could not observe less struggle, more satisfaction, or better exam outcomes for women students when they were placed in heterogeneous groups as the minority, and slightly decreased confidence and satisfaction for men students in men-majority groups (compared to men students in all-men groups). The study, therefore, does not provide evidence in favor of purposely forming men-majority groups. The study does, however, provide some indications for factors to be aware of when men-majority groups are formed. In an echo of Winter et al.: “There are no easy solutions to these complex experiences of being a minority, but clearly much work is needed to consider how to respond and how to foster a more robust and resilient sense of belonging for women within CS” [16].

VI. Threats to Validity

False positives vs false negatives (type I vs type II errors)? As mentioned above, we perform multiple statistical analyses. Consistent with recommendations on statistics, we deliberately do not use a Bonferroni correction since this is an exploratory study where “it is better not to miss a possible effect” [51].

Bias from project work training? In parallel with the CS1 course, our students had a course on “Project Work & Communication” teaching them how to prevent and resolve group conflicts. This may affect the results, although research suggests that teamwork experience and training do not necessarily translate to a more satisfying team work [20].

Are surveyed students representative? As described in Section III-G, the total response rate was 68%. If those who responded are representative of the entire student population, the results will generalize. If, however, there is a correlation between those not responding to our survey versus those struggling (or not struggling), the results may not generalize.

Students interfering with the study? The students were not aware of the objectives of this study so as to avoid or, at least, minimize interference with the results. Also, it is unclear what the student would stand to gain from interfering with the experiment, had they become aware of the study.

Beyond predetermined groups? Importantly, in our study, the groups were formed (randomly) by the teacher. In contrast, groups may also be formed by the students themselves. For this reason, our results might not translate to teams in a professional setting (e.g., a company or an organization), nor to courses where students form groups autonomously.

Beyond groups of three? The study is focused on small groups specifically. In smaller (maximum 4) groups, only the WMMM constellation is directly comparable; we speculate that the findings extend to such groups (Fig. 1). We do not know if or to what extent the results extend to larger groups; especially ones with more than one woman (e.g., WWMMM)?

Beyond CS1? This work was carried out in the context of first-semester CS1 students. We do not know whether the results generalize to more academically mature students; presumably, they have learned to better cope with stress or there may even be a selection pressure against the ones most unable to cope (who may even have dropped out by that time).

VII. Conclusion

In this paper, we explored gender differences in group work in CS1, focusing on men-majority groups. We found statistically significant gender differences along four different parameters, summarized in five observations: (1A): Many women students in men-majority groups struggle with a lack of confidence early in the project work, and this extends to men students in men-majority groups as well. (1B): Women appear to struggle more with personal obligations from other course requirements than men. (2): In men-majority groups, women students are reportedly more involved in “non-technical” tasks, while men students are reportedly more involved in technical tasks. (3): In men-majority groups, women are generally less satisfied with the group work than men in the same groups. (4): We did not find statistically significant gender differences with respect to individual or group grades.

The study was explorative in nature, and we have therefore highlighted several possible hypotheses to explain the observations, which we believe provide important avenues for future research. In a computing educational setting where women are, for the most part, a minority, we find it critical to understand how this impacts their experience and the educational environment in general, as well as what instructors and teachers can do to mitigate negative consequences. We make a cautious recommendation based on our findings:

Recommendaion: Educators of early computing courses involving group projects should be wary of forming smaller men-majority (WMM) groups.

References


[25] D. A. Ertmer and R. J. Newby, “‘you mean you have to work together?!’ a study of the formation and interaction of programming teams in a college course setting,” in Proc. 6th international workshop on computing education research, 2010.