Evidence for Teaching Practices that Broaden Participation for Women in Computing

Briana B. Morrison*
bbmorrison@virginia.edu
University of Virginia
Charlottesville, Virginia, USA

Beth A. Quinn*
beth.quinn@colorado.edu
University of Colorado Boulder
Boulder, Colorado, USA

Steven Bradley
Durham University
Durham, England
s.p.bradley@durham.ac.uk

Kevin Buffardi
California State University - Chico
Chico, California, USA
kbuffardi@mail.csuchico.edu

Brian Harrington
University of Toronto Scarborough
Scarborough, Canada
brian.harrington@utoronto.ca

Helen H. Hu
Westminster College
Salt Lake City, Utah, USA
hhu@westminstercollege.edu

Maria Kallia
University of Glasgow
Glasgow, Scotland
maria.kallia@glasgow.ac.uk

Fiona McNeill
University of Edinburgh
Edinburgh, Scotland
f.j.mcneill@ed.ac.uk

Oluwakemi Ola
University of British Columbia
Vancouver, Canada
kemiola@cs.ubc.ca

Miranda C. Parker
University of California - Irvine
Irvine, California, USA
miranda.parker@uci.edu

Jennifer Rosato
College of St. Scholastica
Duluth, Minnesota, USA
jrosato@css.edu

Jane Waite
Raspberry Pi Foundation
Queen Mary University of London
London, England
j.l.waite@qmul.ac.uk

ABSTRACT

Computing has, for many years, been one of the least demographically diverse STEM fields, particularly in terms of women's participation [12, 36]. The last decade has seen a proliferation of research exploring new teaching techniques and their effect on the retention of students who have historically been excluded from computing. This research suggests interventions and practices that can affect the inclusiveness of the computer science classroom and potentially improve learning outcomes for all students. But research needs to be translated into practice, and practices need to be taken up in real classrooms. The current paper reports on the results of a focused systematic "state-of-the-art" review of recent empirical studies of teaching practices that have some explicit test of the impact on women in computing. Using the NCWIT Engagement Practices Framework as a means of organization, we summarize this research, outline the practices that have the most empirical support, and suggest where additional research is needed.

*Working Group co-leader

ITiCSE-WGR '21, June 26-July 1, 2021, Virtual Event, Germany

ACM ISBN 978-1-4503-9202-0/21/06.
https://doi.org/10.1145/3502870.3506568

KEYWORDS

Gender and Diversity; Learning Environment; Undergraduate Instruction; Curriculum Addressing Gender and Diversity; Teaching Practices; Inclusive Culture

ACM Reference Format:
Briana B. Morrison, Beth A. Quinn, Steven Bradley, Kevin Buffardi, Brian Harrington, Helen H. Hu, Maria Kallia, Fiona McNeill, Oluwakemi Ola, Miranda C. Parker, Jennifer Rosato, and Jane Waite. 2021. Evidence for Teaching Practices that Broaden Participation for Women in Computing. In *2021 ITiCSE Working Group Reports (ITiCSE-WGR '21), June 26-July 1, 2021, Virtual Event, Germany*. ACM, New York, NY, USA, 34 pages. https://doi.org/10.1145/3502870.3506568

1 INTRODUCTION

Computing is a wide-ranging discipline that influences almost every corner of our modern world and provides high paying and influential jobs. It remains, however, one of the least demographically diverse STEM fields [12, 36]. The exact patterns of exclusion vary across national contexts but generally those who are underrepresented are women of all backgrounds and Black and Hispanic men (at least in the workforce, see [12]). That is, in most English-speaking countries, computing tends to be the purview of White and Asian men.

Given the ubiquity and power of computing, many have called for efforts to broaden participation in computing (BPC) [5, 16, 32, 74, 96]. The emergence in the U.S. context of the National Center for Women & Information Technology (NCWIT), the Computing Alliance of Hispanic-serving Institutions (CAHSI), the Institute for African-American Mentoring in Computer Sciences (iAAMCS), AnitaB.org, and many others were a direct response. The U.S. National

Science Foundation's is now requiring "BPC Plans" (i.e., "Broadening Participation in Computing") to be submitted as part of the formal process of applying for research grants across a wide range of computing research proposals [22]. A similar movement toward systemic change can be found in the European Commission's Women in Digital initiative [42].

While systemic, high-level initiatives are crucial, fruitful work is also needed "on the ground." In the case of higher education, one of the most powerful sites of intervention is the classroom. Not only do pedagogy and curricula impact the recruitment and retention of women in computing, but the classroom is an important location where computing culture is created, i.e., where students begin to establish a professional identity, see what good work looks like, and internalize norms of how one treats colleagues. That is, students learn much in college computing classrooms beyond computing. Will they learn that computing is only for a narrow type of person? Or can they learn a more expansive vision of computing in society?

Computer science instructors know the content of the subject they teach. But where do they learn how to teach their students, especially using techniques and practices that are welcoming and inclusive to students who–because of cultural stereotypes–might not see themselves in computing? What can individual instructors do to recruit and retain talented students from all demographic groups? The answer may lie in a growing body of research at the intersections of computer science education and the learning and social sciences. The last decade, in particular, has seen a proliferation of research exploring the effects of new teaching techniques, support strategies, and course sequencing on the retention of students who have been excluded from computing. Recent research suggests there are interventions and practices that instructors can employ that may positively affect the inclusiveness of the computer science classroom [8, 10, 47]. In addition, many of these practices have the additional benefit of improving learning outcomes for all students. The interdisciplinary nature of this growing body of research is a strength, as it benefits from varied theoretical frameworks and diverse methodologies. But this strength presents a problem, especially for practitioners: How does a Computer Science (CS) instructor come to know this research as it spans disciplines distinctly outside their core expertise? And more practically, how do they effectively apply research findings in their courses?

This report is the culmination of work by a 2021 ITiCSE working group convened to "chronicle the evidence for broadening participation" in computing education. Because of the COVID-19 pandemic, ITiCSE working groups worked as remote teams and the conference itself was conducted remotely. Under this model, ITiCSE working groups were required to submit a draft report before the beginning of the conference (June 27, 2021). Given this extremely accelerated time frame, we limited our focus to the issue of women's participation in undergraduate computing. The working group began work in early May and completed initial work for a "state-of-the-art" method of systematic literature review in a little over six weeks. The group worked an additional eleven weeks to complete a first draft of the report that was submitted in early September.

We also limited our review to interventions conducted in English-speaking countries because of the varied ways that gender operates across different linguistic and cultural contexts. Detailed information on search criteria and methodology may be found in the Methods section. We understand that this produced a more limited understanding of broadening (or widening) participation, but we hope the methods modelled here can be used in the future to expand the scope.

To organize our work, we made use of NCWIT's 2013 Engagement Practices Framework [83]. The Framework provided much needed structure and some lightweight theoretical framing. We also subjected it to evaluation based on the research evidence uncovered by our work. In summary, we found some support for all aspects of the Framework but were surprised by the lack of well-designed interventions that could provide unequivocal support for individual practices. Many studies implemented multiple interventions at once and too often the question of the impact on broadening participation seemed an afterthought. Our analysis did uncover research supporting the addition of three new practices to the EP Framework. The Framework is discussed in depth in section 3 and the suggested additions in section 5.4.

1.1 Women are Underrepresented in Computing in the English-Speaking World

Similar patterns of marginalisation of women in computing are seen across the English-speaking world. Consider the U.S. context: in 2019 women earned 57% of all bachelors degrees but only 21% of computing bachelor's degrees [36]. In the U.S. workforce, women hold 57% of professional occupations but only 25% of professional computing occupations. Their representation in leadership positions in computing is even lower, e.g., women hold only 18% of Chief Information Officer (CIO) positions in the Top 1000 Companies [36]. Canadian figures mirror U.S. figures fairly closely, with 18% of CS students being women in 2015 (up from only 15.8% in 2010) [24], and women accounting for 21% of people performing ICT roles in ICT companies. Within the largest 100 Canadian ICT companies, only five women hold CEO positions with an additional woman serving as a co-CEO. Only 15% of all leadership roles in these companies are held by women, with 26% of these companies having no women in senior management roles [122].

In the UK, while women account for 57% of all higher education students, their proportion of computing degrees is similar to the North American context: in 2019 women represented just 19% of CS students and only 16% of CS graduates [52, 112]. In the same year in industry, 16% of IT professionals were women - a decrease both proportionally and in absolute numbers from 2016, when women represented 21% of IT technicians. In Australia, women were awarded 38% of all STEM degrees in 2019 but only 18% of computing degrees [6]. And while women accounted for almost 50% of the Australian workforce in 2020, their representation in CS-related industry was only 21% [7]. The situation in New Zealand is better. Women make up 43% of the workforce as a whole, 36% of IT and Computing students, and 23% of ICT employees [81]. Nearly 40% of technology Small and Medium Enterprises (SMEs) have leadership boards which are either gender balanced or predominantly female, which, whilst still some way from gender balanced, is better than we see in leadership roles in many other English-speaking countries.

The reasons for women's current underrepresentation in computing in the English-speaking world are many, and the exact causation is contested. One potential cause that is easily dismissed is the claim that there is an inherent difference in interest for women versus men [37]. Two main critiques of this perspective are that women have not always been so underrepresented in computing and the pattern of underrepresentation fails to hold across cultures. For example, women accounted for over a third of computer science graduates in the U.S. in the mid-80s [94] and in countries such as India, computing careers are seen as highly appropriate for women, with women accounting for 42% of CS and computer engineering undergraduates in 2011 [117]. Today, researchers generally point to a gradual masculinization of computing as a profession, beginning with the rise of the personal computer [9, 41]. That is, in most of the English-speaking world, computing has come to be seen as something that men do and that women do not. Many professions are gendered (and "raced"). Since this gendering changes across time and culture, researchers suggest that it is not inherent but socially constructed and socially reproduced [56, 76, 100]. The reasons for these changes are complex but the mechanisms by which the gendering is perpetuated–or changed–are well understood. Social actors and social institutions (such as the media, parents, teachers, and employers) take up and reinforce ideas about who is the "normal" occupant of a profession and then explicitly or implicitly encourage some and discourage others. Those who match the expected demographics of the normal occupant, not surprisingly, tend to have a greater sense of belonging, confidence, and interest in the field [51, 76, 100]. Perhaps most powerfully, the culture and processes of these fields (e.g., teaching practices) begin to match the needs and expectations of the dominant group.

2 WHY INSTRUCTORS MATTER

Several large scale interventions have effectively increased the participation rates of women or other underrepresented populations in computing in particular contexts. These tend to be multi-faceted, highly coordinated efforts. Perhaps the most well-known are those at Carnegie Mellon University [74] and Harvey Mudd College [63]. These changes required buy-in and commitment from numerous faculty and administration members. Some interventions, such as those described in Betsy DiSalvo's dissertation [35], require significant resources including faculty time and large sums of money. But how do these initiatives begin? What can be done with limited resources and time? And what can a small group or even an individual do?

Many individual instructors want to make a difference. They want to recruit more women into their classes and they recognize that these students deserve welcoming and inclusive classrooms. Instructors are searching for techniques and practices to implement within their classroom to ensure the most inclusive environment, and that recruited students are retained.

The impact individual instructors can have on students is great. Hattie [50], for example, notes that 30% of the variance in student achievement can be attributed to individual teachers. Good teachers can certainly motivate and inspire their students to persevere and achieve. Positive encouragement from a teacher can literally change the course of a students' life by helping them to consider a career in a field they did not feel possible. In contrast, computing teachers that, for example, decorate their classrooms with 'geeky' cultural references such as Star Wars and Star Trek can send subtle messages to students that only those who like or 'get' these references are welcome in the computing classroom.

This report concentrates on pedagogical practices that are primarily implemented by a single instructor in their own classroom. Our goal is to provide the individual instructor, as well as other researchers, a guide for actions and interventions that can positively impact women students' experiences in their local environment.

3 A PRELIMINARY STRUCTURE: NCWIT'S ENGAGEMENT PRACTICES FRAMEWORK

A challenge for any project seeking to understand the state of research on a particular topic is how to organize that research into a coherent structure. In established fields that are less interdisciplinary, dominant theoretical concepts are often used. Research on broadening participation in computing is, however, a relatively new, interdisciplinary field of study without a shared theoretical frame or even shared terminology. Another approach is to let the frame emerge from the research itself via content coding [39]. While this approach has many benefits, it is notably labor- and time-intensive. Due to the short time frame for this working group, we chose instead to begin with an existing framework developed by the National Center for Women & Information Technology (NCWIT): The Engagement Practices Framework (hereafter, "EP Framework") [83].

The EP Framework was developed in 2013 as part of the EngageCSEdu project. A collaboration between NCWIT and Google, EngageCSEdu aims to improve the recruitment and retention of women in computing by improving teaching at the introductory collegiate level [38, 80]. It is an open-source repository of high quality instructional materials (e.g., assignments) for introductory courses that instructors can access for free. Materials are included in the collection based on assessed quality and alignment with practices that research suggest help recruit and retain women in computing. In 2013, NCWIT and Google convened a team of learning scientists, social scientists, and computer science educators to craft an inclusion rubric for EngageCSEdu in order to seed the collection with an initial set of materials. The interdisciplinary team conducted a limited literature review and reviewed pedagogical and curricular practices recommended by NCWIT, which were based on existing research at the time of publication. The resulting set of practices were content coded into eleven general "engagement practices," such as "Give Effective Encouragement" or "Use Meaningful and Relevant Content." (See Table 2 for a full elaboration of the principles and practices contained in the NCWIT EP Framework.)

These practices are clearly good pedagogical techniques. But the core question for the EngageCSEdu project was to illuminate why these practices also appear to effectively engage and retain women students. Drawing on general literature about the needs and behavior of marginalized groups in STEM fields, the original team categorized the practices into three "engagement principles" that sought to succinctly describe how each practice functions to engage students who, because of stereotypes about computing, may not see themselves in the field and may be more likely to leave. For example, why would having assignments that make explicit interdisciplinary connections from computing to fields such as biology potentially engage and help retain women in computing? We do not need to assume inherent differences in the interests of women compared to men. Rather, making interdisciplinary connections in this context is simply a means of expanding the narrow gendering of computing by connecting it to fields that are less explicitly gendered. Similarly, in this context the practice of using "well-structured collaborative learning" is less about the learning gains that can be realized by practices such as Process Oriented Guided Inquiry Learning (POGIL) or Pair Programming. These practices, if employed with an eye to equity and mindful of group dynamics, can help grow a more inclusive student community, and students who feel that they are part of a learning community are more likely to be retained [33, 40]. Collaborative learning techniques also provide experiences that can help students develop communication and teamwork skills, and an appreciation of diversity.

The EP Framework is only a starting point. This working group was convened out of a recognition that the amount of research has grown significantly since 2013 and that a more systematic review of the literature is needed. Our analysis of the research begins by using the categories and definitions of the EP Framework. But an important goal of this working group is to assess the strength of support for the Framework's recommended practices and to identify promising practices that may be missing from the Framework.

4 METHOD

4.1 Research Questions

The goal of the working group was to conduct a systematic, but bounded, state-of-the-art review of recent studies into teaching and related practices that impact women's participation in computing. We proceeded with the following research questions:

(1) What empirical, intervention-focused research exists to support or refute the principles and practices outlined in the NCWIT Engagement Practices Framework?
(2) What types of outcomes (affective, cognitive, and population) are represented within that evidence?

4.2 Search Approach

A state-of-the-art literature review aims to provide a critical survey of recent literature (e.g., produced in the past decade) as well as a synthesis of the current thinking in the field. According to Grant and Booth [46], this type of review is appropriate for assessing the current state of knowledge, and identifying gaps and priorities for future research. We chose the state-of-the-art method given the time constraints of the working group format and the proliferation of research on the topic in recent years. We followed the steps for a Systematic Literature Review Checklist adapted from the PRISMA 2020 Checklist [87].

After collaboratively articulating the rationale and objectives for the literature review, we specified the criteria for inclusion and exclusion of research papers. Aware that the project could quickly explode in scope, the group sought ways to narrow the search scope while still meeting our objectives. The working group–with perspectives from various disciplines and national contexts–had lively discussions over what should and should not be considered. We decided on the following scope parameters. First, given our goal of assessing the effectiveness of particular instructor practices, we only considered empirical work (both quantitative or qualitative) with an identifiable intervention from peer-review publications. Second, we only considered work from the last decade. Third, while it is likely that some teaching practices are effective across different levels of education and fields, these are questions only answerable by empirical comparison. Time limitations preclude us from doing that comparison so we limited the scope of our review to computing education at the collegiate level. Fourth, we recognized that the ways in which gender and race/ethnicity play out varies by social context, including nationality (where, e.g., legal regimes and cultural histories vary) and this variation may impact how particular interventions work. For this reason, we limited our search to papers from English-speaking countries. Lastly, while we had originally intended to consider research looking at the impacts on all under-served computing populations, we quickly realized that the number of search terms grew almost exponentially with the addition of each sub-group. Because of this, we decided that the present report would focus on research that assesses the impacts of interventions on women's underrepresentation. Where data is presented intersectionally (which is, unfortunately, not common), it is described.

4.3 Databases Searched

Given the interdisciplinary nature of the research we were seeking, we sought to cast the widest possible net. After consulting two librarians and asking working group members to identify databases that they regularly use, we identified 15 databases that span the fields of computing, engineering, education, and the social sciences. The databases are identified in Table 1.

4.4 Search Criteria

We began with the original inclusion criteria and collaboratively generated search terms (including synonyms) for each. For example, for the gender criteria, we settled on the terms "female", "woman", "women", and "gender." In addition, where available, one of the authors researched the associated dictionary (a list of common terms and organizational classification) for each database or found existing articles that matched our inclusion criteria and used the classifications on those papers to inform the search terms used. Because of the differences in how interfaces and search parameters are structured across the various databases, the search parameters we used varied slightly across databases. To limit the search for only undergraduate education, we excluded the terms "high school", "elementary school," "primary grades," and "secondary school," and included "higher education" when needed for specific databases. The method used to limit the queries to computing education varied by database. In the ACM Digital Library, which is almost exclusively computing related, we were able to limit to "cs0", "cs 0", "cs1", "cs 1", and "introductory course". For the Scopus database, we simply used the term "computing education".

The search queries were conducted by two of the authors between May 15, 2021 and May 28, 2021. The exact queries for each database can be found in the appendix. Note that for some databases, multiple queries were conducted to account for initial limitations. For example, a second query for the ACM Digital Library was conducted when we discovered that both "CS0" and "CS 0" (with a space) are used to designate that US-based course type. Because we assumed that most of the relevant research would be catalogued by the ACM Digital Library, we began the search process there to determine the scale of results we would receive. As we queried other databases, duplicates were removed based on previous queries. The number of articles included from each database, as listed in Table 1, represent unique articles not found in previously search databases. Databases are listed in the order they were searched.

Table 1: Papers retrieved per database

Databases	Retrieved	Included
ACM DL	1114	58
ERIC (FirstSearch)	108	4
ERIC	189	2
Scopus	56	0
Web of Science	405	9
IEEE	98	1
EBSCO (Academic Search Premiere)	15	0
EBSCO (Education Source)	168	3
Compendex	22	0
Springer	821	1
Science Direct	898	0
ProQuest	1065	6
PsycInfo	251	3
Google Scholar	78	0
CSEd Research	14	0
ProQuest Sociology	13	0
Total	5315	87

4.5 Summary of Inclusion Criteria

The final inclusion criteria were iteratively developed by the team. The final criteria for inclusion are:

- Empirical works, either quantitative or qualitative, with an identifiable intervention; no meta-analyses or commentary
- Conducted in English-speaking countries
- Published between January 1, 2011 and May 28, 2021
- Original research published in peer-reviewed publications
- Emphasis on activities an individual instructor can implement within a classroom; however, high quality papers where the intervention could require decision-making beyond the individual instructor (e.g., curriculum changes) were accepted
- Studies that make explicit comparisons between women and men students or high quality studies of initiatives targeting only women
- All course levels are included even though the scope of the searches privileged introductory college computing courses.
- Research is evaluated even if no differences or negative results are reported since these also inform practice

4.6 Screening

Using the search criteria outlined above, 5315 papers were initially identified (see Figure 1). After removing duplicates and results that were obviously outside our scope based on the title, 4332 papers remained for further analysis. In the first stage of the review process, group members were assigned sets of papers to assess for inclusion using the paper title and abstract and by applying the inclusion and exclusion criteria outlined in section 4.5. Each paper was categorised as *Include, Exclude,* and *Maybe*. From the 4332 papers reviewed, only 313 papers survived this initial review as "Include" or "Maybe." These 313 papers were then reviewed by a second reviewer, using the same criteria and information. The initial inter-rater reliability was moderate with Kappa = .654 ($p < .05$). All disagreements between reviewers were resolved via discussion (sometimes as a group). After this stage, 133 papers remained.

This set of papers was divided among the authors for a fuller assessment. The following information was gleaned from each paper: 1) the context, including nation and course type, 2) the intervention, 3) the subjects of the intervention, 4) the method of data collection, and 5) the NCWIT Engagement Practice(s) the intervention appeared to use (if any). The individual conducting this review also wrote a summary of the results, with a focus on the impact on women or gender differences. They then entered this data into a Qualtrics form (a survey tool) along with its bibliographic data. This allowed for an annotated bibliography to be produced and for research to be categorized by Engagement Practice.

At this point a deep dive into each paper was conducted to assess its findings and applicability for the final report. Outcomes were assessed by type and direction of outcome. The types of outcomes were classified as affective (e.g., confidence, belonging), cognitive (e.g., learning gains), and population (e.g., improved recruitment or retention). The direction of each outcome (positive, neutral or no effect, negative) was determined for all students, and for women or men as a groups. An additional 46 papers were excluded at this stage when a closer read revealed they did not meet the inclusion criteria, leaving 87 papers to be fully assessed and summarized in the final report. (Reasons for exclusion at this stage varied by paper). Figure 1 provides a graphic depiction of the screening and inclusion process.

A full list of the included papers can be found in the appendix in Table 18. Table 2 shows the categorization of papers according to the NCWIT Engagement Practices Framework, the organization strategy used to structure our discussion of the results. Many papers included interventions that implicated several Engagement Practices so the percentages do not sum to 100%.

5 RESULTS: RESEARCH EVIDENCE

For each Engagement Practice, a synthesis of papers is provided, including tables summarizing the findings by type and direction of outcome (affective, cognitive, population). In the synthesis, a narrative summary highlighting trends is provided. [1]

5.1 Grow An Inclusive Community

The NCWIT Engagement Practices Framework suggests that students who feel they belong to a community they identify with may be more likely to persist through difficulties [83]. According to the

[1] In the results section, we use the demographic terminology that the original authors use. Thus, you will see multiple instances of "female" and "male" used in the context of the discussion of gender. While gender theory suggests it is not accurate to use these terms in the context of gender (they are terms related to biological sex, not gender; appropriate terms are "woman" or "woman-identified"), we felt it was appropriate to reproduce how authors collected and reported their data.

Table 2: Categorization of papers accepted for the systematic literature review

Principle / Practice	# Papers	% Papers
Grow an Inclusive Community	**63**	**72%**
Avoid Stereotypes	5	6%
Use Well-Structured Collaborative Learning	23	26%
Encourage Student Interaction	17	20%
Make It Matter	**46**	**53%**
Use Meaningful and Relevant Content	19	22%
Make Interdisciplinary Connections	11	13%
Address Misconceptions about the Field	9	10%
Incorporate Student Choice	7	8%
Build Student Confidence and Professional Identity	**25**	**29%**
Give Effective Encouragement	4	5%
Offers Student-Centered Assessment	5	6%
Provide Opportunities for Interaction with Faculty	14	16%
Mitigate Stereotype Threat	2	2%
Candidates for New Practices	18	21%

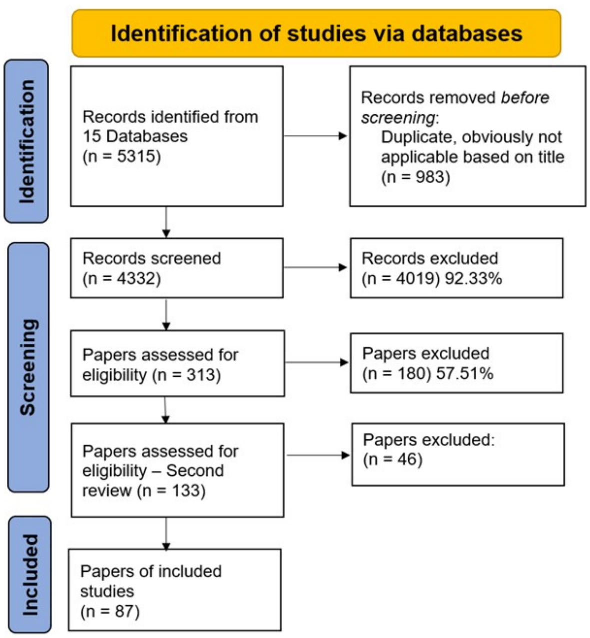

Figure 1: Identification and Screening Process

Table 3: Summary of papers identified within the practice, "Avoid Stereotypes"

Paper	Intervention	Outcomes*		
		A	C	P
[27]	Presentation of stereotyped models of computer scientists	F-		
[29]	Presentation of stereotyped models of computer scientists	F-		
[64]	Incorporation of physical activity into lecture breaks	F+	E+?	
[78]	Removal of gendered examples in textbooks	F+		
[79]	Modification of web resource aesthetics	F+		

* Outcomes are reported aligned to our research questions. A is Affective, C is Cognitive, P is Population. E is for everyone, M is for males, and F is for females. A + symbol means the results indicated the intervention improved outcomes. A - symbol means the intervention yielded negative outcomes. An = symbol means the intervention had no statistically significant effect. A ? symbol indicates the outcomes is either anecdotal or not statistically significant. Presence of multiple symbols (e.g., E+-) indicates mixed results. A blank entry indicates that aspect was not reported.

Framework, educators can encourage an inclusive community in computing programs by using these three practices:

- Avoiding Stereotypes
- Using Well Structured Collaborative Learning, and
- Encouraging Student Interaction.

5.1.1 Avoid Stereotypes. Five papers, as shown in Table 3, include aspects of the "avoid stereotypes" practice. Three different approaches emerged from our analysis. The first approach focused on addressing stereotypes that exist in course materials such as web interfaces and textbooks. For example, Metaxa-Kakavouli et al. [79] compared the impact aesthetic features of web pages had on students' interest in computing. Two versions of the website were compared. Each had identical content and page layouts but the photos and text style differed. The "masculine" version used a Star Trek photo and text evocative of old style computer screens (green on a black background) and the "gender neutral" version used a photo of plants and white font on a green background. The course webpage that reinforced stereotypes about computing (e.g., geek sci-fi culture, only about programming) negatively impacted female students' sense of belonging, interest in CS, and self-efficacy.

Medel and Pournaghshband [78] sought to understand how women were portrayed in course textbooks in cryptography and the impact these (often stereotypical) portrayals might have on women undergraduates in computing. The authors offer an analysis of a classical teaching example used in most textbooks to explain cryptographic protocols that portrays women negatively, e.g., 13% of the female characters had a positive connotation compared to 50% of the male characters [78]. The authors created a new version of the example, replacing human names with animal characters that maintain the "whimsical" aspect of the original example. Animal characters were referred to using the gender neutral term, "they." The authors conducted a small pilot study with two classes of undergraduates with one class using the traditional example and the other the new version. They found that women exposed to the gender-free example scored higher on confidence related to understanding that particular example compared to women exposed to the traditional example; no effect for men was noted. The authors do not, however, provide details on any statistical tests that were performed.

The second approach was to combat stereotypes through role models. While there is an assumption that the gender of role models positively impacts the sense of belonging, these papers examine how the role models' beliefs influence students' interest in computing [27, 29]. A study of 100 female non-computing undergraduate students explored the impact of interacting with role models. They observed that interacting with a role model with stereotypical interests (e.g., favorite magazine → *Electronic Gaming Monthly*) had a negative effect on women's interest in majoring in CS and sense of belonging [27]. Again studying non-computing students, Cheryan et al. observed that stereotypical computing role models negatively affect women's perceived success in the field [29]. They observed that *regardless of the gender of the role model*, stereotypical masculine ideas about computing negatively influence female students' interest in computing.

The third approach focused on incorporating in-class activities that went against the stereotype of computer scientists as unfit, sedentary, and having unhealthy lifestyles. A study of 750 students across five CS1 sections at the University of Toronto had students perform 'fit-breaks', short sessions of stretching and cardiovascular exercise, during mid-lecture intervals, and found that female students in the sections that used fit-breaks instead of generic music breaks showed improvements in perception of agency, usefulness, and enjoyment [64].

Overall, interventions that used gender-neutral course materials demonstrated improvement in women's interest in computing, sense of belonging, and confidence.

5.1.2 Use Well Structured Collaborative Learning. Educators who use well structured collaborative learning opportunities provide guidelines and structure during a formal learning activity for students to work together in class [83]. These guidelines should encourage students to work collaboratively and engage in intellectual discussions with each other. NCWIT recommends discouraging students from taking a "divide and conquer" approach since gender and racial stereotypes can influence group decisions about who does what (e.g., White & Asian men doing the technical tasks, women doing artistic or administrative tasks) [9]. While the goal of the collaboration is generally academic learning, it can also help grow a more inclusive student community by helping students develop communication and teamwork skills and an appreciation for diversity. However, a potential for microaggression and bias within teams exists if not carefully facilitated by instructors. Twelve papers that included this engagement principle involved pair programming and/or peer instruction and the remaining eleven papers studied other approaches such as peer mentoring, flipped classrooms and formal learning groups (see Table 4).

Table 4: Summary of papers identified within the practice, "Use Well Structured Collaborative Learning"

Paper	Intervention	Outcomes*		
		A	C	P
[3]	Pair programming + peer instruction	E=	E=	
[17]	Pair programming	E=	E=	E=
[18]	Linked course learning community			F+
[23]	Flipped class	F-		E-
[25]	Group exams		F+?	
[57]	Peer mentoring			F+
[59]	Having a female partner in pair programming	E+		E+
[65]	Formal learning groups	F+?		
[67]	Peer mentoring	E+?		F+
[69]	Flipped lecture + lab		E+	F+
[68]	Flipped, teams, gamification	F+ ? E+	E+	
[71]	Group projects, pair programming, homework partner option		E+	F+?
[72]	Pair programming		F+?	
[82]	Peer-led team learning	E+?	E+?	E+
[92]	Peer instruction, media computation and pair programming			F+? M+
[105]	Peer instruction, media computation and pair programming		E+ F++	E+ F++
[107]	Living learning community	F+		
[111]	Pair programming		F-	
[119]	Peer mentoring	E+?		
[124]	Peer mentoring + online course	F+		
[125]	Pair programming	F+ E+		
[129]	Peer instruction	E+		
[128]	Peer instruction	E+	E=	

* For explanation of symbols and abbreviations, see Table 3.

Pair programming is the most commonly studied collaborative learning practice. In pair programming, students switch between "driver" and "navigator" roles as they program together. Five papers examined the impacts of pair programming, three papers investigated the effects of adding both pair programming and peer instruction to an introductory computer science class, and one paper studied pair programming as one element in a shift to a more collaborative class. When Ying et al. [125] surveyed their students about pair programming, students reported improved learning experiences, with women more likely to report on the impacts on class atmosphere and networking, and men more likely to report on the impacts on productivity. Jarratt et al. [59] found that having a female partner increased lab attendance and students' confidence for both male and female students, with a greater effect for female students.

Of the papers which concentrated solely on pair programming, three papers compared exam scores of students who pair programmed with exam scores of students who programmed alone and found conflicting results. Maguire et al. [72] reported that female exam scores increased after adding pair programming to a CS1 course, and Smith et al. [111] observed that females who chose to work alone scored higher on their CS2 exams than partnered females. A third study by Bowman et al. [17] randomly assigned students to sections with and without pair programming and found no significant benefits or detriments of pair programming to women. These results suggest that the way in which pair programming is implemented may affect its outcomes. Instructors vary widely in how they facilitate pair programming in the classroom, with some regularly reminding students to switch between roles versus others who offer less guidance. They also differ in how pairings are done and managed. While it was not the primary focus of their study, Latulipe et al. [68] found that female students mostly preferred to work in gender matched pairs during pair programming.

Three papers examined changes in their CS1 outcomes after adding pair programming, peer instruction (PI), and media computation to their first-year courses. Peer instruction is a strategy where lectures are interspersed with peer instruction periods where students first individually answer questions about common misconceptions before discussing their answers within small groups. Over 17 years, one department found positive effects for all students after these interventions. In particular, Salguero et al. [105] found a decrease in failure rates, degree completion time decreased, and upper-division GPAs increased among female students exclusively. Another department that made the same three interventions reported that all CS1 students saw an increase in retention, although the increase by the end of the course was only statistically significant for male students and not females [92]. Alvarado et al. [3], working in the same department, found that grouping students by confidence and prior programming experience had no statistically significant impact on performance, and that high initial confidence predicted exam performance for male students but not females. Lewis et al. [71] found no differences in students' grades in a CS2 class after shifting to a more collaborative class structure including more emphasis on pair programming, more group projects, and an option for paired homework submission, but did find female students' drop rates decreased, although not significantly.

Zingaro [128, 129] reported in two separate papers on the effects of peer instruction on students' self-efficacy, performance, interest, and enjoyment in CS1 classes, finding significant increases in self-efficacy but no statistical differences in exam scores (males continued to perform better than females). He does suggest that peer instruction may have had some effect against the negative influences of performance goals on student exam grades.

Seven other papers reported on gains from making other interventions using collaborative learning. Four of these papers reported on an intervention involving peer mentors leading students through active learning modules to support their learning. Wai-Ling Packard et al. [119] surveyed students at a women's college about their experiences with a peer learning program and found they generally found it helpful for their learning. At a coed institution, Jamieson et al. [57] found that a higher percentage of female students who participated in the interventions went on to CS2. Xu and Jin [124] compared peer mentoring and collaborative work in a face-to-face versus an online environment, and found that female students particularly benefited from face-to-face interactions. Kulkarni et al. [67] found that peer mentoring on group projects in a CS course for non-CS majors resulted in better retention rates than a traditional CS offering.

Three papers examined flipped classrooms, although their emphasis on collaboration in the classroom varied widely. In a 2015 study by Latulipe et al. [68], "lightweight teams" were implemented and student discussion was expected after solo work but otherwise collaboration was not emphasized. In a later study by the same lead author, collaboration in the classroom was strongly structured [69]. In the earlier paper [68], the authors reported no statistical differences between genders. In the later paper with more structured interventions, they reported that female students in the flipped classroom were less likely to switch majors, whereas male students in a traditional classroom were less likely to switch. Performance for all students in the flipped classroom was higher than in the traditional class, and one-year retention rates for women and racial minorities were also higher [69]. In a 2014 study, Campbell et al. [23] found that female students spent more time on the flipped classes and found the classes more difficult.

Krause et al. [65] reported that students, especially female students, enjoyed formal learning groups in their CS1 class, although there was no evidence that the intervention impacted their intent to pursue CS and no statistical analysis was performed. Cao and Porter [25] conducted an intervention where students took two-stage exams in a class already using peer instruction. Students who took a standard individual exam followed immediately by an exam taken in small groups did better than students who took individual exams only. The female students benefited more than men on one topic. However, no gains were present by the end of the term.

5.1.3 Encourage Student Interaction. Encouraging student interaction outside of the classroom is one way to foster inclusive learning communities. Compared to well-structured collaborative learning, the primary focus is on encouraging social connections and building of community rather than improving performance on a specific academic task. The papers under this practice can loosely be grouped into three main approaches: learning communities, support sessions, and conference attendance (see Table 5).

Learning communities are cohort-based groups where students live, study, engage in social activities, or do community service

Table 5: Summary of papers identified within the practice, "Encourage Student Interaction"

Paper	Intervention	Outcomes*		
		A	C	P
[1]	Conference attendance	F+		F+
[18]	Learning Community			F+
[54]	Peer Mentoring	E+	E+	
[57]	Discussion group			F+
[58]	Conference attendance	F=		F+?
[67]	Cohorting by major	E+?		F+
[68]	Gamification	F+? E+	E+	
[71]	Group projects, pair programming		E+	F+?
[82]	Learning Community	E+?	E+?	E+
[84]	Peer mentoring		E+	E+? F+
[85]	Peer feedback	F+		
[101]	Learning Community	E+	E+?	E+
[107]	Learning Community	F+		
[108]	Learning Community	E+?	E=	
[109]	Learning Community		E-?	E-?
[123]	Learning Community	F-		
[127]	Gamification	F-	E+	

* For explanation of symbols and abbreviations, see Table 3.

together. The common thread that unites students in these communities is their courses or major. Several papers [18, 107–109] describe learning communities in which students enrolled in a first-year programming course, also have at least one other course in common (i.e., linked). In addition to the linked courses, students also participate in extra-curricular activities organized by the course instructors. In Narayanan et al. [82] students from underrepresented minorities were grouped based on prior experience, and in [67] students not majoring in computer science were grouped based on their area of study.

In general, retention rates varied by cohort. Brinkman and Diekman [18] reported that cohorts with more first-generation students had lower retention rates than the general population highlighting the impact of intersectionality on diversity efforts. At the same time, the population outcomes were better for some cohorts [18, 109]. Across multiple cohorts, evidence was found for learning communities having positive impacts on attitudes and study habits [108, 109], as well as improved sense of community and improved interpersonal relationships [107]. However, these changes did not tend to persist over time. Wright et al. [123] observed similar phenomena, noting that students who participated in their learning communities had a stronger intention to persist in computing than did students who did not participate in the community. Learning communities at this institution were characterized by common residence and mentoring and outreach activities. Over time, however, learning community students and students in the comparison group were both negatively impacted by environmental threats as their psychological and aspirational outcomes decreased over time. Rheingans et al. [101] reports on a learning community that was also residence-based, noting that students who participated in the learning communities tended to have more inclusive beliefs about the challenges and abilities of female students. This observation is important because as we strive to create inclusive environments we need to consider how the beliefs (and behaviors) of students in majority groups can influence the sense of belonging of students in underrepresented groups.

The second approach includes interventions that were tangentially related to coursework including peer mentoring activities, workshops, and help sessions. Hug et al. [54] observed the impact of participating as peer mentors for 1st-year courses. They observed that leaders had positive affective and cognitive outcomes. No statistically significant differences by gender were noted, but they suggested that these findings may indicate that serving as a leader can improve the retention of peer leaders from minority groups. Narayanan et al. [82] used several peer enrichment activities covering CS fundamentals, and found positive impacts on transfer and graduation rates, as well as a non-statistically significant improvement on standardized tests, though the impacts are difficult to attribute to any specific intervention.

Jamieson et al. [57] reported on a program in which students were enrolled in a 1-credit seminar that met for two hours twice a week to discuss programming concepts. These sessions were in addition to regular coursework but were not considered to be remedial sessions. Students were primarily graded on participation (credit/no credit) and the focus was to provide students from minority groups with spaces to meet peers and form friendships. The retention rates for women who completed the program and went on to take CS2 and CS3 were significantly higher than the general student population. Newhall et al. [84] provided help sessions with student mentors for students in CS1 and CS2 and noticed similar improved population outcomes.

Zahedi et al. [127] and Latulipe et al. [68] used online gamification environments to increase student engagement and build social connections with other students. While the primary focus was to deliver supplemental content to students, they also included engagement and social activities designed to motivate students. Zahedi et al. reported an increase in CS identity development and self-efficacy for both genders, but it is not clear that this result was due to the gamification elements. Latulipe et al. [68] found that gamification resulted in improved future marks for all students.

The third approach focused on conferences and social gatherings that occurred outside of the university. Both Alvarado et al. [1] and Janzen et al. [58] observed that sending students to the Grace Hopper celebration–a conference for women in technology–had a positive impact on retention efforts. While we highlight the benefits of this approach, we note that the previous papers presented multiple interventions, so it is difficult to determine the specific impact of sending students to such conferences. Janzen et al. specifically asked students to what extent the conference had on them remaining in the major, and the feedback from all participants was positive, but not significantly different by gender.

Learning communities seem to have a positive effect on the affective outcomes of students; however, these attitudinal changes may not persist over time. These communities are also beneficial in improving the retention of students from traditionally underrepresented groups, including women. Learning support sessions have a positive impact on population outcomes and possibly have cognitive benefits for students. Attending conferences and social gatherings positively impact retention and the affective outcomes of students from minority groups. Overall, interventions that encourage student interaction are mostly beneficial for students regardless of their gender, improving affective, cognitive, and population outcomes. However, it is worth highlighting that many of the interventions in this section are costly especially given that questionable long-term impacts. More exploration and research is needed on more affordable ways to foster student interaction.

5.1.4 Grow an Inclusive Community: summary. There is a large body of research across this principle, but the evidence is largely clustered around implementation of a few specific practices. Even then, the evidence is sometimes contradictory and results are often difficult to attribute to any specific practice.

The evidence for avoiding stereotype was quite strong, and mostly involved clear and specific implementations of the practice of mitigation or direct evidence of the impact of negative stereotypes. It is clear that the use of materials and activities (e.g., assignments, images) that are stereotypically associated with men (or particular subsets of men, i.e., geek culture, Brogrammers) or that portray women in a negative light have negative affective impacts on female students. There appears, however, to be relatively few studies providing direct evidence of the impact of stereotypes on cognitive or population outcomes of female students (but see [28]). While it is natural to link stereotypes with affect, there is clearly an opportunity to study other areas of impact. As we discuss later in this report, however, the relative sparseness of studies around this practice may be an artifact of our method of including and excluding studies.

By far the most studied practice in our sample was collaborative learning. The reported outcomes were generally positive. While the implementations of peer mentoring and peer instruction varied quite widely across studies—making comparisons difficult—the results were almost uniformly positive across all types of outcomes (affective, cognitive, population). Outcomes for pair programming were more mixed and may hinge on the level of care taken to match pairs appropriately and to help students learn to work effectively in pairs. That said, the mechanisms by which the practices impact women's participation is unclear. Is it because the practice is a more effective teaching strategy and boosts efficacy and confidence? Or does the practice work because it helps students form relationships with other students who then provide support and encouragement? Or is it some combination of those elements and others? These are open questions.

Most directly related to the goal of creating community were studies of learning communities which groups students into cohorts across multiple courses. Most of these papers reported positive overall results for students, but it is difficult to attribute results to any specific component of the interventions, as these were almost always very complex interventions that spanned many practices from multiple principles. In addition, there is some evidence that the impacts of these communities may not persist.

5.2 Make It Matter

The NCWIT Engagement Principle, Make It Matter, recognizes previous research that suggests that students are more motivated, perform better, and are more likely to persist when they can see how a lesson connects to their experiences, interests, goals, and values. In addition, students who do not fit the stereotype of someone who does computing may need even more explicit connections for them to envision themselves in the field. Educators are encouraged to "help students connect to computing by connecting computing to their students' lives" [83]. There are four practices associated with this principle, and papers addressing each of these practices are discussed in the following subsections:

- Use Meaningful and Relevant Content
- Make Interdisciplinary Connections to CS
- Address Misconceptions About the Field of CS
- Incorporate Student Choice

5.2.1 Use Relevant and Meaningful Content. Students are more likely to persist in the face of a challenge when what they are learning is relevant to their life experiences and goals. The practice "Use Relevant and Meaningful Content" challenges instructors to find projects, assignments, and examples that are culturally and personally relevant to their students, to explain to students how a particular computing idea is relevant in different contexts and fields, and to broaden students' understanding of what computing is and who does it.

A large number of papers address this practice, so we have separated the discussion into those that address this practice alone and those that address this practice along with one or more practices from another principle. There are also multiple papers that address both this practice and others within the "Make It Matter" principle. These are discussed within the sections addressing other practices, to avoid repetition while keeping the description of the principle coherent.

Use Relevant and Meaningful Content only. We first consider the papers that only employ an intervention that falls within the "Use Relevant and Meaningful Content" practice, as summarised in Table 6. It should be noted that some papers explicitly justify why their interventions should be relevant and meaningful to their students. However, some papers refer to capturing students' interest and engaging them with novel interventions, while requiring the reader to infer why the approach may connect with students. This section includes all papers that express the intention of connecting with the students' interests, regardless of how well they justified the particular intervention.

One key paper by Guzdial [47] reviews evidence for the effectiveness of using Media Computation (MediaComp) as a way to engage students. This approach "introduces computing through exploration of data abstraction related to digital media" such as images, video, sound and web pages. The author considers papers from over a decade and suggest that there is support for the claim that "contextualization aspect of MediaComp" effectively connects with female students' perception of relevance. This is supported, in part, by the fact that the course regularly attracts a substantial number of female students (e.g., the course regularly enrolled more than 40% women). Female students were also found to have success

Table 6: Summary of papers identified only within the practice, "Use Relevant and Meaningful Content"

Study	Intervention	Outcomes*		
		A	C	P
[13]	Spiral-based curriculum design	E+, F+		
[21]	Robotics in software sys course	E+?, F+?	E+?, F+?	
[31]	Challenging course assignments	E+-, F+-		
[44]	Utility value	F+	E=	
[47]	Media computation	F+		F+
[48]	Physical computing	E+	E+	F+, M+?
[53]	Using data and real-world problems		E=	
[66]	Mindset and metacognition tasks		E=, F+?	
[70]	Combining video games with lectures		E+	
[77]	Robotics in prog course	F+		
[88]	Story programming		F+?	
[97]	Food-focused activities			F+
[99]	Animation and robotics programming		F+	
[102]	Animation & robotics in CS1	E+?	E=	
[106]	Using creative videos to teach algorithms		E+	

* For explanation of symbols and abbreviations, see Table 3.

rates in the course similar to male students, and across multiple US-based institutions.

Remshagen and Rolka [99] discuss an intervention where part of the Media Computation approach had been adopted through the use of animation to teach programming (Greenfoot) in a CS0 course. Student outcomes, across the board, had been poor, with a 50% failure rate. To address this, they introduced robotics (IPRE Scribbler) alongside animation in the course. While female students were more likely to find robot exercises difficult (by a slight margin), the intervention did result in a statistically significant difference between male and female students, with female students performing better. The paper recommends the use of robots only in "selected exercises," partly because of frustrations that can occur because of technical problems using robots. They also recommend allowing students to take robots home for at least a day so that they can work on them at home and "show off their work to their friends." Effectively, this is a way to make the work socially relevant to students.

Another study involving robotics in a U.S. CS0 course (again with Scribbler) was carried out by McGill [77], specifically to examine the effect on student motivation and how it varied with gender. Using Keller's Instructional Materials Motivation Survey, the author found no significant difference between male and female students on measures of motivation. They did find that female students were more likely than male students to agree with the statements "learning how to program with the robot was useful for me to learn how to program effectively" and "it was a pleasure to work with the robot to learn how to program." Interestingly, women were also more inclined to agree that "using the robots to learn how to program was intimidating to me." (Differences were statistically significant despite a relatively small sample size of 13 female and 22 male students.) Potential technical problems associated with robotics were also highlighted in this paper.

Caldwell and Jones [21] adopted Parallax Sumobot kits to engage students in programming their robots in wrestling and dancing competitions for second-year Introduction to Computer Software Systems Course with assembly programming. The course was implemented at a Historically Black College or University (HBCU) in the U.S. with 87% male and 13% female students. The authors report that female students were initially hesitant to engage in building the robots but that gradually all of the students grew excited and engaged with building, programming, and decorating the robots. The paper reports anecdotal, positive observations from the instructor as well as nearly unanimous positive responses from student surveys. Female students averaged 92% and male students averaged 84% on a quiz for assembling the robots, however, no statistical tests of differences were reported.

A lower-cost approach to physical computing is outlined in Hallak et al. [48] where Raspberry Pi computers were used both in face-to-face and online contexts. Traditional programming courses that used C++ were adapted by using colored LEDs (connected via GPIO) as an alternative output for various algorithmic problems. Compared with previous offerings of the introductory programming courses, grades are reported to have increased by 25% across the class (with no gender differences observed), and lab attendance for female students increased by 10% (the figure for males students was not reported). Attitudes towards the course improved for all students. The only gender difference on affective outcomes was that female students tended to be more positive in their answers to the question "does the use of Raspberry Pi mini-computer in the lab as a group project help increase your communication level with your classmates?"

Integrating programming lectures with virtual robotics using Cargo-Bot was studied by Lee et al. [70]. Cargo-bot provides a simple visual programming language with which to control a virtual robot to solve problems involving stacking crates. Exercises with Cargo-bot were woven into four lectures on recursion with Java within a CS2 course, while a control group received the standard Java lectures only. Performance gains within the intervention group were greater than for the control group but the difference was not statistically significant. Students who completed eight or nine CargoBot puzzles scored higher than the control group by a statistically significant margin, but gender was not found to be a significant factor in the improvement.

Another approach is using creative videos to teach algorithms, studied by Schreiber and Dougherty [106]. In order to increase conceptual understanding and confidence in binary search and selection sort algorithms, the authors created a video series for each algorithm. The researchers were particularly interested in investigating the effects of these videos on students' technical grasp and self-confidence about understanding and applying the algorithms.

They used pre- and post-surveys in which students answered questions about their familiarity with programming, music, and the given algorithm, and how confident they were with their understanding of the algorithm. Students showed substantial improvement in their knowledge of the complexity of each algorithm but the authors do not report gender differences. Regarding students' perceived confidence, there was a significantly positive difference between pre-survey and post-survey questions for the entire sample, evident in both genders.

Rolka and Remshagen [102] assessed different contextualized learning tools in order to explore whether they increased student performance in introductory CS courses by influencing attendance. Female students had overall high grades in the course as well as higher attendance but it is difficult to assess whether the contextualized learning tools are the cause.

Krause-Levy et al. [66] investigated whether combining previously identified promising intervention techniques within a standard structure of a CS1 course significantly increased students' performance and what value, if any, students perceived from these interventions. The interventions implemented were: 1) mindset interventions; 2) thinkathons, where on-computer coding exercises were replaced with paper-based scaffolded code comprehension and writing exercises targeting the same learning outcomes; 3) metacognitive tasks, used to support study skill development and inclusion and values reflections to foster growth mindset and belonging. The interventions did not generally appear to make a difference and few gender-based differences were observed. The exception was in exam performance, where the female students showed a small but statistically significant improvement (7%) while male students showed a small and not statistically significant decline (7%). The paper does not provide sufficient evidence to determine whether these interventions would, in fact, improve performance. However, the paper provides a basis for further investigation, especially around how it could potentially improve the performance of female students.

The potential to leverage programming difficulty to improve understanding and perceptions of non-majors was explored by Crosby et al. in [31]. The CS0 course for non-majors was redesigned by replacing the standard intro to CS course — similar but less deep than that for CS majors — with a course focusing on introducing Python with a variety of lab assignments that included ASCII art and games of chance. Analyses of survey data showed that female students were more likely to dislike the design part of the assignments, while preferring coding and debugging. Interestingly, student attitudes towards coding changed from the beginning to the end of the course in various ways, including by gender. For example, female students interest in coding tended to increase across the course, whereas male student interest tended to drop. However, it's not clear whether a more standard course would also have had this affect on student views.

Story programming, an approach to introduce computing concepts in the context of fictional stories without using a computer, was explored in Parham-Mocello et al. [88]. An introductory orientation course was redesigned to introduce computational thinking. No differences by gender were found in dropout rates, grade distribution, or in the ratings of the computational thinking or coding activities was seen in the different groups. However, when the data from all sections were combined, females were significantly more likely to report higher engagement in coding activities than males. Overall, it appeared that the choice of programming language had greater impact on student outcomes than the approach (story programming vs traditional approach); students using Haskell had stronger positive reactions to the CT activities than students using Python.

Geerling et al. [44] explored how exposure to a topic's utility value can lead to positive outcomes for women in computing. In a small introductory CS course focusing on web programming, students were randomly assigned to an intervention or control group. Students in the intervention group were provided with a list of nine applications that the course would cover and asked to select the three that were most personally useful to them. Students in the control group read only about the general usefulness of the material and were not prompted to connect the uses to their personal interests. Students were surveyed throughout the semester asking about how they were feeling, including interest and confusion. Women in the intervention reported less confusion than the men, but women in the intervention that reported high levels of confusion were more likely to have an increase in interest over time. In contrast, men exhibited this connection between high confusion and high interest regardless of being exposed to the treatment. No gender differences in performance were found between the control and treatment groups but the sample size was very small.

The impact of using web services to incorporate data and real world services was explored by Hosack et al. in [53]. Over two years, 586 students participated in the intervention. Each semester, two sections were offered: the experimental section used web services problems and the control section did not. Scores on a common final exam were analyzed, taking into account major, gender, GPA, and class rank. Gains were observed in three semesters; however, the poor results of the third semester cancelled out the gains in the other three. No significant differences were observed by gender.

The use of food-focused activities to improve Computational Algorithmic Thinking (CAT) in a cohort of African-American women was explored by Rankin et al. [97]. Students participated in a "Dessert Wars Challenge" designed to support the development of CAT skills. Thirty-five participants across two years used self-reflection journals with guided prompts. High retention rate of students – 96% and 100% for the two experimental semesters – support the authors' conclusion that the intervention provided motivation and situated context for students to develop their CAT skills. In contrast, when the course was taught the following year without the food-related activities, the retention rate decreased to 79%.

An approach to teaching spiral-theory-based cybersecurity models within existing CS courses is outlined by Basu et al. [13]. A pre- and post-course test was used to explore whether students found the course useful and interesting, as well as evaluating subject knowledge. The students met the learning objectives of the course and no difference were observed between men and women, with the exception of women scoring higher on one learning level.

Use Relevant and Meaningful Content with other principles. In this section we review papers that use "Relevant and Meaningful Content," along with other engagement practices from principles other than "Make It Matter" [83]. These are summarised in Table 7.

Table 7: Summary of papers identified within the practice, "Use Relevant and Meaningful Content" also involving practices from other principles

Study	Intervention	Outcomes*		
		A	C	P
[30]	Small CS class with a range of interventions			F+
[68]	Gamification	F+?, E+	E+	
[92]	Media computation, pair programming, peer instruction			F+?, M+
[105]	Media computation, pair programming, peer instruction		E+, F++	E+, F++

* For explanation of symbols and abbreviations, see Table 3.

Combining the influential Media Computation course with other practices of pair programming and peer instruction, Porter and Simon [92] describe the effects on retention after three years of running the new course, including comparisons with a control group using the original curriculum and the same instructor. The authors report positive results on retention and in students passing the course, which improved from 71% to 89%. Effects were more positive for male students than for female students but the difference was not statistically significant. When taking a broader view of retention, i.e., from all students enrolled at week 1, statistically significant improvements were noted for students in the intervention section compared to those in the control group; no gender differences were noted.

A later paper by Salguero et al. [105] examines the effect of the same changes over a longer period of time, looking both at retention/failures and GPA. Again there are significant improvements in retention and failure rates among female students in the intervention group, and these were mirrored among male students. Interestingly, over the same period there was also a statistically significant improvement in the failure rate of female students within the control group, but no improvement in retention of females. GPA scores were similarly and significantly improved for female students in both the intervention and control groups over the same period. The authors conclude that in "examining the scale of the benefits, women appear to benefit more for some metrics and men for others." They also note that "we cannot distinguish which [practices] were more important or if they were needed in combination."

The same trio of practices (Media Computation, pair programming, peer instruction) were also adopted by Latulipe et al. [68], with the addition of lightweight teams and gamification. Since the data presented focuses on student opinions about the use of teams, a detailed discussion of the paper is included in section 5.1.

Cohoon and Tychonievich [30] also explore the impact of a combination of practices, most significantly in the areas of giving effective encouragement and providing opportunities for interaction with faculty but also by carefully considering the impact of different types of application on the interest of students. This new course (CS1X) had significantly more participation from female students (64% of the class) than comparable CS1 courses. A detailed

Table 8: Summary of papers identified within the practice, "Make Interdisciplinary Connections to CS"

Study	Intervention	Outcomes*		
		A	C	P
[11]	Theme-based computing curriculum	E+		E+
[14]	Computer Science Principles	E+	E+	F+
[15]	Open-ended game development project		F-?	
[58]	Multi-intervention effort	F=		F+?
[67]	Cohort-based multi-intervention effort	E+?		F+
[86]	Writing-intensive CS1	E+-, F+		
[93]	Personally relevant projects	E+, F+		
[113]	Animation module	E+	E+?	E+?
[114]	Serious game development	F+		
[115]	Interdisciplinary problem-based approach			F+?
[120]	Music-based programming	E+		E+

* For explanation of symbols and abbreviations, see Table 3.

analysis of the popularity of different application topics by gender is provided, showing marked differences between men and women on interest, and a general trend of men being "generally more pleased by applications than were females." The authors conducted focus groups to garner student reaction to the various practices adopted, and found that the relevance of the content did not come up often in students' comments. No gender differences in this qualitative data were noted.

5.2.2 Make Interdisciplinary Connections to CS. This practice involves connecting computer science to other fields, such as medicine, the humanities, and media. By showing how computer science concepts and skills are used in other fields, especially those that tend to engage a lot of women students (e.g., biology, humanities), those who may not have considered computer science might find it newly relevant to their lives and career interests [83]. The papers identified as testing this practice are show in Table 8.

Integrating multiple other disciplines, Stone and Clark [113] developed Problem-Oriented Animated Learning Modules for Introductory Computer Science (PALMS for CS1)/ a collection of animated examples of problem solving, algorithms, and C++ programming from other disciplines. PALMS particularly featured examples from multiple areas of the natural sciences to appeal to non-majors in their CS1 course. The authors report data from two years where PALMS was adopted, finding no significant difference in course withdrawal and overall course average for men and women.

In Sweedyk [114], an existing game development course was adapted to focus specifically on "serious games", including bringing in topics from other disciplines e.g. history or physics. Although the sample size was very small, positives outcomes were reported for female students including higher engagement, better course results, and positive attitudes towards serious game development.

In a CS2 course, Black [15] introduced Android development with an example application featuring GPS usage. Across seven semesters, students (n=141) also completed open-ended projects and the paper reports anecdotal observations of projects that incorporated connections to other disciplines. The authors noted lower than average scores on project completeness and correctness for women and minority men but these differences were not statistically significant.

Used in both high school and college-level computing courses, EarSketch is a learning environment that connects music composition, production, and remixing with introductory computing [120]. Over three semesters from Spring 2016 to Spring 2017, Earsketch was studied among students at the college-level, with a total of 206 students receiving Earsketch lessons and 163 students in a comparable control group. The intervention had a positive impact on the students' intent to persist, but this was independent of gender and underrepresented minority status. Furthermore, multilevel modeling methods showed that attitudes towards computing influenced students' intent to persist.

Rader et al. [93] explored differences in interest between male and female students in a range of topics used in CS1 and Software Engineering classes. Some topics were game-related and others had humanitarian or other practical aspects to them. Overall, women rated the humanitarian projects higher than male students, but the highest ratings went to the projects they found to be personally relevant. However, the sample size was small, and no analyses of statistical significance were reported. Similarly, adding interdisciplinary topics to a problem-based Introduction to CS course are described in Tartaro and Cottingham [115]. Over a period of three years they achieved much higher female enrollment rates than the national average by including topics such as bioinformatics, business analytics and creativity for entertainment.

Papers that investigate this practice alongside other practices within the "Make It Matter" principle are discussed in the following section.

5.2.3 Address Misconceptions About the Field of CS. This NCWIT Engagement Practice focuses on addressing misconceptions students have about the field that may prevent them from taking computing courses or pursuing a career in computer science. Many students have overly narrow ideas of what computing is about, what the work is like, and the kinds of people who typically do it. NCWIT suggests that educators can address these misconceptions by 1) illustrating the diversity and breadth of work a computer scientist can do and as well as the diversity of people who can do it, 2) emphasizing that success comes from practice, and 3) discussing the advantages and rewards of computing careers [83]. Ideally, misconceptions can be addressed by simply showing things counter to the misconception without actually invoking the misconception and risk reinforcing a stereotype. The papers identified as testing this practice are show in in Table 9.

Table 9: Summary of papers identified within the practice, "Address Misconceptions About the Field of CS"

Study	Intervention	Outcomes*		
		A	C	P
[1]	Multi-intervention effort	F+, E+		F+
[2]	Grace Hopper attendance	F+		F+
[14]	Computer Science Principles	E+	E+	F+
[18]	Cohort-based service learning program			F+
[57]	Enrichment workshop			F+
[62]	Research experience for undergraduates	E+?, F+?	E+?	
[67]	Cohort-based multi-intervention effort	E+?		F+
[86]	Writing-intensive CS1	E+-, F+		
[123]	Living learning community	F-		

* For explanation of symbols and abbreviations, see Table 3.

Brinkman and Diekman [18] investigated how to address the lack of diversity in technology by applying communal goal congruity, with an emphasis on working with others and in the service of others. This approach addresses the misconception that computing is only utilitarian rather than humanitarian, i.e., it is not the kind of work that can be socially relevant. At a public North American research university, the authors implemented a service-learning program with cohorts in computer science, computer engineering, electrical engineering, and software engineering. They piloted the program by partnering with a local service organization that teaches coding clubs for primary school students. The program also offered scholarships (with award amounts determined by level of financial need) as well as counseling and cohort community-building activities. They found that the matriculating cohort was more diverse than the admitted cohort and only a weak association between the amount of funding offered and whether or not the student matriculated into the program.

Alvarado and Judson [2] report on the effects of attending the Grace Hopper Celebration (GHC) on female students' decision to major in computer science and follow corresponding career paths. GHC is an annual conference celebrating the accomplishments of women in CS, combining technical talks, targeted workshops, panels focused on issues facing women in the field, and networking events. Student participants are exposed to diverse women role models and provided with networking opportunities with other women and potential employers. The researchers surveyed 60 female GHC attendees from Harvey Mudd College before and after the conference trips in 2009 and 2010. The results indicate that GHC had a positive and statistically significant effect on attendees, especially those with previous CS experience or interest. The authors found that students who attended GHC enrolled in a second CS course at a significantly higher rate than those who did not attend GHC, and 25% of those who were not considering a CS major prior to attending GHC chose to major in CS after attending the conference.

Behnke et al.'s [14] work also had positive outcomes on student's attitudes towards CS by introducing a curriculum framework of Computer Science Principles, in which computational thinking is explored with a focus on creativity and the impact of CS on the world. Although there are no comparative statistics reported on how male and female students' opinions changed, the affective outcomes for all students were positive and with female enrollment at 29%, there were definite gains in participation.

Another approach that demonstrated a positive impact was described by Kulkarni et al. [67]. The paper reports on a pilot program, Promoting INclusivity in Computing (PINC) at San Francisco State University that aimed, among other things, to improve diversity in their computing program. PINC is a set of computing courses for students majoring in other fields such as Biology and Biochemistry, majors which enjoy fairly high enrollments by women and students from underrepresented minority groups. The program included cohort-based program structure with a dedicated curriculum, near-peer mentoring, project-driven and cooperative learning. Surveys of participating students showed that the majority felt less intimidated about computer programming, and felt they understood its utility and necessity. The gender composition of PINC program was 73% female compared to the corresponding 16% in the traditional CS course. Moreover, while the dropout rate for female students in the traditional CS course was 15% by the end of the program, for the PINC program, it was only 2%. Similar trends were observed in the numbers of URM students. Given the range of practices that were adopted it is not possible to pin down the success of the approach to a single practice, but seems to have a similar approach to the "Green" curriculum at Harvey Mudd [1].

Another method of addressing misconceptions about computing is through workshops that complemented a CS1 course [57]. CS1 students were recruited to participate in enrichment workshops–patterned after those developed by Uri Triesman at UC Berkeley–which focus on group work and community building more than technical details. In these two-hour twice-a-week workshops, students learned CS1 concepts but in a way that emphasized teamwork and interdisciplinary skills. Women and non-Asian men of color who participated in the intervention were more likely to go on to CS2 than those not in the workshops, but this outcome may be due to selection bias of who chose to participate in the program. No changes, or differences with non-participants, in attitudes toward computing or improvements in grades were observed after participation in the workshop.

Wright et al. [123] evaluates the impact on three cohorts of first-year women at Rutgers University who took part in a Computer Science Living-Learning Community (LLC). The participants were recruited from women who indicated their intent to major in CS on their admission applications. In the LLC, students live together as well as study together, and are provided with a range of support. While the majority of the intervention focused on community (NCWIT principle 3: Grow an Inclusive Student Community), some work on addressing misconceptions about the fields of computing was included. The women involved were more likely to engage in extracurricular computing activities, but some negative affective outcomes were noted.

O'Hara et al. [86] evaluated a three-week writing-intensive course, "Language & Thinking," required of all freshmen at the liberal arts-focused Bard College that had newly integrated a computing component. The change was made to expose first-year students to coding and computational thinking with the intent of positively influencing the attitudes of participating students. Students were given attitude surveys before and after learning through unplugged activities and in contexts relevant to the students. The authors report "modest changes in attitudes" including female students developing a greater appreciation of computing. Importantly, a mean increase in the belief that "computer science is just learning how to program in different languages" was observed. The authors attribute this to holding coding sessions outside of the normal class period. We suggest that it may also be due to the coding-intensive focus of the new material. These findings suggest that the content of a program/intervention, as well as the way it is structured, are important considerations when attempting to increase interest in computing. *Some interventions may, in fact, reinforce stereotypes about computing rather than dispel them.*

5.2.4 Incorporate Student Choice. The final engagement practice within the "Make it Matter" principle–incorporate student choice–suggests that allowing students to choose from among a set of problems or assignments, or to choose the topic area to be addressed in a particular assignment helps make computing topics more relevant by allowing students to make the connection to their own interests [83]. Seven papers, as shown in Table 10, included aspects of this practice.

Table 10: Summary of papers identified within the practice, "Incorporate Student Choice"

Study	Intervention	Outcomes*		
		A	C	P
[11]	Theme-based CS1 courses	E+		F+
[15]	Open-ended game development project		F-?	
[58]	Multi-intervention effort	F=		F+?
[67]	Cohort-based multi-intervention effort			F+
[86]	Writing-intensive CS1	E+-, F+		
[110]	Open-ended assignments		F+	
[116]	Robot-based CS2 course	E+		F+

* For explanation of symbols and abbreviations, see Table 3.

One approach to student choice involves allowing students to choose a recurrent theme to explore throughout a course. For example, Barr [11] describes CS0 and CS1 courses where students choose from themes of big data, robotics, game development, artificial intelligence, media computation, and engineering applications. A study of 58 sections across six years found an increase in overall enrollment in intermediate courses, in general, as well as an increase in the percentage of women in the major. Surveys revealed that 57% of students enrolled in themed courses expressed interest in CS or applications of computing while 58% indicated they may take additional CS courses.

Overall, interventions that highlight students choosing between alternative themes or offerings for introductory courses consistently highlight improvements in recruiting female students. However, more work is necessary to reconcile potential negative impacts on students' performance in subsequent classes, and to gather more insights into impacts on affective outcomes.

Besides offering students a choice of course theme, several studies investigated incorporating student choice in course projects. In a within-subject, quasi-experimental design , Sharmin et al. compared open-ended project assignments to control projects with conventional requirements in a CS1 course [110]. For the open-ended projects, students were afforded freedom to choose the application for their assignment. Note, however, that all assignments, regardless of condition, had to be game-related. The study reports on students' self-efficacy measured using a pre-test survey and found that it was associated with higher grades on assignments. While gender had a significant effect on the assignment grade (female students scoring significantly higher), there was not a significant difference between the experimental (open-ended) and control (conventional) treatments. It may be that confining choice to game design and implementation may have suppressed the desired effect and that providing options to, e.g., design an application for social good, may be more effective.

Ustek et al. [116] describes a robot-based second programming course designed and developed at Grinnell College with the aim to increase students' interest in CS courses. The course was developed with the assistance of four students; specifically, the students were responsible for designing, testing and refining the course materials by drawing on their previous experiences with CS courses. The evaluation of the new course was based on surveys of the enrolled students, course performance data, and end-of-course evaluations. While students' performance did not significantly change compared to the previous years, 65% of the students replied positively to a question on the usefulness of the robot-design course to their understandings. Course enrollment increased moderately, but no statistical measures are reported. The authors suggest that a part of this increase may be due to the enrollment of more females in the whole introductory three course sequence.

Some papers (e.g., [58, 86]), did include some student choice, but this was not the only NCWIT practice to be addressed and there was little or no data on how effective the student choice element was by itself.

5.2.5 Make it Matter: summary. There is a substantial body of work across this principle, with a particular focus on using meaningful and relevant content. Overall, the results are mixed, and one issue that has become clear is that there is a need for further research into this principle. Many papers suggest interesting results but need further work to determine whether these results are statistically significant, apply across different student groups, or can be attributed to the intervention discussed rather than to confounding factors. Several papers are unclear about reporting the significance of their results, or were carried out on very small populations. Carrying out conclusive research in this field is challenging, but we emphasise the benefits that can arise through doing this.

There was, nevertheless, evidence of significant positive impact of interventions using the NCWIT recommended practices under this principle. Twenty-three out of 46 papers reported some form of positive impact on women's enrollment and engagement in CS courses. More than half (26) of the papers noted improvements in students' attitudes to CS, split between those that specifically improved the attitudes of female students and those that improved it for all. However, several papers reported mixed findings where attitudes improved in some respects and remained the same or worsened in others. Fourteen papers reported positive impacts of interventions on test and exam scores; these improvements were frequently observed for all students, rather than only for women.

5.3 Build Student Confidence & Professional Identity

Because computing has come to be associated with some fairly strong stereotypes about who does computing (i.e., white and Asian men), anyone who does not fit the stereotype may have difficulty seeing themselves in the field. They may also be less likely to have people supporting them in their pursuit of computing. Instructors can help by building student confidence, modeling inclusive behavior, and providing experiences that support all students developing computing identities [83]. The NCWIT engagement principle, Build Student Confidence & Professional Identity, includes four practices, and papers addressing each of these practices are discussed in the following subsections:

- Give Effective Encouragement
- Offer Student-Centered Assessment
- Provide Opportunities for Interaction with Faculty
- Mitigate Stereotype Threat

5.3.1 Give effective encouragement. Effective encouragement is critical for students' self-efficacy, and as such, its role in retaining women in male-stereotyped fields is important. Given the popularity of interventions such as growth mindset in other fields, we were surprised that there were so few papers in our sample that investigated this practice (see Table 11). We also observed that effective encouragement was often studied as one intervention among many, complicating the assessment of this practice.

Table 11: Summary of papers identified within the practice "Give Effective Encouragement"

Paper	Intervention	Outcomes*		
		A	C	P
[19]	Growth Mindset	E+	E=?	E=?
[60]	Problem-Solving Strategy Detection and Development approach	E+	E+	
[104]	Learning community	F+?	F+	E+
[123]	Living learning community	F-+		

* For explanation of symbols and abbreviations, see see Table 3.

One study that explicitly looked at a growth mindset intervention is Burnette et al. [19]. Using a rigorous double-blind experimental design, they sought to understand the impact of a set of online growth mindset modules on improving performance and promoting interest in computer science. The content of the modules included explaining the benefits of a growth mindset, reinforcing the message that ability for computer science can be developed, presentations of role models that suggested tips for success and effective learning strategies, and a "saying is believing" writing exercise to reinforce the growth mindset message. Introductory computing students from seven US colleges were assigned to control or experimental conditions. Of the 500 participating students, 29% identifying as women. Results suggest that the mindset intervention increased students' career interests and intrinsic value in the field but it did not appear to impact final grades. There was no support for the hypothesis that growth mindset training would offset women's tendency to be less interested in STEM fields or ameliorate performance gaps. However, the authors note that students in this sample tended to score highly on growth mindset to begin with which would limit the impact of the intervention. The intervention might well be effective for those starting with more fixed mindsets.

Another more systemic approach to encouragement are learning communities. Russel [104] studied the effects of a learning community at MIT, called "an experimental study group" (ESG), on students' achievement and major choice. ESG is a multi-faceted approach, including small classes, mostly female instructors, and targeted advising and mentoring. No significant benefits to the general population were noted; however, participating in these learning communities resulted in an average increase in academic performance for women. The author, however, is reluctant to make generalisations due to the small sample size. Also, given the multifaceted approach, it is difficult to separate the effects of of the different interventions. The author does suggest that having female instructors is particularly beneficial to female students but it is not possible to know whether this is due to the types of encouragement they may provide, a role model effect, or some other aspect.

A similar study conducted by Wright et al. [123] looked at a CS living-learning community (LLC) for first-year women. This LLC also incorporated multiple interventions including mentoring, academic and professional development activities, a community of peers for friendship and academic support, and exposure to the issues and applications of CS. Results were mixed regarding students' CS experiences. This intervention could have encouraged students' self-efficacy and fostered the development of growth mindsets; however, their evaluation showed weaker computing self-efficacy and stronger beliefs that computing ability is innate at the end of the experience.

5.3.2 Offer Student-Centered Assessment. Student-centered assessment is any practice that helps students examine their own learning and knowledge. Because of a number of social factors, individuals who may wonder if they "fit" in a major may get discouraged by difficult problems and set backs. That is, they may internalize the difficulty as something about them, rather than see it as a feature of the task. Providing feedback that is timely, contextualized, and actionable–and assistance on how to reflect on it–can encourage persistence [83].

Four papers implemented interventions that highlighted this principle (see Table 12). First, in a web-based cybersecurity course, Raina et al. [95] redesigned modules to implement the principles of segmentation to reduce students skipping the content: this included breaking the content into sections with checkpoint questions which provided immediate feedback on submission and students could not progress to the next section until all the questions were answered correctly. They also included text-based checklists, multiple choice, and constructed response questions, for which students received immediate feedback until the third attempt and immediate elaborate feedback thereafter. The authors argued that the module led to significant increase in student engagement, both for women and men.

Table 12: Summary of papers identified within the practice "Offer Student-Centered Assessment"

Paper	Intervention	Outcomes*		
		A	C	P
[95]	Segmentation and Feedback	E+F+		
[113]	Problem-Oriented-Animated learning Modules	E+	E+?	E+?
[69]	Gamification	F+?,E+	E+	
[127]	Gamification		F-	E+
[49]	Exam question ordering	F+	F+	

* For explanation of symbols and abbreviations, see see Table 3.

Stone and Clark [113] report on a two-year approach that involved Problem-Oriented-Animated Learning Modules (PALMs) for CS1. This is a set of animated learning modules that aims to enhance students' engagement, success and retention by including animation, video, audio and storytelling. Each PALMs module included assignments and progress indicators/benchmarks that allowed students to chart their progress and review questions. They also included a set of graded lab exercises and in-class quizzes. The students found the built-in review questions–which allowed them to check their knowledge–to be one of the most helpful components of PALM modules. Male students were significantly more likely to pass the class than other students but no significant gender differences were noted in course withdrawal and overall course average. That said, the low proportion of women in the sample (16%) limits generalizability. Also, given the multiple elements in this intervention, it is not possible to attribute these findings to the assessment elements alone.

Also implemented in an introductory course, Harrington et al. [49] investigated the role of question ordering on students' performance and confidence. The authors assigned students randomly to exam conditions with questions being ordered either from easy to difficult or difficult to easy. At the end of the exam, the students were asked to predict their marks. Interestingly, while ordering the questions from hardest to easiest had a negative impact on the performance of international students, it significantly raised both the performance and confidence of female students. The intervention had virtually no effect on the accuracy of student predictions.

In recent years, "gamification" has been a popular intervention hypothesized to enhance students' interest and boost their confidence partially because of the performance feedback it can offer. The results of these studies, however, are inconclusive and tend to point to negative impacts for some groups. For instance, in a study conducted by Latulipe et al. [69], the authors employed gamification (e.g., stamps, leaderboard and tokens) in an attempt to improve students' motivation, encourage them to work harder, and make the course more engaging. The results suggest that the gamification elements had a positive influence on the students' work ethic and self-reliance, however, the authors argued that the use of these elements needs further refinement. Along the same lines, Zahedi et al. [127] examined the effects of gamification (virtual points and leaderboard) on women's CS identity development and self-efficacy. The authors found no indication that gamification affected computing identity and self-efficacy. The quantitative analysis showed that total virtual points positively correlated with students' grades, but the qualitative analysis revealed that female participants were indifferent or negatively oriented towards gamification. Overall, their results do not provide concrete evidence of positive effects of gamification on women's CS identity and self-efficacy.

5.3.3 Provide Opportunities for Interaction with Faculty. The NCWIT Engagement Practices Framework posits that interacting with faculty can be a powerful way to give students encouragement and support, to impart tacit professional knowledge, and to help students begin to see themselves as part of a community of computing professionals. These interactions can take place in a variety of settings (e.g., class or lab, office hours, other informal settings), and can be important for sustaining student interest in computing [83]. In addition to faculty, interaction with other instructors, such as teaching assistants, and with industry professionals may be similarly important and so studies addressing these types of interactions are also considered here.

Sixteen papers tested interventions related to the practice of "opportunities for interaction" with these types of individuals. An overview of these papers is provided in Table 13 with papers grouped by the primary type of initiative. The first set are in-class activities that are specifically designed to allow more intense interaction with instructors. The second set are studies comparing face-to-face courses with online courses and, indirectly, the impact the latter may have on students-instructor interaction. The third set are studies of larger educational initiatives, such as Living Learning Communities. In these studies, opportunities for faculty or professional interaction were one part of multi-pronged initiatives that makes it difficult to assess the independent effect of providing opportunities for faculty and professional interaction. About half of these studies also incorporated interaction with industry professionals. The last section are studies that looked mainly at these kinds of interactions.

Table 13 also provides information on who the student is interacting with: teaching assistants (TAs), administrative staff (Ad), faculty (F) or IT professionals (IT). "Faculty" may be instructors or other faculty with whom students would encounter (e.g., advisors, learning community leaders). "TAs" are students who are employed or receive academic credit to support teaching and learning; they include undergraduate students who have completed the same course as those they support to more experienced PhD students. The roles that TAs play in the initiative are also noted: designing courses (D), running teaching sessions (T), or mentoring (M) one-on-one or in small groups.

Interventions that sought, in part, to increase interaction with faculty ranged from one-off activities such as:

- students visiting instructor's homes for events [107, 109],
- start of course instructor-student interviews that established relationships and expectations [26],
- dedicated faculty to support learning communities [82, 91, 104, 123],
- extra academic support sessions [101]

to whole-scale changes such as:

- changing the format of interaction with instructors (from face to face to online) [55, 124],
- increasing the amount of contact (by redistribution of time or changing class sizes) [30, 55, 82],
- teaching and learning activities to women-only classes with women faculty [126].

In one of the few interventions specially designed to increase interaction with the instructor, Case et al. [26] describes a one-off, in-class activity where students engaged in mutual in-class interviews with the instructor. The objective was for staff and students to get to know each other and set expectations. The intervention was used on the first day of class by four instructors in four different fields: education, psychology, women's studies, and computer science. Across nine courses, 149 students (73% women) completed surveys about their experience with the activity. No gender differences were found but computer science students tended to rate the activity lower than students in other fields. Students generally reported that the exercise increased their knowledge about the course and the authors suggest it can be an effective starting point for establishing a supportive classroom community.

In an intervention that redesigned the structure of an entire course to increase community and interaction with instructors, Alvarado et al. [4] examine the impact of reorganizing a traditionally large lecture class into small micro-classes of 25-30 students taught by TAs. While the primary intent of the intervention was to build student community, the design also meant that students had closer contact with TAs. The intervention resulted in more female students enrolling in the micro-class and in an improved sense of community for all students compared to those in the traditional lecture class. No differences in academic outcomes were noted. It is not possible to know, however, if the outcomes were due to closer interaction with instructors or to the small class size, or to a combination of those, and possibly other, factors.

Two papers compare online courses with face-to-face courses–designs that can shed some light on the impact of interaction with faculty. Xu and Jin [124] initially set out to incorporate extra interaction with TAs in their pilot study of 4-session Games Development workshops. But due to COVID-19 they pivoted online partway through the sessions. This allowed them, unexpectedly, to capture data comparing face-to-face and online delivery, in addition to TA-designed and delivered game development courses. From an analysis of simple end-of-session surveys (between 22 and 34 respondents per session), the authors reported that female participants had enjoyed the face-to-face peer mentor sessions more than males. In addition, while students on average reported negative impacts of moving online, it was more significant for females [124].

In another comparison study made possible by the COVID pandemic, Irani and Denaro [55] report that when they migrated a large-scale discrete math courses online, they saw no degradation in student academic outcomes and student satisfaction, irrespective of gender or other demographics. However, the online course was carefully migrated and heavily scaffolded, and explicit attention was given to maintaining student-instructor interaction. Interestingly, in migrating online, the instructor reported spending more time answering student questions than in the face-to-face version of the course [55]. However, rather than provide insight into the ways in which interaction with faculty can positively impact students, this

Table 13: Summary of papers identified within the practice "Provide Opportunities for Interaction with Faculty"

Study	Intervention*							Outcomes**		
	Type of Initiative	Who is interacting?						A	C	P
		TA			Ad	F	IT			
		D	T	M						
[26]	First day instructor-student interview					✓		E+?		
[4]	TA Led Micro-classes		✓	✓				E+	E=	F+?
[124]	TA design and deliver the course & face-to-face to online course conversion	✓	✓					F+		
[55]	Face-to-face to online course conversion					✓		E=	E=	
[123]	Living Learning Community		✓	✓	✓	✓	✓	F-		
[101]	Living Learning Community			✓	✓	✓	✓	E+	E+?	E+
[107]	Linked-courses learning community					✓	✓	F+		
[109]	Linked-courses learning community					✓	✓		E-?	E-?
[82]	Cohort-based degree (including PLTL)		✓	✓	✓	✓	✓	E+?	E+?	E+
[67]	Cohort-based degree (including PLTL)		✓	✓				E+?	E+	F+
[30]	Small CS class with a range of interventions		✓	✓		✓		E+?		F+
[126]	Female taught, all women, small class					✓		F+	F+	F+
[104]	STEM Learning Community & Research Experiences		✓	✓		✓		F+?	F+?	E+?
[91]	STEM Learning Community & Research Experiences					✓	✓	E+?	E+?	F+?
[90]	Work placements						✓		E+	
[61]	Industry internships						✓	E+?	E+?	

* Intervention is reported by categories with whom the interaction is with TA = Teaching Assistants, Ad = Admin Staff, F = Faculty (instructors), IT = IT Professionals. Within TA interaction, TAs take on different roles of D = Course Designers, T = Teachers, M = Mentors or Tutors.
** For explanation of other symbols and abbreviations, see see Table 3.

study addresses the ways that instructors may mitigate decreases in interaction that may happen with when courses are put online.

Many of the studies testing the effects of interaction with faculty were part of large scale department-level initiatives such as learning communities or cohort-based degrees (see Table 13). Unfortunately, often insufficient detail is provided about individual component interventions to allow replication, and given the simultaneous implementation of multiple initiatives, it is not possible to evaluate the independent effects of initiatives such as increasing interaction with faculty.

Nine studies incorporated increased interaction with TAs. In this group two small to medium size studies [67, 82] reported using Peer-Led Team Learning (PLTL) [45] and one [67] also used Affinity Research Groups [118]. Another approach was to have micro-classes, small groups in large classes, led by specially trained TAs [4]. Within specialist programmes, such as living learning communities, several studies reported selected peers (TAs) also provide general mentoring and administrative support [101, 123].

A recurrent sub-theme across TA interaction studies was maximising TA effectiveness through training. One introductory CS1 course study described how a designated member of the staff provided weekly training of TA tutors in how to tutor the current topic and trained PLTL TAs who guided first-year students through weekly collaborative problem sets [82]. Monthly PLTL TA training in a different study by Kulkarni et al. [67], describes having the following goals: 1) supporting mentor TAs to assume ownership of their teaching; 2) develop strategies to identify and resolve learning challenges; and 3) co-discover methodologies to resolve student issues. TAs were also encouraged to share their personal experiences and struggles of learning CS with their assigned students. As well as introducing extra TA support in the first year, in the second year of Kulkarni et al.’s study, a group of more experienced peers, senior or graduate students, acted as ARG (Affinity Research Groups) mentors for capstone projects. In employing TAs to support students, a key benefit was reported to mitigate stereotypes [67] (see Section 5.3.4). However, what impact each form of TA training or different intervention is having is not clear; further research is needed.

With regard to administration interaction, a "one-stop-shop" for all administrative and academic queries is described in one initiative to give first-generation college students help navigating their cohort-based CS degree [82]. In other studies, living-learning communities sometimes provide dedicated support staff [123].

Increasing interaction with IT professionals across the reported studies was afforded through a range of interventions from optional employer visits [101, 107, 109, 123], industry networking [101, 123], industry-supported leadership training [101], female industry professionals teaching students [101] to optional [61, 90] or mandatory internships [82]. All but two of these studies were part of large scale initiatives. The impact of engagement with IT professionals in the large scale interventions was not distinctly measured; therefore, it is difficult to conclude about the contribution of this component. Targeted research is needed to investigate the forms of, and impact of, interaction between students and IT professionals.

In summary, few studies have sought to understand the unique impact of interactions with faculty, instructors, and other IT professional on women's sense of belonging, performance, and retention in computing. As such, the role of these interactions is an open question and more targeted research is needed to evaluate instructor-student engagement approaches.

5.3.4 Mitigate Stereotype Threat. Stereotype threat occurs when we fear that our actions will confirm negative stereotypes of an identity group to which we belong. When activated, stereotype threat can negatively affect performance and motivation by reducing feelings of competence, belonging, and trust. It has been suggested that stereotype threat can be mitigated by reframing tasks to remove associations with stereotypes, by giving effective encouragement, and with self-affirmations. Only two papers tested some aspect of stereotype threat mitigation. See Table 14 for a list of the papers.

Table 14: Summary of papers identified within the practice "Mitigate Stereotype Threat"

Paper	Intervention	Outcomes*		
		A	C	P
[28]	Classroom environment	F+		F+
[114]	Game Development	F+		

* For explanation of symbols and abbreviations, see see Table 3.

Cheryan et al.'s [28] study highlights the way that a traditional CS classroom may reproduce stereotypes and intensifies stereotype threat. The authors experimentally examined enrollment intentions, ambient belonging, and expected success in 3D virtual classrooms. A total of 159 undergraduates who were not already majoring in computer science were studied across three experiments. The "stereotypical" classroom included science fiction books, computer parts, electronics, software technology magazines, video games, and computer books. The non-stereotypical classroom included nature and art posters, plants, lamps, general magazines, and water bottles. Women who interacted with the non-stereotypical classroom were more likely to indicate an intention to take CS courses and to anticipate greater success than those who interacted with the stereotypical classroom. The women in the latter condition also reported significantly lower ambient belonging. The different environments did not appear to have significantly different effects on men. The study highlights that the design of learning environments (their layouts and objects) can affect students' identities and sense of belonging.

In a different vein, Sweedyk [114] looked at game development projects in CS1. While games can be powerful tools for engagement, they may have adverse effects on non-gamers and the content of the games may reinforce gender (and other) stereotypes. To this end, the author changed the focus of the games from types that traditionally appeal to men to that of serious games for real clients. For instance, one team of women developed an educational game to teach Greek mythology to middle school students. Survey results indicate that women agreed that serious games made good course projects and were more involved in these than were men.

5.3.5 Build Student Confidence & Professional Identity: Summary. There is limited empirical research testing the practices within this principle, with most of the papers being categorised in the *Provide Opportunities for Interaction with Faculty* practice. Nonetheless, interventions across this principle showed some positive impact. Particularly in regards to the practice *Provide Opportunities for Interaction with Faculty*, the results highlight that increased interaction with faculty, including teaching assistants and IT professionals, appears to positively impact all students, regardless of gender, and particularly, in regards to affective outcomes. However, quite often details about implementation were not offered by the authors, and thus it would be difficult to replicate many of these studies. It should also be noted that some studies were part of multi-faceted interventions, and as such, any reported positive impact can not easily be attributed to these practices alone. Therefore, further research needs to be conducted to investigate the independent impact of these practices on female students' outcomes, or whether these practices act more as supportive mechanisms and should better be regarded in multi-faceted intervention settings.

Regarding the other three practices, *Give Effective Encouragement, Offer Student-centred Assessment*, and *Mitigate Stereotype Threat*, empirical research is scarce. The limited existing research paid particular emphasis on reporting effects on students' self-efficacy and confidence, engagement and interest, performance, identity and sense of belonging, but the results were mixed across all these practices; some papers noted positive impact on a specific population and negative or no impact on others. The results from some papers are not highly generalizable due to small sizes. Overall, computing education research conducted over the last decade provides little conclusive evidence for these practices.

5.4 Potential New Practices

In the process of reviewing the literature, some themes emerged that did not easily fit into the Engagement Practices Framework, or seemed so important that we felt they should not be subsumed under a more general practice. In two instances, we recommend adding a new practice, and in one–lowering risk–the evidence is insufficient to recommend adding a new practice but the relevant papers are discussed here.

5.4.1 Provide Practical Experiences. Several papers in our set had an overarching initiative of a research or practical experience [43, 62, 91, 103, 104]. For example, a study by Kim et al. [62] sought to evaluate the Research Experience for Undergraduate (REU) students. REUs are learning activities that follow a research process, are intended to increase interest in graduate studies, and usually involve solving real-world problems. They surveyed 117 NSF REU program directors, interviewed 20 of those who submitted a survey, and studied the experiences of 96 students who participated in evaluations of the program sites through surveys, focus groups, observations, and journal entries. The assumption is that, given hands-on experiences and collaboration with faculty, graduate students, other undergraduate students, female students in REUs can understand what a computing career can look like and directly experience some of the advantages and rewards of these pathways. Findings indicate the importance of combining multiple principles and practices to produce positive outcomes. Qualitatively, several participants thought that the projects "broke the mold of the 'typical' computer scientist or engineer" by balancing contributions from students from different disciplines.

Of the other research experience papers, two are components within large-scale community programs [91, 104], two are synthesis of courses from many institutions over several years with varying outcomes [62, 103] and one is a small scale pilot [43]. While these differences make it is difficult to generalise, studies of REUs (see Table 15) point to promising increases in affective outcomes and some evidence of cognitive and population increases, either for all students or for females.

Table 15: Summary of papers identified within the practice "Practical Experience"

Paper	Intervention	Outcomes*		
		SE	C	P
[1]	Research experiences and a range of interventions	E+	E=	F+?
[62]	Research Experiences for Undergraduates	E+? F+?	E+?	
[43]	Research Experiences for Undergraduates	F+		F+
[91]	STEM Learning Community and Research Experiences	E+?	E+?	F+?
[104]	STEM Learning Community and Research Experiences	F+?	F+?	E+?
[103]	Research Experiences for Undergraduates	E+ F++	E+	
[90]	Work placement		E+	
[54]	Leaders of Peer-Led Team Learning (PLTL)	E+	E+	

* For explanation of symbols and abbreviations, see see Table 3.

One element of the multi-pronged approach taken at Harvey Mudd described by Alvarado et al. [1] was a summer research experience that requires no background beyond CS1. With these research experiences, the authors aimed to enhance women's experiences with real computing problems and support the growth of confidence in their abilities to contribute meaningfully to these problems. Other interventions (discussed elsewhere in this report) included taking a broader perspective of CS and including a third "Green" section of the course focused on interdisciplinary applications in biology. They also created research projects for students in early stages of their education to apply their computing skills. The report could not attribute the impact of each intervention independently, but the combination of interventions facilitated a dramatic increase in participation among women in CS. The multi-pronged approach taken at Mudd is obviously powerful "on the ground" but poses difficulties in attributing impacts to any one initiatives.

Patel et al. [90] conducted an archival study of the impacts of doing work placements (e.g., an internship or co-op) as part of a degree course on academic outcomes. The authors analyzed the academic records of 290 undergraduate CS students in the UK, comparing those who did a placement with those who did not. They found that more of those completing placements achieved a higher classification of degree (i.e., better performance) than those who did not take a placement. Analyses suggest that the increase was not because students with higher academic performance take placements. Gender was a peripheral question for the authors: rather than considering the potentially differential impacts of placement on women, the authors considered gender to be a confounding factor. While fewer women than men participated in work placement, the difference was not significant and they observed no differences in outcomes for women.

Another way to provide practical experience to undergraduate computing students is to engage them as teaching assistants in introductory courses. Hug et al. [54] reported on a Peer Led Team Learning (PLTL) project where TAs were required to produce feedback and reflective weekly reports. Analysis of survey data indicated that undergraduates who served as leaders in PLTL reported improved content knowledge and improved confidence in computing, irrespective of gender.

5.4.2 Changes to Introductory Curriculum. Several papers described initiatives that were more curricular than pedagogical, particularly around the redesign of introductory course sequences (see Table 16). The existing NCWIT Framework, largely because of its origin as a rubric for EngageCSEdu, does not include anything around these types of initiatives. For example, Alvarado et al. [1] discusses a series of interventions employed at Harvey Mudd College beginning in 2005 that resulted in almost tripling women's representation in the CS major to almost matching their proportion in the overall student body (50%). One important intervention was a set of changes to the introductory course sequence, including creating distinct sections of the course for students with previous programming experience and for those without. This intervention addressed the demoralization that can occur when students with a little experience in computing enter introductory courses with students with ample experience.

Cohoon and Tychonievich [30] implemented a similar intervention in the context of a US research university where they designed and offered a computer science course explicitly for students with no previous programming experience. In essence, the lecture and lab sessions were combined into a single set of instructional periods. The course also included more programming projects and assignments per class period and a smaller cohort of students. Assignments were smaller and more frequent at the start of the course with students engaging in two projects per week by the end of the semester. Simple in-class exercises and homework assignments were also part of the instructional period with teaching assistants providing help to students, as needed. The analysis of their findings demonstrates that in semesters when this course was offered, women and minority students were overrepresented. Additionally, 70% of the students mentioned that the instructor's role was very encouraging and all students who intended to major in computing argued that they felt confident and sure that they would succeed (the results were not gender-specific).

Similarly, Parham-Mocello et al. [89] provided options for students to choose between three variations of CS0 courses: coding first, delayed coding, and coding-free. The paper reported that female students were more interested in the coding-free course. However, the study (n=191) did not indicate whether the population differences in enrollment translated into improved recruitment/retention, nor whether there were differences in cognitive or affective outcomes.

Dickerson [34] details how a private liberal arts college in the US implemented an alternate to Python-based CS1 courses that leveraged the NetLogo programming language to develop multi-agent simulation and agent-based modeling. The study found that the NetLogo course (n=37) had a higher proportion of female enrollment (49%) than the Python courses (31% and 39%). However, students in the NetLogo course were less likely to indicate CS as their intended major and did not perform as well in the subsequent CS2 course, averaging a grade point average of 3.49 compared to 3.67 and 3.70 from the Python courses.

Ying et al. [126] describe an intervention where women were offered enrollment in an alternative, small women-only class in place of the traditional large lecture course (with over 600 students). The curriculum and assessments were the same in both classes and the instructors for the respective courses were both women. Compared to the experiences of women in the large lecture course, students in the women's class reported greater social connections and comfort collaborating with their peers, a more enjoyable and welcoming classroom climate, being more confident in the course material, and being more supported from the women in the class. Moreover, the drop rate was significantly lower in the non-traditional class. The authors suggest that women-only educational interventions can benefit and support women in CS. Unfortunately, it is impossible to disentangle the effect of the women-only environment with the effect of the small, intimate setting.

A few papers fit less well within this new practice category but deserves mention as a potentially effective technique to use in *introductory courses.* Marguliuex et al. [75] describe the effects of including subgoal labels where portions of code within a program are labeled with the function that the code implements. In the large scale study of 256 CS1 students in the US, the authors showed that the group of students, irrespective of gender, who used worked examples with subgoals were more successful in formative assessment tests than students who used worked example without subgoal labels. However, in the final summative assessment no difference in performance were noted. With respect to population effects, students, irrespective of gender, in the subgoal group were less likely to drop out of the course than the control group [75]. Another paper with a promising practice for introductory courses described the effects of changing a CS systems module in which content transitioned from an emphasis on C language mechanics to systems concepts such as basic computer organisation and memory management. The authors report that following the change, females were more likely to major in CS [98]. Finally, through an iterative Problem-Solving Strategy Detection and Development approach, Jin et al. [60] improved the performance and students' motivation and confidence in an introductory programming course which was particularly evident for the female students.

Table 16: Summary of papers identified within the practice "New Course Structures"

Paper	Intervention	Outcomes*		
		SE	C	P
[30]	CS1 course for inexperienced	E+		F+
[1]	Multi-intervention effort	F+		F+
[98]	Content change from programming mechanics to systems concepts			F+
[34]	CS1 course on multi-agent simulation		E-?	F+
[89]	Unplugged and story-based CS0		F+	
[75]	Subgoal labeled worked examples		E+?=	E+
[126]	Small cohort women's class	F+	F+	F+

* For explanation of symbols and abbreviations, see see Table 3.

5.4.3 *Lower the Risk.*

Another theme that emerged from our analysis of the literature is the idea of lowering the risk to engaging with, or trying out, computing as a means of increasing the recruitment and engagement of women. However, the evidence for its effectiveness in increasing women's engagement, performance, and persistence is not strong, but we include a discussion of the papers here since they are part of our sample.

One way to lower risk at the course level, is to provide pass/fail options. Malan [73] presents an alternative grade initiative in which students, instead of a letter grade, receive a grade of Pass/Fail or Satisfactory/Unsatisfactory. The authors report their long term initiative which started in 2010, but it was only when the authors made Satisfactory/Unsatisfactory the course's default in 2017 that the number of students taking the course increased by 31%, with the number of women in the course rising to 44%. Depending on institutional context, this intervention simply may not be practical.

At a more micro level, Butler et al. [20] describe using paper-and-pencil puzzles in Introductory Computer Science (CS1/2) courses over the span of three years with approximately 500 students. The proportion of women students ranged from 11% to 15%. The authors suggest that these types of puzzle assignments are a leveling context for the instruction of CS topics (because no programming is required) and can be effective in teaching students of varying experience levels. In this sense, they represent a kind of lower risk in engaging with computing. However, no gender differences were observed for grades, student perceptions, or student reflective thinking.

Another study introduced "Hour of Code" puzzle-based activities from code.org to undergraduates taking an introduction to computer science course [121]. Wimmer reports that females performed as well as males, except for one code comprehension question where females on average performed better than males. However, the evidence for an impact on women's interest in computing was not supported with males indicating they were more likely to follow the course with further programming activities than were females [121].

6 DISCUSSION

6.1 Assessment of the Research Questions

In this report we have sought to evaluate the empirical evidence for the principles and practices outlined in the NCWIT Engagement Practices Framework, as well as determine if there are additional promising practices. We conducted a systematic review of the intervention-focused research published in the last decade that included assessment of impacts on women. To assess the strength of the evidence in the context of computing, only studies implemented in the context of computing education were assessed. The impacts reported for all students and for women, if available, were reviewed and categorized as affective (e.g., enjoyment, belonging), cognitive (e.g., improved performance in a course), or population-level (e.g., more women taking additional computing courses).

We found general support for the NCWIT Engagement Practices Framework, with some caveats for particular practices, and with the suggestion of adding two additional practices. In this section, we review the general evidence for practices within each principle.

Growing Inclusive Community. Overall, the research supports that this principle is an important one. In addition, of the three principles, this principle directly addresses the instructional environment rather than the qualities or skills of individual students or groups of students, and as such, has more potential for resulting in long term, systemic change. The limitation, however, is that many of the associated practices–and the related research–are designed to impact individual learning; the impacts on student community and culture is considered secondarily, or more problematically, must be inferred. One exception are studies of learning communities. These, however, are often so multifaceted that it is difficult to disentangle independent effects and the research we reviewed showed mixed, and potentially fleeting, effects. In implementing and researching collaborative learning, learning communities, and other practices where students are regularly interacting, more systematic attention needs to be paid to informal social interactions among students, including how student microaggressions, "showboating," and other power plays can negatively impact the experiences of women and other students who are in the minority in computing. In tandem, we need more theoretically informed research that examines *directly* how instructors can create structures and cultural practices that help students learn to work effectively in diverse groups and to combat sexist and racist beliefs and practices.

In addition, there is a great deal of variety in the way all of these practices are implemented and evaluated. Avoiding stereotypes has clear and well supported affective benefits to students, but additional research by the computing education research community in needed to understand the impacts on cognitive and population outcomes. Interventions that promote collaborative learning and student engagement clearly have benefits when properly implemented, but poorly planned implementations can have negative impacts on women (and others) students. Further research is needed that replicates specific interventions so that results can be more directly compared, and for more established interventions, such as pair programming, the community needs to develop, test, and seriously engage with best practices. Clearly establishing best practices for interventions would not only allow for replication and more direct comparison of studies, but also ensure that the negative outcomes are minimized.

Making it Matter. Our review of the papers that employed practices associated with "making it matter" indicate that they can result in significant improvements for women students in terms of changing attitudes, improving grades, and increasing their representation. Much work remains, however, to understand more clearly which interventions are most effective and under what circumstances, and how to avoid potential negative impacts. Interventions which intuitively seem very beneficial for women students do not always have the expected impact, and interventions do not always appear to have the same impact when done in slightly different circumstances. But the crux of it may lie here: few studies attempted to empirically determine the interests of particular students before an intervention was designed. That is, interventions were created based on assumptions instructors made about what might engage students, e.g., robotics or gamification, with little attention paid to whether certain groups of students (e.g., women) might actually be interested in them. One exception is the study by Xu and Jin [124] where undergraduate TAs co-designed the intervention. But we urge caution: are the TAs representative of the kinds of students you are seeking to engage? Students who are currently succeeding in computing–in the current environment and with the current pedagogy and curriculum–may not have similar interests to those who have not (yet) chosen computing. The latter might just thrive if the environment was also designed for them in mind. Cohoon and Tychonievich [30] are a model, however, as they have developed a method for gathering students interests directly.

Building Confidence and Professional Identity. Increased interaction with faculty including TAs and IT professionals appears to positively benefit all students, irrespective of gender, and particularly with respect to affective outcomes. But targeted research is needed to investigate the effectiveness of specific interventions. There were few studies in our sample that tested the effectiveness of the other practices under this principle. Some of this may be a function of our search methodology and focus. There is, for example, a large body of research on growth mindset, but to date few studies looking specifically at its impact in computing. Likewise, stereotype threat and its mitigation is a well-researched area in social psychology but we found few studies looking at it in the context of computing.

6.2 Limitations of Our Approach

As an ITiCSE Working Group project, we are necessarily operating under time and resource constraints. This required limiting our focus and this means that the conclusions we can draw have associated limitations. First, we cannot provide evidence for the effectiveness of these practices beyond the English-speaking world; in fact, many of the studies were conducted in North America so even application to educational contexts in, e.g., Great Britain, may be limited. Second, some of the practices, such as mitigating stereotype threat and giving effective encouragement (such as growth mindset), are general educational practices related to promoting inclusion and encouraging persistence. Because of our inclusion criteria focused on identifiable interventions within computing contexts, we missed a great deal of research supporting these practices. While research within the computing context would be useful, we may not need it in order to recommend these practices. Third, our choice to look only at research with specified interventions means we excluded other research that could have shed light on topics such as how negative stereotypes impact students' perceptions of what computing is and who does it, or the impact of student interaction, culture, and community. Lastly, because we privileged the classroom context we potentially missed research on other forms of student interaction that might be facilitated outside a formal classroom but may be equally important (e.g., we likely missed much of the research on women's groups).

In terms of our original goal of assessing the evidence for teaching practices that broaden participation, the narrowing of our focus means that we cannot speak to how the effectiveness of these practices may vary across ethnic or racial groups, for first generation students, or for students with disabilities. And because few papers discussed their data intersectionally (beyond looking for interaction effects), we are unable to assess how these practices might differentially impact *women from various ethnic and racial groups*.

Finally, we suggest that the three principles of the Engagement Framework are useful quasi-theoretical ways of structuring our thinking about the mechanisms by which the practices instructors employ work to increase and improve women's participation in computing. That said, we acknowledge that our use of the Framework as an organizational tool for our work may have blinded us to seeing other ways ("principles") that instructor practices may positively impact women's participation and experience of computing.

7 RECOMMENDATIONS FOR INSTRUCTORS

Based on this review of empirical studies over the last decade, we recommend that instructors:

(1) Assess all course materials, webpages, and other student-facing communication for stereotypes; stereotypes hurt in multiple ways and can readily be corrected.
(2) Try collaborative learning, especially peer instruction, but provide structure and training on how to be a good group member; be attentive to student interaction, in particular, be on the look-out for microaggressions and biased or exclusionary behavior.
(3) Make connections from computing to your students' lives and interests (Make it Matter) but don't assume you know what those interests are; find out!
(4) Having meaningful, positive interactions with faculty, teaching assistants, and other computing professionals can powerfully impact students. Find ways to intentionally bake these interactions into your initiatives. But make sure these individuals are educated about the ways that biases and stereotypes can come into play in these interactions and how they can be mitigated.

8 FUTURE WORK

8.1 More Sophisticated Empirical & Theoretical Approaches are Needed

Intersectional Analyses Are Needed. Few to no studies used an intersectional approach in analyzing potential gender differences (e.g., Black women, Asian men, first generation women vs. men). Sexism and gender stereotypes do not always function in the exact same way for women from different racial/ethnic and class backgrounds. We should then not assume that interventions to increase women's participation in computing would work in the same way for different women. Granted, the lack of intersectional analyses may be due to small numbers within subgroups, but we should, at the very least, report the racial/ethnic backgrounds of the men and women who are being compared so that context is explicit. Too often White and Asian women are standing in for "all women" and White and Asian men for "all men."

Built-in Sampling Limitations. The power of studying practices in real educational settings is obvious but it comes with a built-in limitation when the goal is diversifying your student body. If there are already few women (or, e.g., Black men, etc.) in a course, the sample size for comparisons will be small which limits the ability to detect differences. Perhaps more problematic is that we are studying the minority of women who already are choosing computing. We do not know if they are representative of the greater population of women who might be successful in computing if they were encouraged, recruited, or given a chance. A few studies did look at students who were not (yet) computing students and tested directly the effect of interventions on their interest in computing (e.g., Cheryan et al. [28]). We are aware that much of the current research in this area is done with students prior to attending university, however, we need more of this kind of research at all levels.

In addition, we need many more replication studies that clearly define the intervention, including details about implementation, and implement across a variety of educational settings. While there are notable exceptions, most of the work we reviewed evaluated interventions implemented at only one institution and many did not have the detail necessary to implement an effective replication of the study. Even in cases where the intervention is fairly well-specified and well-known, such as peer instruction and pair programming, lack of detail on the exact implementation limits comparisons across studies. We suggest that the traditional practice within computing disciplines of publishing research in very short conference proceedings may be part of the problem. While it may be appropriate to keep papers to six pages or less for technical topics, this short format does not provide the space to fully specify most learning and social science research projects. (Note that most social science research is published in journals, not proceedings, with a standard length of 20 pages.)

Theory Helps. Perhaps one of the most oft repeated frustrations from the group as we reviewed the literature is the difficulty of disentangling effects. Since many papers described and assessed multiple interventions at the same time, attributing outcomes to a particular practices is difficult, if not impossible. We harken back to Guzdial's admonition in 2013 for computing education research to move away from its habit of doing "Marco Polo" style research, i.e., "the researchers tried something and reported what happened"

[47] and to move toward more systematic, hypothesis- and theory-driven research. In addition, while there is value in implementing in authentic educational environments, we may need more controlled experimental studies to disentangle effects.

Similarly, research that makes gender comparisons will be improved if it draws more explicitly on existing social science and learning theories; for example, how professions become gendered, and how that gendering is reproduced; how microaggressions and implicit bias work to depress women's engagement, confidence, and persistence; and how particular social supports can mitigate against sexism. More specifically, the field needs to stop the all too common practice of conflating sex with gender. The vast majority of studies used the terms male and female (terms related to biological sex) when they were really looking at gender (where "men" and "women" are the correct terms).

8.2 Expanding the Scope

We would encourage future work using a technique similar to the one used for this report. This work might focus on demographics other than gender (or be explicitly intersectional with gender), expand the scope beyond English-speaking countries, focus more intentionally beyond introductory courses, expand beyond identifiable interventions to include qualitative and survey studies of student experiences, or assess the framework for application to the primary or K-12 context.

The majority of the interventions were carried out in universities in the United States and it is possible that because of specific factors in that environment (e.g., academic pathways, the existence of minors) these practices will not translate into universities in other countries. Clearly, more research is needed. The field also needs more research that tests interventions in more than one environment, be it macro-level differences such as cultural or national context, meso-level differences such as institutional type, or micro-level such as the type of course.

REFERENCES

[1] Christine Alvarado, Zachary Dodds, and Ran Libeskind-Hadas. 2012. Increasing women's participation in computing at Harvey Mudd College. *ACM Inroads* 3, 4 (Dec. 2012), 55–64. https://doi.org/10.1145/2381083.2381100

[2] Christine Alvarado and Eugene Judson. 2014. Using targeted conferences to recruit women into computer science. *Commun. ACM* 57, 3 (March 2014), 70–77. https://doi.org/10.1145/2500883

[3] Christine Alvarado, Cynthia Bailey Lee, and Gary Gillespie. 2014. New CS1 pedagogies and curriculum, the same success factors?. In *Proceedings of the 45th ACM technical symposium on Computer science education (SIGCSE '14)*. Association for Computing Machinery, New York, NY, USA, 379–384. https://doi.org/10.1145/2538862.2538897

[4] Christine Alvarado, Mia Minnes, and Leo Porter. 2017. Micro-Classes: A Structure for Improving Student Experience in Large Classes. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education (SIGCSE '17)*. Association for Computing Machinery, New York, NY, USA, 21–26. https://doi.org/10.1145/3017680.3017727

[5] William Aspray. 2010. Participation in computing. *The National Science Foundation's Expansionary Programs. Springer International Publishing, Switzerland* (2010).

[6] Australian Government, Department of Industry, Science, Energy and Resources. 2019. University enrolment and completion in STEM and other fields. https://www.industry.gov.au/data-and-publications/stem-equity-monitor/university-enrolment-and-completion-in-stem-and-other-fields#university-enrolment-and-completion

[7] Australian Government, Department of Industry, Science, Energy and Resources. 2020. STEM-qualified occupations. https://www.industry.gov.au/data-and-publications/stem-equity-monitor/stem-qualified-occupations

[8] Lecia Barker, Christopher Lynnly Hovey, and Leisa D Thompson. 2014. Results of a large-scale, multi-institutional study of undergraduate retention in computing. In *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*. IEEE, New York, 1–8. https://doi.org/10.1109/FIE.2014.7044267

[9] Lecia J. Barker and William Aspray. 2008. *The state of research on girls and IT*. MIT Press. 3–54 pages.

[10] Lecia J Barker, Charlie Mcdowell, and Kimberly Kalahar. 2009. Exploring Factors that Influence Computer Science Introductory Course Students to Persist in the Major. *ACM SIGCSE Bulletin* 41, 2 (2009), 282–286.

[11] Valerie Barr. 2016. Disciplinary Thinking, Computational Doing: Promoting Interdisciplinary Computing While Transforming Computer Science Enrollments. (2016), 10.

[12] Valerie Barr. 2018. Different denominators, different results: reanalyzing CS degrees by gender, race, and ethnicity. *ACM Inroads* 9, 3 (2018), 40–47.

[13] Debarati Basu, Harinni K. Kumar, Vinod K. Lohani, N. Dwight Barnette, Godmar Back, Dave McPherson, Calvin J. Ribbens, and Paul E. Plassmann. 2020. Integration and Evaluation of Spiral Theory based Cybersecurity Modules into core Computer Science and Engineering Courses. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20)*. Association for Computing Machinery, New York, NY, USA, 9–15. https://doi.org/10.1145/3328778.3366798

[14] Kara Alexandra Behnke, Brittany Ann Kos, and John K. Bennett. 2016. Computer Science Principles: Impacting Student Motivation & Learning Within and Beyond the Classroom. In *Proceedings of the 2016 ACM Conference on International Computing Education Research (ICER '16)*. Association for Computing Machinery, New York, NY, USA, 171–180. https://doi.org/10.1145/2960310.2960336

[15] Michael David Black. 2016. Seven Semesters of Android Game Programming in CS2. In *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education*. ACM, Arequipa Peru, 5–10. https://doi.org/10.1145/2899415.2899470

[16] Kamau Bobb. 2016. Broadening Participation in Computing: A Critical Perspective. *ACM Inroads* 7, 4 (2016), 49–51.

[17] Nicholas A. Bowman, Lindsay Jarratt, KC Culver, and Alberto M. Segre. 2021. The Impact of Pair Programming on College Students’ Interest, Perceptions, and Achievement in Computer Science. *ACM Transactions on Computing Education* 21, 3 (May 2021), 19:1–19:19. https://doi.org/10.1145/3440759

[18] Bo Brinkman and Amanda Diekman. 2016. Applying the Communal Goal Congruity Perspective to Enhance Diversity and Inclusion in Undergraduate Computing Degrees. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education (SIGCSE '16)*. Association for Computing Machinery, New York, NY, USA, 102–107. https://doi.org/10.1145/2839509.2844562

[19] Jeni L. Burnette, Crystal L. Hoyt, V. Michelle Russell, Barry Lawson, Carol S. Dweck, and Eli Finkel. 2020. A Growth Mind-Set Intervention Improves Interest but Not Academic Performance in the Field of Computer Science. *Social Psychological and Personality Science* 11, 1 (Jan. 2020), 107–116. https://doi.org/10.1177/1948550619841631

[20] Zack Butler, Ivona Bezakova, and Kimberly Fluet. 2017. Pencil Puzzles for Introductory Computer Science: an Experience- and Gender-Neutral Context. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. ACM, Seattle Washington USA, 93–98. https://doi.org/10.1145/3017680.3017765

[21] E. Rebecca Caldwell and Elva J. Jones. 2013. Beyond wrestling: using sumobots to engage students in the computer science classroom. *Journal of Computing Sciences in Colleges* 29, 2 (Dec. 2013), 132–139.

[22] Tracy Camp, Wendy DuBow, Diane Levitt, Linda J Sax, Valerie Taylor, and Colleen Lewis. 2019. The new NSF requirement for broadening participation in computing (BPC) plans: Community advice and resources. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. 332–333.

[23] Jennifer Campbell, Diane Horton, Michelle Craig, and Paul Gries. 2014. Evaluating an inverted CS1. In *Proceedings of the 45th ACM technical symposium on Computer science education - SIGCSE '14*. ACM Press, Atlanta, Georgia, USA, 307–312. https://doi.org/10.1145/2538862.2538943

[24] Canadian Government, Statistics Canada. 2019. Persistence and representation of women in STEM programs. https://www150.statcan.gc.ca/n1/pub/75-006-x/2019001/article/00006-eng.htm

[25] Yingjun Cao and Leo Porter. 2017. Evaluating Student Learning from Collaborative Group Tests in Introductory Computing. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education (SIGCSE '17)*. Association for Computing Machinery, New York, NY, USA, 99–104. https://doi.org/10.1145/3017680.3017729

[26] Kim Case, Robert Bartsch, Lillian McEnery, Sharon Hall, Anthony Hermann, and David Foster. 2008. Establishing a Comfortable Classroom from Day One: Student Perceptions of the Reciprocal Interview. *College Teaching* 56, 4 (Sept. 2008), 210–214. https://doi.org/10.3200/CTCH.56.4.210-214

[27] Sapna Cheryan, Benjamin J. Drury, and Marissa Vichayapai. 2013. Enduring Influence of Stereotypical Computer Science Role Models on Women's Academic Aspirations. *Psychology of Women Quarterly* 37, 1 (March 2013), 72–79. https://doi.org/10.1177/0361684312459328

[28] Sapna Cheryan, Andrew N. Meltzoff, and Saenam Kim. 2011. Classrooms matter: The design of virtual classrooms influences gender disparities in computer science classes. *Computers & Education* 57, 2 (Sept. 2011), 1825–1835. https://doi.org/10.1016/j.compedu.2011.02.004

[29] Sapna Cheryan, John Oliver Siy, Marissa Vichayapai, Benjamin J. Drury, and Saenam Kim. 2011. Do Female and Male Role Models Who Embody STEM Stereotypes Hinder Women's Anticipated Success in STEM? *Social Psychological and Personality Science* 2, 6 (Nov. 2011), 656–664. https://doi.org/10.1177/1948550611405218

[30] James P. Cohoon and Luther A. Tychonievich. 2011. Analysis of a CS1 approach for attracting diverse and inexperienced students to computing majors. In *Proceedings of the 42nd ACM technical symposium on Computer science education (SIGCSE '11)*. Association for Computing Machinery, New York, NY, USA, 165–170. https://doi.org/10.1145/1953163.1953217

[31] Ralph W. Crosby, Stephanie Valentine, and Tiffani L. Williams. 2014. Leveraging programming difficulty to improve understanding and perceptions of non-majors. *Journal of Computing Sciences in Colleges* 29, 4 (April 2014), 27–35.

[32] Jan Cuny. 2010. Changing what's happening in HS and introductory college CS: keynote address. *Journal of Computing Sciences in Colleges* 26, 2 (2010), 5–6.

[33] Linda Darling-Hammond, Lisa Flook, Channa Cook-Harvey, Brigid Barron, and David Osher. 2020. Implications for educational practice of the science of learning and development. *Applied Developmental Science* 24, 2 (2020), 97–140. https://doi.org/10.1080/10888691.2018.1537791

[34] Matthew Dickerson. 2014. Multi-Agent Simulation, NETLOGO, and the Recruitment of Computer Science Majors. *Computing Sciences in Colleges* (2014), 9.

[35] Betsy DiSalvo, Mark Guzdial, Amy Bruckman, and Tom McKlin. 2014. Saving Face While Geeking Out: Video Game Testing as a Justification for Learning Computer Science. *Journal of the Learning Sciences* 23, 3 (2014), 272–315. https://doi.org/10.1080/10508406.2014.893434 arXiv:https://doi.org/10.1080/10508406.2014.893434

[36] Wendy DuBow and Jax J. Gonzalez. 2020. NCWIT Scorecard: The Status of Women in Technology. Boulder, CO: NCWIT.

[37] Wendy Dubow, Alexis Kaminsky, and Joanna Weidler-Lewis. 2017. Multiple factors Converge to Influence Women's Persistence in Computing: A Qualitative Analysis. *Computing in Science & Engineering* (5 2017), 30–39. www.computer.org/cise

[38] Wendy M DuBow, Beth A Quinn, Gloria Childress Townsend, Rosario Robinson, and Valerie Barr. 2016. Efforts to make computer science more inclusive of women. *ACM Inroads* 7, 4 (2016), 74–80.

[39] Vincent Duriau, Rhonda Reger, and Michael Pfarrer. 2007. A Content Analysis of the Content Analysis Literature in Organization Studies: Research Themes, Data Sources, and Methodological Refinements. *Organizational Research Methods* 10 (01 2007), 5–34. https://doi.org/10.1177/1094428106289252

[40] Karen Kangas Dwyer, Shereen G. Bingham, Robert E. Carlson, Marshall Prisbell, Ana M. Cruz, and Dennis A. Fus. 2004. Communication and connectedness in the classroom: Development of the connected classroom climate inventory. *Communication Research Reports* 21, 3 (2004), 264–272. https://doi.org/10.1080/08824090409359988

[41] Nathan Ensmenger. 2010. *Making programming masculine.* IEEE-CS/Wiley. 115–141 pages.

[42] European Commission. 2020. Women in Digital. https://ec.europa.eu/digital-single-market/en/women-digital-0

[43] Ann Quiroz Gates, Sarah Hug, Heather Thiry, Richard Aló, Mohsen Beheshti, John Fernandez, Nestor Rodriguez, and Malek Adjouadi. 2011. The Computing Alliance of Hispanic-Serving Institutions: Supporting Hispanics at Critical Transition Points. *ACM Transactions on Computing Education* 11, 3 (Oct. 2011), 16:1–16:21. https://doi.org/10.1145/2037276.2037280

[44] Danielle Geerling, Jonathan Butner, Tamra Fraughton, Sungchoon Sinclair, Joseph Zachary, and Carol Sansone. 2020. The Dynamic Association of Interest and Confusion: The Potential for Moderation by Utility Value and Gender. *The Journal of Experimental Education* 88, 3 (May 2020), 407–430. https://doi.org/10.1080/00220973.2018.1561403

[45] David K Gosser. 2001. *Peer-led team learning: A guidebook.* Prentice Hall Upper Saddle River, NJ.

[46] Maria J Grant and Andrew Booth. 2009. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health information & libraries journal* 26, 2 (2009), 91–108.

[47] Mark Guzdial. 2013. Exploring hypotheses about media computation. In *Proceedings of the ninth annual international ACM conference on International computing education research (ICER '13)*. Association for Computing Machinery, New York, NY, USA, 19–26. https://doi.org/10.1145/2493394.2493397

[48] Hussam A Hallak, Soad Ibrahim, Christy Low, and Ayman El Mesalami. 2019. The Impact of Incorporating Hands-on Raspberry Pi Projects with Undergraduate Education in Boosting Students' Interest in Scientific/Engineering Majors and Encouraging Women and Minorities to Advance their Integration in Practical Fields. (2019), 8.

[49] Brian Harrington, Jingyiran Li, Mohamed Moustafa, Marzieh Ahmadzadeh, and Nick Cheng. 2019. On the Effect of Question Ordering on Performance and Confidence in Computer Science Examinations. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. ACM, Minneapolis MN USA, 620–626. https://doi.org/10.1145/3287324.3287398

[50] J. A. C. Hattie. 2003. Teachers make a difference: What is the research evidence?. In *Building Teacher Quality: What does the research tell us ACER Research Conference*. Melbourne, Australia. http://research.acer.edu.au/research_conference_2003/4/

[51] Sylvia Ann Hewlett, Carolyn Buck Luce, Lisa J. Servon, Laura Sherbin, Peggy Shiller, Eytan Sosnovich, and Karen Sumberg. 2008. The Athena Factor: Reversing the Brain Drain in Science, Engineering, and Technology. , 100 pages. https://www.researchgate.net/publication/268325574

[52] Higher Education Statistics Agency. 2019. Higher Education Student Statistics: UK, 2017/18 - Student numbers and characteristics. https://www.hesa.ac.uk/news/17-01-2019/sb252-higher-education-student-statistics/numbers

[53] Bryan Hosack, Billy Lim, and W Paul Vogt. 2012. Increasing Student Performance Through the Use of Web Services in Introductory Programming Classrooms: Results from a Series of Quasi-Experiments. *Journal of Information Systems Education* 23 (2012), 12.

[54] Sarah Hug, Heather Thiry, and Phyllis Tedford. 2011. Learning to love computer science: peer leaders gain teaching skill, communicative ability and content knowledge in the CS classroom. In *Proceedings of the 42nd ACM technical symposium on Computer science education (SIGCSE '11)*. Association for Computing Machinery, New York, NY, USA, 201–206. https://doi.org/10.1145/1953163.1953225

[55] Sandy Irani and Kameryn Denaro. 2020. Incorporating Active Learning Strategies and Instructor Presence into an Online Discrete Mathematics Class. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20)*. Association for Computing Machinery, New York, NY, USA, 1186–1192. https://doi.org/10.1145/3328778.3366904

[56] Leslie Irvine and Jenny R. Vermilya. 2010. Gender work in a feminized profession: The case of veterinary medicine. *Gender and Society* 24 (2 2010), 56–82. Issue 1. https://doi.org/10.1177/0891243209355978

[57] Alan C. Jamieson, Lindsay H. Jamieson, and Angela C. Johnson. 2012. Application of non-programming focused treisman-style workshops in introductory computer science. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education (SIGCSE '12)*. Association for Computing Machinery, New York, NY, USA, 271–276. https://doi.org/10.1145/2157136.2157219

[58] David S. Janzen, Sara Bahrami, Bruno C. da Silva, and Davide Falessi. 2018. A Reflection on Diversity and Inclusivity Efforts in a Software Engineering Program. In *2018 IEEE Frontiers in Education Conference (FIE)*. IEEE, San Jose, CA, USA, 1–9. https://doi.org/10.1109/FIE.2018.8658677

[59] Lindsay Jarratt, Nicholas A. Bowman, K.C. Culver, and Alberto Maria Segre. 2019. A Large-Scale Experimental Study of Gender and Pair Composition in Pair Programming. In *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '19)*. Association for Computing Machinery, New York, NY, USA, 176–181. https://doi.org/10.1145/3304221.3319782

[60] Wei Jin, Cynthia L. Johnson, and Sonal Dekhane. 2020. A Guided Inquiry Approach for Detecting and Developing Problem-solving Strategies for Novice Programming Students. In *Proceedings of the 2020 ACM Southeast Conference*. ACM, Tampa FL USA, 211–217. https://doi.org/10.1145/3374135.3385289

[61] Amanpreet Kapoor and Christina Gardner-McCune. 2020. Exploring the Participation of CS Undergraduate Students in Industry Internships. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. ACM, Portland OR USA, 1103–1109. https://doi.org/10.1145/3328778.3366844

[62] Karen A. Kim, Amy J. Fann, and Kimberly O. Misa-Escalante. 2011. Engaging Women in Computer Science and Engineering: Promising Practices for Promoting Gender Equity in Undergraduate Research Experiences. *ACM Transactions on Computing Education* 11, 2 (July 2011), 8:1–8:19. https://doi.org/10.1145/1993069.1993072

[63] Maria Klawe. 2013. Increasing Female Participation in Computing: The Harvey Mudd College Story. *Computer* 46, 3 (March 2013), 56–58. https://doi.org/10.1109/MC.2013.4 Conference Name: Computer.

[64] Alyona Koulanova, Ary Maharaj, Brian Harrington, and Jessica Dere. 2018. Fit-breaks: incorporating physical activity breaks in introductory CS lectures. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE 2018)*. Association for Computing Machinery, New York, NY, USA, 260–265. https://doi.org/10.1145/3197091.3197115

[65] Julie Krause, Irene Polycarpou, and Keith Hellman. 2012. Exploring formal learning groups and their impact on recruitment of women in undergraduate CS. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education (SIGCSE '12)*. Association for Computing Machinery, New York, NY, USA, 179–184. https://doi.org/10.1145/2157136.2157192

[66] Sophia Krause-Levy, Leo Porter, Beth Simon, and Christine Alvarado. 2020. Investigating the Impact of Employing Multiple Interventions in a CS1 Course. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20)*. Association for Computing Machinery, New York, NY, USA, 1082–1088. https://doi.org/10.1145/3328778.3366866

[67] Anagha Kulkarni, Ilmi Yoon, Pleuni S. Pennings, Kazunori Okada, and Carmen Domingo. 2018. Promoting diversity in computing. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*. ACM, Larnaca Cyprus, 236–241. https://doi.org/10.1145/3197091.3197145

[68] Celine Latulipe, N. Bruce Long, and Carlos E. Seminario. 2015. Structuring Flipped Classes with Lightweight Teams and Gamification. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education (SIGCSE '15)*. Association for Computing Machinery, New York, NY, USA, 392–397. https://doi.org/10.1145/2676723.2677240

[69] Celine Latulipe, Audrey Rorrer, and Bruce Long. 2018. Longitudinal Data on Flipped Class Effects on Performance in CS1 and Retention after CS1. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE '18)*. Association for Computing Machinery, New York, NY, USA, 411–416. https://doi.org/10.1145/3159450.3159518

[70] Elynn Lee, Victoria Shan, Bradley Beth, and Calvin Lin. 2014. A structured approach to teaching recursion using cargo-bot. In *Proceedings of the tenth annual conference on International computing education research (ICER '14)*. Association for Computing Machinery, New York, NY, USA, 59–66. https://doi.org/10.1145/2632320.2632356

[71] Colleen M. Lewis, Nathaniel Titterton, and Michael Clancy. 2012. Using collaboration to overcome disparities in Java experience. In *Proceedings of the ninth annual international conference on International computing education research (ICER '12)*. Association for Computing Machinery, New York, NY, USA, 79–86. https://doi.org/10.1145/2361276.2361292

[72] Phil Maguire, Rebecca Maguire, Philip Hyland, and Patrick Marshall. 2014. Enhancing Collaborative Learning Using Pair Programming: Who Benefits? *AISHE-J: The All Ireland Journal of Teaching and Learning in Higher Education* 6, 2 (2014), 25.

[73] David J. Malan. 2021. Toward an Ungraded CS50. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. ACM, Virtual Event USA, 1076–1081. https://doi.org/10.1145/3408877.3432461

[74] Jane Margolis and Allan Fisher. 2002. *Unlocking the clubhouse: Women in computing*. MIT press.

[75] Lauren E. Margulieux, Briana B. Morrison, and Adrienne Decker. 2020. Reducing withdrawal and failure rates in introductory programming with subgoal labeled worked examples. *International Journal of STEM Education* 7, 1 (Dec. 2020), 19. https://doi.org/10.1186/s40594-020-00222-7

[76] Patricia Yancey Martin. 2003. "Said and done" versus "saying and doing": Gendering practices, practicing gender at work. (2003). https://doi.org/10.1177/0891243203251716

[77] Monica M. McGill. 2012. Learning to Program with Personal Robots: Influences on Student Motivation. *ACM Transactions on Computing Education* 12, 1 (March 2012), 4:1–4:32. https://doi.org/10.1145/2133797.2133801

[78] Paola Medel and Vahab Pournaghshband. 2017. Eliminating Gender Bias in Computer Science Education Materials. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education (SIGCSE '17)*. Association for Computing Machinery, New York, NY, USA, 411–416. https://doi.org/10.1145/3017680.3017794

[79] Danaë Metaxa-Kakavouli, Kelly Wang, James A. Landay, and Jeff Hancock. 2018. Gender-Inclusive Design: Sense of Belonging and Bias in Web Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3173574.3174188

[80] Alvaro E Monge, Cameron L Fadjo, Beth A Quinn, and Lecia J Barker. 2015. EngageCSEdu: engaging and retaining CS1 and CS2 students. *ACM Inroads* 6, 1 (2015), 6–11.

[81] Myob.com. 2020. Women in Tech Report. https://www.myob.com/content/dam/public-website/docs/misc/myob-women-in-tech-report-2019.pdf?_ga=2.94641819.683107675.1624717016-243744843.1624717016

[82] Sathya Narayanan, Kathryn Cunningham, Sonia Arteaga, William J. Welch, Leslie Maxwell, Zechariah Chawinga, and Bude Su. 2018. Upward Mobility for Underrepresented Students: A Model for a Cohort-Based Bachelor's Degree in Computer Science. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE '18)*. Association for Computing Machinery, New York, NY, USA, 705–710. https://doi.org/10.1145/3159450.3159551

[83] National Center for Women & Information Technology. 2020. Engagement Practices Framework. https://www.ncwit.org/engagement-practices-framework

[84] Tia Newhall, Lisa Meeden, Andrew Danner, Ameet Soni, Frances Ruiz, and Richard Wicentowski. 2014. A support program for introductory CS courses that improves student performance and retains students from underrepresented groups. In *Proceedings of the 45th ACM technical symposium on Computer science education (SIGCSE '14)*. Association for Computing Machinery, New York, NY, USA, 433–438. https://doi.org/10.1145/2538862.2538923

[85] Jacqueline Ng Lane, Bruce Ankenman, and Seyed Iravani. 2018. Insight into Gender Differences in Higher Education: Evidence from Peer Reviews in an Introductory STEM Course. *Service Science* 10, 4 (Dec. 2018), 442–456. https://doi.org/10.1287/serv.2018.0224

[86] Keith J. O'Hara, Kathleen Burke, Diana Ruggiero, and Sven Anderson. 2017. Linking Language & Thinking with Code: Computing within a Writing-Intensive Introduction to the Liberal Arts. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education*. ACM, Bologna Italy, 269–274. https://doi.org/10.1145/3059009.3059018

[87] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, Roger Chou, Julie Glanville, Jeremy M Grimshaw, Asbjørn Hróbjartsson, Manoj M Lalu, Tianjing Li, Elizabeth W Loder, Evan Mayo-Wilson, Steve McDonald, Luke A McGuinness, Lesley A Stewart, James Thomas, Andrea C Tricco, Vivian A Welch, Penny Whiting, and David Moher. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* (March 2021), n71. https://doi.org/10.1136/bmj.n71

[88] Jennifer Parham-Mocello, Shannon Ernst, Martin Erwig, Lily Shellhammer, and Emily Dominguez. 2019. Story Programming: Explaining Computer Science Before Coding. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19)*. Association for Computing Machinery, New York, NY, USA, 379–385. https://doi.org/10.1145/3287324.3287397

[89] Jennifer Parham-Mocello, Martin Erwig, and Emily Dominguez. 2019. To Code or Not to Code? Programming in Introductory CS Courses. In *2019 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, Memphis, TN, USA, 187–191. https://doi.org/10.1109/VLHCC.2019.8818909

[90] Nayna Patel, Willem-Paul Brinkman, and Jane Coughlan. 2012. Work placements and academic achievement: undergraduate computing students. *Education & Training* 54, 6 (2012), 523–533. https://doi.org/10.1108/00400911211254299

[91] Jon K. Piper and Dwight Krehbiel. 2015. Increasing STEM Enrollment Using Targeted Scholarships and an Interdisciplinary Seminar for First- and Second-Year College Students. *Journal of STEM Education : Innovations and Research* 16, 4 (Dec. 2015), 36–43. http://www.proquest.com/docview/1767144450/abstract/5BB8B973D9046FEPQ/1

[92] Leo Porter and Beth Simon. 2013. Retaining nearly one-third more majors with a trio of instructional best practices in CS1. In *Proceeding of the 44th ACM technical symposium on Computer science education (SIGCSE '13)*. Association for Computing Machinery, New York, NY, USA, 165–170. https://doi.org/10.1145/2445196.2445248

[93] Cyndi Rader, Doug Hakkarinen, Barbara M. Moskal, and Keith Hellman. 2011. Exploring the appeal of socially relevant computing: are students interested in socially relevant problems?. In *Proceedings of the 42nd ACM technical symposium on Computer science education (SIGCSE '11)*. Association for Computing Machinery, New York, NY, USA, 423–428. https://doi.org/10.1145/1953163.1953288

[94] National Public Radio. 2014. Episode 576: When Women Stopped Coding. https://www.npr.org/sections/money/2014/10/17/356944145/episode-576-when-women-stopped-coding

[95] Sagar Raina, Blair Taylor, and Siddharth Kaza. 2015. Security Injections 2.0: Increasing Engagement and Faculty Adoption Using Enhanced Secure Coding Modules for Lower-Level Programming Courses. In *Information Security Education Across the Curriculum (IFIP Advances in Information and Communication Technology)*, Matt Bishop, Natalia Miloslavskaya, and Marianthi Theocharidou (Eds.). Springer International Publishing, Cham, 64–74. https://doi.org/10.1007/978-3-319-18500-2_6

[96] Nancy Ramsey and Pamela McCorduck. 2005. Where are the women in information technology. *Report of Literature Search and InterviewsPrepared for the National Center for Women & Information Technology* (2005).

[97] Yolanda A. Rankin, Jakita O. Thomas, and India Irish. 2019. Food for Thought: Supporting African American Women's Computational Algorithmic Thinking in an Intro CS Course. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19)*. Association for Computing Machinery, New York, NY, USA, 641–646. https://doi.org/10.1145/3287324.3287484

[98] Katie Redmond, Sarah Evans, and Mehran Sahami. 2013. A large-scale quantitative study of women in computer science at Stanford University. In *Proceeding of the 44th ACM technical symposium on Computer science education (SIGCSE '13)*. Association for Computing Machinery, New York, NY, USA, 439–444. https://doi.org/10.1145/2445196.2445326

[99] Anja Remshagen and Christine Rolka. 2014. Contextualized learning tools: animations and robots. In *Proceedings of the 2014 ACM Southeast Regional Conference (ACM SE '14)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/2638404.2638458

[100] Barbara F. Reskin and Patricia A. Roos. 1990. *Job Queues, Gender Queues Explaining Women's Inroads into Male Occupations*. Temple University Press.

[101] Penny Rheingans, Anne Brodsky, Jill Scheibler, and Anne Spence. 2011. The Role of Majority Groups in Diversity Programs. *ACM Transactions on Computing Education* 11, 2 (July 2011), 11:1–11:15. https://doi.org/10.1145/1993069.1993075

[102] Christine Rolka and Anja Remshagen. 2015. Showing Up is Half the Battle: Assessing Different Contextualized Learning Tools to Increase the Performance in Introductory Computer Science Courses. *Georgia Educational Researcher* 9, 1 (Jan. 2015). https://doi.org/10.20429/ijsotl.2015.090110

[103] Audrey Smith Rorrer, Joseph Allen, and Huifang Zuo. 2018. A National Study of Undergraduate Research Experiences in Computing: Implications for Culturally

Relevant Pedagogy. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. ACM, Baltimore Maryland USA, 604–609. https://doi.org/10.1145/3159450.3159510

[104] Lauren Russell. 2017. Can learning communities boost success of women and minorities in STEM? Evidence from the Massachusetts Institute of Technology. *Economics of Education Review* 61 (Dec. 2017), 98–111. https://doi.org/10.1016/j.econedurev.2017.10.008

[105] Adrian Salguero, Julian McAuley, Beth Simon, and Leo Porter. 2020. A Longitudinal Evaluation of a Best Practices CS1. In *Proceedings of the 2020 ACM Conference on International Computing Education Research (ICER '20)*. Association for Computing Machinery, New York, NY, USA, 182–193. https://doi.org/10.1145/3372782.3406274

[106] Benjamin J. Schreiber and John P. Dougherty. 2017. Assessment of Introducing Algorithms with Video Lectures and Pseudocode Rhymed to a Melody. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education (SIGCSE '17)*. Association for Computing Machinery, New York, NY, USA, 519–524. https://doi.org/10.1145/3017680.3017789

[107] Amber Settle, James Doyle, and Theresa Steinbach. 2017. The Effect of a Computing-focused Linked-courses Learning Community on Minority and Female Students. In *Proceedings of the 18th Annual Conference on Information Technology Education (SIGITE '17)*. Association for Computing Machinery, New York, NY, USA, 153–158. https://doi.org/10.1145/3125659.3125679

[108] Amber Settle, John Lalor, and Theresa Steinbach. 2015. Evaluating a Linked-courses Learning Community for Development Majors. In *Proceedings of the 16th Annual Conference on Information Technology Education (SIGITE '15)*. Association for Computing Machinery, New York, NY, USA, 127–132. https://doi.org/10.1145/2808006.2808031

[109] Amber Settle and Theresa Steinbach. 2018. Retention Rates for the First Three Years of a Linked-courses Learning Community. In *Proceedings of the 19th Annual SIG Conference on Information Technology Education (SIGITE '18)*. Association for Computing Machinery, New York, NY, USA, 166–171. https://doi.org/10.1145/3241815.3241854

[110] Sadia Sharmin, Daniel Zingaro, Lisa Zhang, and Clare Brett. 2019. Impact of Open-Ended Assignments on Student Self-Efficacy in CS1. In *Proceedings of the ACM Conference on Global Computing Education*. ACM, Chengdu,Sichuan China, 215–221. https://doi.org/10.1145/3300115.3309532

[111] Max O. Smith, Andrew Giugliano, and Andrew DeOrio. 2018. Long Term Effects of Pair Programming. *IEEE Transactions on Education* 61, 3 (Aug. 2018), 187–194. https://doi.org/10.1109/TE.2017.2773024

[112] STEM Women. 2021. Percentages of Women in STEM Statistics. https://www.stemwomen.co.uk/blog/2021/01/women-in-stem-percentages-of-women-in-stem-statistics

[113] Jeffrey A. Stone and Tricia K. Clark. 2011. The impact of problem-oriented animated learning modules in a CS1-style course. In *Proceedings of the 42nd ACM technical symposium on Computer science education (SIGCSE '11)*. Association for Computing Machinery, New York, NY, USA, 51–56. https://doi.org/10.1145/1953163.1953182

[114] Elizabeth Sweedyk. 2011. Women build games, seriously. In *Proceedings of the 42nd ACM technical symposium on Computer science education (SIGCSE '11)*. Association for Computing Machinery, New York, NY, USA, 171–176. https://doi.org/10.1145/1953163.1953218

[115] Andrea Tartaro and Haley Cottingham. 2014. A problem-based, survey introduction to computer science for majors and non-majors. *Journal of Computing Sciences in Colleges* 30, 2 (Dec. 2014), 164–170.

[116] Dilan Ustek, Erik Opavsky, Henry M. Walker, and David Cowden. 2014. Course development through student-faculty collaboration: a case study. In *Proceedings of the 2014 conference on Innovation & technology in computer science education - ITiCSE '14*. ACM Press, Uppsala, Sweden, 189–194. https://doi.org/10.1145/2591708.2591723

[117] Roli Varma and Deepak Kapur. 2015. Decoding Femininity in Computer Science in India. *Commun. ACM* 58 (4 2015), 56–62. Issue 5. https://doi.org/10.1145/2663339

[118] Elsa Q Villa, Kerrie Kephart, Ann Q Gates, Heather Thiry, and Sarah Hug. 2013. Affinity research groups in practice: Apprenticing students in research. *Journal of Engineering Education* 102, 3 (2013), 444–466.

[119] Becky Wai-Ling Packard, Jaemarie Solyst, Anisha Pai, and Lu Yu. 2020. Peer-Designed Active Learning Modules as a Strategy to Improve Confidence and Comprehension Within Introductory Computer Science. *Journal of College Science Teaching* 49, 5 (June 2020), 76–83. http://search.ebscohost.com/login.aspx?direct=true&db=eue&AN=143188756&site=ehost-live&scope=site

[120] Dana Linnell Wanzer, Tom McKlin, Jason Freeman, Brian Magerko, and Taneisha Lee. 2020. Promoting intentions to persist in computing: an examination of six years of the EarSketch program. *Computer Science Education* 30, 4 (Oct. 2020), 394–419. https://doi.org/10.1080/08993408.2020.1714313

[121] Hayden Wimmer. 2019. Hour of Code: A Study of Gender Differences in Computing. (2019), 10.

[122] Women in Communication and Technology. 2018. Where are the Women in the Canadian ICT Industry. https://www.wct-fct.com/sites/default/files/wct-upthenumbers-report-2018-final.pdf

[123] Rebecca N. Wright, Sally J. Nadler, Thu D. Nguyen, Cynthia N. Sanchez Gomez, and Heather M. Wright. 2019. Living-Learning Community for Women in Computer Science at Rutgers. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19)*. Association for Computing Machinery, New York, NY, USA, 286–292. https://doi.org/10.1145/3287324.3287449

[124] Xin Xu and Wei Jin. 2021. Game development workshops designed and delivered by peer mentors to increase student curiosity and interest in an introductory programming course. In *Proceedings of the 2021 ACM Southeast Conference (ACM SE '21)*. Association for Computing Machinery, New York, NY, USA, 87–92. https://doi.org/10.1145/3409334.3452046

[125] Kimberly Michelle Ying, Lydia G. Pezzullo, Mohona Ahmed, Kassandra Crompton, Jeremiah Blanchard, and Kristy Elizabeth Boyer. 2019. In Their Own Words: Gender Differences in Student Perceptions of Pair Programming. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19)*. Association for Computing Machinery, New York, NY, USA, 1053–1059. https://doi.org/10.1145/3287324.3287380

[126] Kimberly Michelle Ying, Fernando J. Rodríguez, Alexandra Lauren Dibble, Alexia Charis Martin, Kristy Elizabeth Boyer, Sanethia V. Thomas, and Juan E. Gilbert. 2021. Confidence, Connection, and Comfort: Reports from an All-Women's CS1 Class. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. Association for Computing Machinery, New York, NY, USA, 699–705. https://doi.org/10.1145/3408877.3432548

[127] Leila Zahedi, Jasmine Batten, Monique Ross, Geoff Potvin, Stephanie Damas, Peter Clarke, and Debra Davis. 2021. Gamification in education: a mixed-methods study of gender on computer science students' academic performance and identity development. *Journal of Computing in Higher Education* (March 2021). https://doi.org/10.1007/s12528-021-09271-5

[128] Daniel Zingaro. 2014. Peer instruction contributes to self-efficacy in CS1. In *Proceedings of the 45th ACM technical symposium on Computer science education (SIGCSE '14)*. Association for Computing Machinery, New York, NY, USA, 373–378. https://doi.org/10.1145/2538862.2538878

[129] Daniel Zingaro. 2015. Examining Interest and Grades in Computer Science 1: A Study of Pedagogy and Achievement Goals. *ACM Transactions on Computing Education* 15, 3 (July 2015), 14:1–14:18. https://doi.org/10.1145/2802752

A APPENDIX

Table 17. Database Queries

Database (s)	Date Executed	Query
ACM DL	5/15/2021	[[All: women] OR [All: female*] OR [All: gender]] AND [[All: cs0] OR [All: cs1] OR [All: cs2] OR [All: "introductory course*"]] AND [[All: survey] OR [All: empirical] OR [All: interview] OR [All: "focus group"]] AND NOT [[Abstract: k12] OR [Abstract: "high school"] OR [Abstract: "elementary school"] OR [Abstract: "primary grades"] OR [Abstract: "secondary school"]] AND [Publication Date: (01/01/2011 TO 12/31/2021)]
	5/20/2021	[[All: women] OR [All: female*] OR [All: gender]] AND [[All: "cs 0"] OR [All: "cs 1"] OR [All: "cs 2"]] AND [[All: [survey] OR [All: empirical] OR [All: interview] OR [All: "focus group"] OR [All:]]] AND [All: abstract:] AND [Publication Date: (01/01/2011 TO 12/31/2021)]
ERIC (FirstSearch)	5/26/2021	((el= "higher education")) and ((su= "computer science education") or (su= "computer science" and su= "education")) and (ab: female* or ab: women or ab: gender) and yr: 2011- and ln= "english" and pr: Yes.
ERIC	5/26/2021	(descriptor:"computer science education") and (educationlevel:"higher education") AND (ab: female* OR ab: women OR ab: gender) and peer reviewed only since 2012
Scopus	5/20/2021	(ABS ("computing education")) AND (ABS (wom?n) OR ABS (female*) OR ABS (gender)) AND PUBYEAR >2010 AND (LIMIT-TO (PUBSTAGE , "final")) AND (LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SRCTYPE , "p") OR LIMIT-TO (SRCTYPE , "j"))

Continued on next page

Table 17 – continued from previous page

Database (s)	Date Executed	Query
Web of Science	5/26/2021	TOPIC: (wom?n OR female* OR gender) AND TOPIC: ("computing education" OR "computer science") AND TOPIC: (educ*) NOT TOPIC: ("middle school" OR "high school" OR "secondary education" OR "elementary school") Refined by: LANGUAGES: (ENGLISH) AND [excluding] Databases: (MEDLINE) AND [excluding] DOCUMENT TYPES: (BOOK OR ABSTRACT OR BIOGRAPHY OR EDITORIAL MATERIAL OR DATA SET) AND COUNTRIES/REGIONS: (USA OR ENGLAND OR SCOTLAND OR AUSTRALIA OR UK OR CANADA OR NEW ZEALAND OR WALES OR IRELAND) Timespan: 2011-2021. Databases: WOS, BIOSIS, CCC, DRCI, DIIDW, SCIELO. Search language=Auto
IEEE Xplore	5/26/2021	("IEEE Terms":Computer science education) AND ("Abstract":wom?n OR "Abstract":female OR "Abstract":gender)
EBSCO (Academic Search Premiere)	5/26/2021	TX (wom?n OR female* OR gender) AND AB ("computing education" OR "computer science" OR "computing course*") AND AB students NOT TX ("middle school" OR "high school" OR "secondary education" OR "elementary school" OR "K-8" OR "K12" OR "K-12") Limiters - Scholarly (Peer Reviewed) Journals; Published Date: 20110101-20211231 Expanders - Apply equivalent subjects Narrow by Language: - english

Continued on next page

Table 17 – continued from previous page

Database (s)	Date Executed	Query
EBSCO (Education Source)	5/26/2021	computer science education AND AB (women or female or woman or females or gender) AND NOT TI(high school or middle school or secondary education or elementary education) Peer reviewed journals Pub year 2011-2021 Pub type: peer-reviewed journal Language: english
Compendex	5/20/2021	found in Compendex for 1884-2022: (((({COMPUTER SCIENCE - EDUCATION}) WN CV))
Springer	5/26/2021	Discipline: Education Subdiscipline: Educational Technology Date Published: 2011-2021 Content Type: Article (822) ("computer" or "computing") AND (wom?n OR female OR gender) AND (undergrad* OR post-secondary) AND (NOT (medical) OR NOT(tenure) OR NOT(wage))
Science Direct	5/26/2021	Year: 2011-2021 Article Type: Research Articles Publication Title: Computers & Education (685), Procedia - Social and Behavioral Sciences (169), and Computers in Human Behavior (44) Subject areas: Social Sciences (898), Computer Science (169) ("computer education" OR "computing education" OR "computer science education") AND (women OR female OR gender) AND (NOT (elementary OR secondary))

Continued on next page

Table 17 – continued from previous page

Database (s)	Date Executed	Query
ProQuest	5/27/2021	(wom?n OR female* OR gender) AND noft("computing education" OR "computer science" OR "computing course*") AND noft(students) NOT ("middle school" OR "middle school students" OR "high school" OR "high school students" OR "secondary education" OR "elementary school" OR "elementary school students" OR "K-8" OR "K12" OR "K-12") AND noft(course OR class) Peer-reviewed English only Excluded in post-search filters: from "subject" NOT (middle school students AND high school students AND secondary school students) Excluded all non-English speaking countries
PsycInfo	5/27/2021	computer science education AND AB (women or female or woman or females or gender) Peer reviewed journals Pub year 2011-2021 Pub type: peer-reviewed journal Language: english Age groups: Adulthood (18 yrs & older) Population group: human

Continued on next page

Table 17 – continued from previous page

Database (s)	Date Executed	Query
	5/27/2021	(noft((women OR female OR gender)) AND noft(("computing education" OR "computer science" OR "computing course*")) NOT noft((elementary OR K12 OR primary OR secondary OR "middle school" OR "high school" OR doctoral))) AND lo.Exact("US Virgin Islands" OR "Europe" OR "Wales" OR "England" OR "Scotland" OR "United Kingdom" OR "Northern Ireland" OR "North America" OR "New Zealand" OR "Great Britain" OR "Appalachia" OR "Ireland" OR "British Virgin Islands" OR "Georgia" OR "Australia" OR "US") AND cl("Cognitive Processes" OR "Curriculum & Programs & Teaching Methods" OR "Learning Disorders" OR "Learning & Motivation" OR "Professional Personnel Attitudes & Characteristics" OR "Community & Social Services" OR "Industrial & Organizational Psychology" OR "Personnel Management & Selection & Training" OR "Social Processes & Social Issues" OR "Occupational Interests & Guidance" OR "General Psychology" OR "Social & Instinctive Behavior" OR "Classroom Dynamics & Student Adjustment & Attitudes" OR "Social Psychology" OR "Group & Interpersonal Processes" OR "Professional Education & Training" OR "Sex Roles & Women’s Issues" OR "Social Perception & Cognition" OR "Occupational & Vocational Rehabilitation" OR "Educational/Vocational Counseling & Student Services" OR "Social Structure & Organization" OR "Communication Systems" OR "Engineering & Environmental Psychology" OR "Educational Administration & Personnel" OR "Educational Psychology" OR "Organizational Behavior" OR "Developmental Psychology" OR "Lifespace & Institutional Design" OR "Educational Measurement" OR "Learning & Memory" OR "Management & Management Training" OR "Culture & Ethnology" OR "Academic Learning & Achievement") DATE: 2011-01-01 - 2021-12-31 Peer-reviewed Age group: 18+ adults and subcategory “young adults” English only
Google Scholar	5/26/2021	women OR OR OR female OR OR OR gender "computer science education" -secondary -OR -primary -OR -elementary -OR -"high school" -OR -doctoral -OR -"middle school" -OR -"IGI Global" English Only

Continued on next page

Table 17 – continued from previous page

Database (s)	Date Executed	Query
CSEd Research	5/20/2021	First pass filters: Undergraduate; boy/men, girl/women, transgender Second pass: using filter only for girl/women yielded a couple more items Conservatively excluded items that were obviously not relevant per our inclusion criteria Searched on "women" (without filters) and compared with previous lists, as a check.
ProQuest Sociology	5/28/2021	noft((gender OR women OR female)) AND noft(("computing education" OR "computer science" OR "computing course*")) NOT noft((elementary OR K12 OR primary OR secondary OR "middle school" OR "high school" OR doctoral)) English only Peer-reviewed 2011-2021 Removed irrelevant subfields

Table 18: Annotated Bibliography of Papers Analyzed

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
Christine Alvarado, Zachary Dodds, and Ran Libeskind-Hadas. 2012. Increasing women's participation in computing at Harvey Mudd College. ACM Inroads (Dec. 2012). https://doi.org/10.1145/2381083.2381100 [1]	USA	intro computing courses (CS0, CS1, CS2)	449 graduates	gender	Research experience	percentages of women in cs course	The paper reports on an intervention implemented in Harvey Mudd College over a period of four years: the intervention focused on impacting three key factors contributing to why females do not major in CS namely: increasing potential female CS majors' interest, confidence, and sense of belonging. Their intervention included the following: a. altering the introductory cs course so as to give first-year students a broader view of CS, demonstrate the breadth of the discipline, and quickly immerse students in various core topics and activities offering each female first-year student, independent of planned major, the opportunity to attend the Grace Hopper Celebration of Women in Computing which makes students aware of the variety of jobs available within the discipline, showcases role models at all career stages, and offers an effervescent and welcoming culture and c. offering research projects to first-year students which gives them the chance to apply their knowledge, boost their confidence, and deepen their interest in the discipline. The results of the intervention were dramatic: the percentage of women CS majors climbed from 10 percent in 2005, to 20 percent in 2006, to the current annual average of between 37 and 50 percent. In just four years, HMC more than tripled its percentage of women CS majors.
Christine Alvarado and Eugene Judson. 2014. Using targeted conferences to recruit women into computer science. Commun. ACM (March 2014). https://doi.org/10.1145/2500883 [2]		intro computing courses (CS0, CS1, CS2)		gender		student survey(s)	The papers reports on the effects of attending a conference intervention had on female students' decision to major in computer science and follow corresponding career paths. The intervention conducted at Harvey Mudd College in which female students attended Grace Hopper Celebration (GHC) conference. GHC is an annual conference celebrating the accomplishments of women in CS, combining technical talks, targeted workshops, panels focused on issues facing women in the field, and networking events. The researchers conducted a survey on GHC attendees to try to understand the effect the trip had on the students who attend. They first administered a pre-survey prior to students' attendance to gather specific student profile and attitudinal information. Following attendance at GHC, the authors administered the post-survey to gather information about attendees' sentiments regarding the degree to which GHC participation had affected their views of CS. The results indicate GHC had a positive effect on attendees, especially those with some CS experience or interest. By examining enrolment patterns, the authors also found attendees were much more likely to take another course and major in CS than female non-attendees. 25% of those not considering a CS major and attended GHC ended up majoring in CS after attending the conference. Finally, those attending GHC take a second CS course at a significantly higher rate than those who do not attend GHC.

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
Christine Alvarado, Cynthia Bailey Lee, and Gary Gillespie. 2014. New CS1 pedagogies and curriculum, the same success factors?. SIGCSE '14. https: //doi.org/10.1145/2538862. 2538897 [3]		intro computing courses (CS0, CS1, CS2)	313M 120F	gender	Pair Programmi ng/ Peer Instruction students paired by confidence level	student survey(s), performance indicators (e.g., grades, test scores)	Context: US university CS1 Intervention: Pair Programming and Peer Instruction Participants: F120 M313 Method:sStudents were put into pairs/groups for Pair Programming (PP) and Peer Instruction (PI) with differing or similar confidence and experience by design. Student results and opinions were compared by group and gender Results. For men, confidence predicts exam performance, for for women it does not. No significant differences between response to survey questions for any of the groups by gender.
Christine Alvarado, Mia Minnes, and Leo Porter. 2017. Micro-Classes: A Structure for Improving Student Experience in Large Classes. SIGCSE '17. https://doi.org/10.1145/30 17680.3017727 [4]	USA		200 (intervent ion) 276 (control)	gender	micro classes taught by TA	student survey(s), performance indicators (e.g., grades, test scores)	CONTEXT: Study created "micro-classes"—small groups of 25-30 students within a large class (200+ students) using additional graduate TAs and undergraduate tutors -- in an upper division advanced data structures course. Examined two research questions: 1) Will students in the micro-classes section report a stronger sense of classroom community? and 2) Will students in the micro-classes section achieve better academic performance, specifically higher student retention and better performance on assessments? POPULATION: approx. 676 students at a large public research university in the US, 200 of which were in a course section with the micro-class format. METHOD: Data were collected from 3 course sections: one using the micro-classes format, one using a more traditional format taught in the same term and another taught in a previous term. Performance data included attendance, grades on each assessment, and overall course grades. Additional data included end-of-term evaluations (student satisfaction and perceived learning) and the Classroom Community Scale (CCS) instrument. RESULTS: The micro-class section had a significantly higher proportion of women enrolled than either control (23.9% vs 13.3% and 15.5%, p<0.05). The micro-class course design was explained to students on the first day of course and there was some switching of sections by students during the first week. The community score was significantly higher in the micro-class section than either control section. Community results by gender were not significant. There were no significant differences in course performance by gender, CS major status, and GPA for all graded assessments. Of note is that the control courses already use active learning strategies such as peer instruction and so the comparison was not against a traditional lecture-based course.
Valerie Barr. 2016. Disciplinary Thinking, Computational Doing:	USA	intro computing courses (CS0,	6 years (no		Theme-based	student survey(s), course and	CONTEXT: Replaced traditional CS1 and CS0 courses with 6 theme-based CS1 courses from which students can select. Six intermediate theme-based courses were also developed, though they are more targeted at non-majors interested in

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
Promoting Interdisciplinary Computing While Transforming Computer Science Enrollments. ACM Inroads (2016). [11]		CS1, CS2), other computing courses, non-CS courses	student numbers)		computing curriculum	major enrollments	interdisciplinary applications of computing. POPULATION: Small liberal arts college in USA. 58 sections of introductory courses (no N for students provided?) over 6 years METHOD: 41% of intro course sections surveyed (pre and post course). Intro course themes included : big data, robotics, game development, artificial intelligence, media computation, and engineering applications. Pedagogy used in the course also changed from lecture with a separate lab to class sessions that included lecture with hands-on activities (supported with a student assistant) and limited to 24 students/section. Intermediate course themes included web programming, CS of games, bioinformatics, NLP, modeling and simulation, and data visualization. RESULTS: Intro courses were 38% female and 72% white (mirroring college's race/ethnicity demographics). Over 57% of students expressed interest in CS or applications of computing and 58% indicated they might take additional CS courses (answered yes or maybe). Student interest data not provided by gender. Enrollments in the intermediate courses also showed increases in the proportion of female students from the first offering to the most recent offering. Student interest in computing in the intermediate courses was also very positive. Reports an increase in the percentage of majors overall and the percentage of female majors though specific data is not included.
Debarati Basu, Harinni K. Kumar, Vinod K. Lohani, N. Dwight Barnette, Godmar Back, Dave McPherson, Calvin J. Ribbens, and Paul E. Plassmann. 2020. Integration and Evaluation of Spiral Theory based Cybersecurity Modules into core Computer Science and Engineering Courses. SIGCSE '20. https: //doi.org/10.1145/3328778. 3366798 [13]	USA	other computing courses, cyber-security	512 students across 6 courses	gender, race. ethnicity	Spiral-based curriculum design	student survey(s), performance indicators (e.g., grades, test scores)	Context US undergraduates addition of cybersecurity modules to existing CS modules. Method: Pre and post-course test that included questions on whether students found the course useful and interesting as well as about subject knowledge. Intervention: Added a new course and included spiral theory. However, there was no previous course or control group. Results. Students met the learning objectives of the course and there was little difference in measured outcomes between gender and ethnicity groups.
Kara Alexandra Behnke, Brittany Ann Kos, and John K. Bennett. 2016. Computer Science Principles: Impacting Student Motivation \& Learning		intro computing courses (CS0, CS1, CS2)		gender		performance indicators (e.g., grades, test scores); student surveys	Behnke et al studied impacts of University of Colorado-Boulder's adoption of a Computer Science Principles class on students (n=94) from several different majors. They reported positive outcomes for students' motivations based on surveys. They also found no significant difference in grade letters achieved by students' reported genders (binary selection of male or female).

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
Within and Beyond the Classroom. ICER '16. https://doi.org/10.1145/2960310.2960336 [14]							
Michael David Black. 2016. Seven Semesters of Android Game Programming in CS2. ITiCSE '16. https://doi.org/10.1145/2899415.2899470 [15]		intro computing courses (CS0, CS1, CS2)		gender		performance indicators (e.g., grades, test scores)	The paper reports on an open-ended Android game development project incorporated in the introduction to computer science course at a small computer science department at American University. The aim was to engage students , motivate them to study computer science and give them practical experience. This Android game project primarily consisted of a three-week open-ended project where students, mostly working in teams, developed an Android game of their choosing. Students were required to write a proposal describing their intended game, were given two weeks of development time, and demonstrated their game to the class during the final lecture period. Bonus points were awarded if the students published their game to the Google Play Store. The goals of the project were: 1) to engage students by offering them a "real world" platform, 2) to show students that they were able to do mobile development, 3) to reward students by publishing their work, and 4) to give them a marketable skill that could lead to internships. The authors provide examples of projects that exemplified some specifically relevant factors that motivate students and engage particular female students as stemming from the EngageCSSEdu initiative and include: "Student-Faculty Interaction", "Student-Student Interaction", "Incorporating Student Choice", "Make Interdisciplinary Connections to CS", and "Problem-based Learning". The authors explain that their intervention had an impact on students pursuing jobs and internships related to mobile design. Regarding minority and gender differences, the authors do not provide much information other than that the score of these groups was lower than the average score on completeness of the project and correctness, but the difference doesn't appear to be significant.
Nicholas A. Bowman, Lindsay Jarratt, KC Culver, and Alberto M. Segre. 2021. The Impact of Pair Programming on College Students' Interest, Perceptions, and Achievement in Computer Science. ACM TOCE (May 2021). https://doi.org/10.1145/3440759 [17]	USA	intro computing courses (CS0, CS1, CS2)	622 students in pairs, 576 control		pair programming	student survey(s), performance indicators (e.g., grades, test scores)	The present study explores the causal effects of pair programming on undergraduates students' performance, interest, plans for future coursework, confidence, comfort, and anxiety with computer science as it has been widely implemented in computer science (CS) coursework. Specifically, almost 100 computer science lab sections were randomly assigned to use pair programming or individual programming, thereby avoiding problems with self-selection and limited sample sizes that are pervasive in previous research. The results identified virtually no significant benefits of pair programming across the desired outcomes. This pedagogy also does not appear to affect long-standing inequities in computer science, since the results were similar regardless of students' sex, race, first-generation status, U.S. citizenship, year in college, prior programming experience, and the type of introductory course. This study provides intriguing evidence that pair programming, at least under certain conditions, may not

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							promote desired student outcomes or reduce equity gaps in computer science coursework.
Bo Brinkman and Amanda Diekman. 2016. Applying the Communal Goal Congruity Perspective to Enhance Diversity and Inclusion in Undergraduate Computing Degrees. (SIGCSE '16). https://doi.org/10.1145/2839509.2844562 [18]		other computing courses	19	gender, race. ethnicity	cohort-based service-learning program - students live together, take 2 courses in common during 1st semester. one service activity/yr, dedicated study space, talks from guest speakers in service	other archival	The paper focuses on communal goal congruity as a contributing factor to the lack of diversity in IT fields. After presenting evidence that demonstrates that there is a general view that tech careers do not include the career goals like "working with others and in the service of others", the authors report the results of a long-term intervention designed based on the communal goal affordance theory, a theory that supports that communal goals (a goal that focuses on working with, or in the service of, others, with an altruistic or outward focus) in particular influence STEM decisions. The program's communal focus was on providing cohort dorm living, courses in common, peer advisors, service projects, dedicated group study space, and career counselling. To demonstrate the effectiveness of their approach, the authors compared the diversity levels of their program with the diversity levels of the baseline population representing students at their university majoring in computer engineering, computer science, electrical engineering, and software engineering. The results of their study depict that the demographics of the group that chose to apply and admitted to the program were already significantly more diverse than the baseline population. Without at any point explicitly focusing on diversity, the authors argued that they have managed to achieve diversity in our S-STEM program simply by designing and marketing the program around communal goal affordances.
Jeni L. Burnette, Crystal L. Hoyt, V. Michelle Russell, Barry Lawson, Carol S. Dweck, and Eli Finkel. 2020. A Growth Mind-Set Intervention Improves Interest but Not Academic Performance in the Field of Computer Science. Social Psychological and Personality Science (Jan. 2020). https: //doi.org/10.1177/1948550619841631 [19]	USA	intro computing courses (CS0, CS1, CS2)	493 introductory computer science students	gender	growth mindset		Context: seven US universities/colleges, introductory CS Intervention: online, 4-session growth mindset intervention Participants: 143F 348M, 68% white 7 institutions, 16 classes. Split evenly between intervention and control Method: pre- and post- survey with control Results: students in the growth mindset condition reported stronger growth mindsets at posttest compared to control; growth mindset intervention failed to significantly predict grades; growth mindset condition reported greater career interest at posttest; contrary to expectations performance of the intervention did not depend on gender
Zack Butler, Ivona Bezakova, and Kimberly Fluet. 2017. Pencil Puzzles for Introductory Computer Science: an Experience- and Gender-Neutral Context.		intro computing courses (CS0, CS1, CS2)		gender		student survey(s), performance indicators (e.g., grades, test scores),	Context US Undergraduate CS1/2 over three years Population: Three groups responded to the surveys each within a cohort of @500 students. Groups were 340 student (12.6% female), 80 (15%), 64 (11%). Method: Surveys and grades of a sample

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
SIGCSE ’17. https://doi.org/10.1145/3017680.3017765 [20]						Student comments	Intervention: Using pencil puzzles to learn cs concepts. Results: There was no association between gender and grades, student perceptions, or student reflective thinking. Pencil puzzle based assignments can be effective in teaching students of varying experience levels. This indicates that pencil puzzles are a leveling context for the instruction of CS topics.
E. Rebecca Caldwell and Elva J. Jones. 2013. Beyond wrestling: using sumobots to engage students in the computer science classroom. Journal of Computing Sciences in Colleges (Dec. 2013). [21]	USA	intro computing courses (CS0, CS1, CS2)	Not provided	gender	Robotics in a computer software systems course	student survey(s), performance indicators (e.g., grades, test scores)	Context: US university, All African-American students Intervention: robotics are introduced into a computer systems course Participants: 4F 26M 1 institution 1 class. control group is previous cohort (no gender breakdown) Method: Marks for different parts of the process were described. Post-survey of student opinion, not reported by gender. Marks for the course were compared with previous cohore Results. Females students average score on the assembling the sumobot quiz was 92/100 while the average score for the males was 84/100 - no discussion of significance.
Jennifer Campbell, Diane Horton, Michelle Craig, and Paul Gries. 2014. Evaluating an inverted CS1. SIGCSE ’14. https://doi.org/10.1145/2538862.2538943 [23]		intro computing courses (CS0, CS1, CS2)		gender		student survey(s), performance indicators (e.g., grades, test scores)	Context: CS1 course at R1 University in North America that users Python and takes an objects-early classes-late approach. They evaluated the impact of an inverted classroom offering. Population: 132 students completed the pre and post term survey. Method: Course was evaluated on a number of factors – attendance, lecture preparation, online helpfulness, enthusiasm, enjoyment, workload, comparable performance, pass rate, Result: Other than for workload there was no mention of gender-specific differences URM relevant result, “Women found the course more difficult and spent more time on it than men” (see section 4.6)
Yingjun Cao and Leo Porter. 2017. Evaluating Student Learning from Collaborative Group Tests in Introductory Computing. SIGCSE ’17. https://doi.org/10.1145/3017680.3017729 [25]	USA	intro computing courses (CS0, CS1, CS2)	228	gender	Group exams	performance indicators (e.g., grades, test scores)	CONTEXT: Impacts on student learning are examined using a two-stage collaborative group exam. In a two-stage exam, students take the first part individually and a second part is taken in a group where students collaborate to submit one answer. The research questions studied were: 1) Do students learn more from the group-retest than an individual-retest? and 2) Do different student groups benefit more from the group-retest than others? POPULATION: CS1.5 course at a large research university in the US with 228 students in two sections of the course.

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							METHOD: Designed as a randomized, crossover study with two experiments (midterm exams 1 and 2). Each midterm exam included two topics and both topics were tested as part of the individual portion of the exam (pre-test). Students were randomly assigned to one topic as the control (re-test as an individual) and the other topic as the experiment (re-test as a group). A quiz was administered about two weeks following the exam (post-test). Scores from an individual final exam were also collected. RESULTS: In the first midterm, students in the group re-test had significantly more learning gains than those in the individual re-test (p=0.003), however, neither group experienced learning gains in the second midterm. The learning gain effect was not seen in the final exam for topics on either midterm. Women benefited more than men from groups on the polymorphism topic on midterm 2. It was noted that over 50 statistical tests were performed and as such some results may be significant, however, when Bonferroni corrections were applied, this gender difference was not significant.
Kim Case, Robert Bartsch, Lillian McEnery, Sharon Hall, Anthony Hermann, and David Foster. 2008. Establishing a Comfortable Classroom from Day One: Student Perceptions of the Reciprocal Interview. College Teaching (Sept. 2008). https://doi.org/10.3200/CTCH.56.4.210-214 [26]	USA	intro computing courses (CS0, CS1, CS2), lecture, lab, psychology, education, women's studies	unclear how many CS students (one of four groups in 120 students) no control	gender, race. ethnicity	A one-off first day activity where students interview the instructor and vice versa.	student survey(s)	Context: one institution (not stated,presumed US) Intervention: Early in the course faculty interview (survey) students e.g. on their goals for the course and then allowed students to interview (ask questions) of the professor. Participants: 160F 40M across multiple disciplines. Number in CS is unclear. Method: after the intervention a Likert survey was used to evaluate the activity. Results: no significant difference between genders either in the main variables (average general attitude, discipline) or in combination (2-way ANOVA). 2-way ANOVA does show significant difference in attitude to the intervention and ethnicity, with "Minority" students having a more positive attitude the "White", although this is across all disciplines, not specifically CS
Sapna Cheryan, Andrew N. Meltzoff, and Saenam Kim. 2011. Classrooms matter: The design of virtual classrooms influences gender disparities in computer science classes. Computers & Education (Sept. 2011). https://doi.org/10.1016/j.compedu.2011.02.004 [28]	USA	Pre-enrollment experiences about potential CS courses	59 students	gender, race. ethnicity	classroom environment	student survey(s)	NOTE: Rather than avoiding stereotypes (NCWIT practise) this intervention points to introducing stereotypes - classrooms with "feminine" artefacts. Women were more likely to enrol in a "feminine" non-stereotype CS classroom than a "masculine" CS stereotype classroom but least likely to enrol in a neutral classroom. "Context: Three experiments exploring 3D virtual classrooms with stereotypical vs non-stereotypical artefacts in them. Each experiment was in the context of a CS course (pre-enrolment) experience that is presented to students using 3D virtual classrooms for Non CS university students. Population: 156 students took part. Experiment 1 59 (male and female) non-cs majors. Experiment 2. 62 (male and female) non-cs majors. Experiment 3 35

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							females non cs-majors. Method: 1. The two virtual classrooms were pre-tested to evaluate if it was "stereotypical" by a post-use survey of 70 undergraduates (48 females) who rated the two online classrooms as stereotypical or not and how masculine or feminine the environment was. The stereotypical classroom had items such as science fiction books, computer parts, electronics, software technology magazines, video games, computer books, Star Wars and Star Trek items. The non CS stereotypical had nature and art posters, plants, lamps, general magazines and water bottles. 2. The interventions were evaluated through post-intervention surveys which took place after the participants had moved around the 3D classrooms. The surveys investigated a) enrollment intentions for related CS courses (all interventions) b) ambient belonging in the virtual classrooms (all interventions) c) expected success in the classroom (intervention 2 and 3) c) Mediation analysis was used to examine whether ambient belonging mediated the gender disparity in enrollment intentions and expected success. Intervention: Intervention 1: Non-CS majors moved around the two virtual classrooms (it is not clear for how long). To ensure that students engaged with the "spaces" they were asked questions about their knowledge of the classrooms. Intervention 2: Same as Intervention as 1 except, presentation of rooms was counter-balanced and asked questions on how likely they were to take EACH class and anticipated success in each class. Intervention 3. Same as Intervention 1, except a third room with no gendered artefacts was presented and the order of presentation of rooms was counterbalanced. Results: Across all three interventions. 1. A majority of men and a minority of women chose to take CS in the stereotypical classroom. 2. Women expressed greater intention to take CS in the non-stereotypical classroom. 3. Women felt significantly lower ambient belonging in the stereotypical classroom than the non-stereotypical. 4. Men felt no difference in ambient belonging between the two classrooms. 5. Mediation analysis indicated that ambient belonging significantly predicted intention to enrol in the stereotypical class. 6. Women reported increased anticipated success in the non-stereotypical classroom, and men marginally reported the opposite.

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							7. Women were significantly more interested in enrolling in the non-stereotypical class over both the stereotypical and neutral class and marginally less interested in the neutral class compared to the stereotypical class. 8. Women anticipated performing better in the stereotypical class than in the non-stereotypical and neutral class (with no difference in anticipated success between the two). 9. Women reported a significantly greater sense of ambient belonging for the non-stereotypical class than the other two classes (with no difference between the two). The upshot is to include artefacts that women associate with their own gender to attract more women to enrol, help them feel like they belong and will succeed. However, this could mean some men may be less likely to enrol or may feel like are marginally less likely to succeed. (There were mixed findings across the two experiments on these aspects.) However, men may feel like they belong more in the non-stereotypical classrooms too.
Sapna Cheryan, Benjamin J. Drury, and Marissa Vichayapai. 2013. Enduring Influence of Stereotypical Computer Science Role Models on Women's Academic Aspirations. Psychology of Women Quarterly (March 2013). https: //doi.org/10.1177/0361684 312459328 [27]		Non-computing majors (Psychology majors)		gender		student survey(s)	Context: This study investigates the effects of stereotyped role models on female students interest in majoring in CS. Population: 100 female undergraduate students not majoring in CS Methods: Participants and student actors asked each other provided questions. The student actors all had the same responses to demographic questions, and then the actors' answers to hobbies and interests varied according to stereotypicality. For example, stereotypical responses to favorite magazine were "Electronic Gaming Monthly," where a non-stereotypical response was "Rolling Stone." After interacting, participants filled out a questionnaire that asked them to rank how likely they were to major in CS, how much they've considered majoring in CS, to measure interest and sense of belonging in CS. Participants were sent an identical questionnaire to fill out two weeks after the laboratory experiment. Results: Interacting with a stereotypical role model reduced women's interest in majoring in CS and sense of belonging in CS.
Sapna Cheryan, John Oliver Siy, Marissa Vichayapai, Benjamin J. Drury, and Saenam Kim. 2011. Do Female and Male Role Models Who Embody STEM Stereotypes Hinder Women's Anticipated Success in STEM? Social Psychological		Non-CS majors		gender	Presentation of stereotyped models of computer scientists	student survey(s)	Context: How does gender and stereotypicality of role models affect women's beliefs about being successful in computer science? Are these effects different for men? This paper is two separate laboratory studies that explore these question. Population: Study 1 - 85 female non computer science majors, Study 2 - 68 non computer science majors, 40 of which were women Methods: Study 1 - Participants and student actors asked each other provided questions.

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
and Personality Science (Nov. 2011). https://doi.org/10.1177/1948550611405218 [29]							The student actors all had stereotypical outfits, the same responses to demographic questions, and then the actors' answers to hobbies and interests varied according to stereotypicality. For example, stereotypical responses to favorite magazine were "Electronic Gaming Monthly," where a non-stereotypical response was "Rolling Stone." After interacting, participants filled out a questionnaire that asked them to rank how well they thought they would do if they majored in CS. Study 2- Participants created an avatar in a virtual world (Second Life) and then interacted with a partner's avatar whose stereotypically was manipulated for each participant. They asked each other the same questions as Study 1 and, afterwards, completed the same questionnaire as Study 1 participants. Results: Study 1 - Stereotypical computer science role models negatively affect women's perceived success in CS. Study 2 - Men's beliefs in succeeding in CS are not affected by stereotypical CS role models. Across both studies, female role models are no more effective than male role models in improving women's success beliefs around CS.
James P. Cohoon and Luther A. Tychonievich. 2011. Analysis of a CS1 approach for attracting diverse and inexperienced students to computing majors. (SIGCSE '11). https://doi.org/10.1145/1953163.1953217 [30]		intro computing courses (CS0, CS1, CS2)	210 (intervention) 604 (control)	gender, race. ethnicity	Small CS class with a range of interventions	enrollments	CONTEXT: A CS1 course was developed for students with no prior programming experience. The course included an integrated lecture and lab, many small examples and assignments, student participation, and limited enrollment to 70 students. POPULATION: students at a US state university (over 6 years), N is not clear -- reported as 1,378 in Table 1 and 4,395 in Table 2. The experimental group (CS1X) included 210 students (Table 1) and 604 (Table 2). METHOD: The adaptations made to the intervention -- CS1X course for students with no prior programming experience, admitted via an interview -- include a class size capped at 70 students in a room where the lab and lecture activities were able to be integrated and TAs are available to help in class, a focus on small programming examples and assignments that were meaningful and relevant to students, personal encouragement from the instructor and/or teaching assistants, and a classroom culture that emphasized learning. RESULTS: Overall enrollment in CS1 has increased over the six years for which data was reported and the proportions of black and female students in CS1X as compared to other CS1 sections has increased significantly (1.6x female, p-value=0.01; 1.2x black, p=0.004). Data on intent to major in CS after the course is difficult to compare directly to other sections, which tend to attract students with experience having already declared CS majors. However, CS1X shows a similar percentage of students who were undeclared at the beginning of the course intending to major in CS. Focus group results indicate most students were

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							encouraged by the instructor and intend to take additional CS courses or major in CS and felt confident they could be successful. An extensive table is included which provides data on preferences for assignment topics by gender.
Ralph W. Crosby, Stephanie Valentine, and Tiffani L. Williams. 2014. Leveraging programming difficulty to improve understanding and perceptions of nonmajors. Journal of Computing Sciences in Colleges (April 2014). [31]	USA	intro computing courses (CS0, CS1, CS2)		gender	Challenging course assignment s	student survey(s), performance indicators (e.g., grades, test scores)	
Matthew Dickerson. (2014). Multi-agent simulation, netlogo, and the recruitment of computer science majors. Journal of Computing Sciences in Colleges. [34]		intro computing courses (CS0, CS1, CS2)		gender		student survey(s), enrollments and retention	CONTEXT: New CS1 course (CSCI 190) based on Multi-agent simulation (MAS) / agent-based modeling (ABM) paradigm using the NetLogo programming language, one of three CS1 courses that all feed into the same CS2 course. The other two CS1 classes are taught in Python. POPULATION: undergraduates in a private liberal arts college in the US over three years (2011-12, 2012-13, and 2013-14) In 2013-14, there were 37 students in CSCI 190, vs 32 and 51 in the other two sections of CS1). METHOD: Student enrollments in CS1, retention rates of students taking additional CS classes, performance of students in CS2 class, and perception questions on entry and exit exams for CS1 students. RESULTS: Female enrollments in this NetLogo version of CS1 was higher (49% vs 31% and 39%) than the other two sections of CS1 in 2013-14. Fewer of the CSCI 190 students indicated CS as an intended first major, compared to the other CS1 sections, suggesting this CS 1 is a better recruitment course, but makes it difficult to compare retention between CS1 sections. Students from CSCI 190 did not perform as well as students from the other classes when moving to CSCI 201—a more traditional CS2 course taught in Java—though they did not fare poorly. The GPAs in CSCI 201 were 3.67, 3.70, and 3.49 respectively for CSCI 101, CSCI 150, and CSCI 190.
Ann Quiroz Gates, Sarah Hug, Heather Thiry, Richard Aló, Mohsen Beheshti, John Fernandez, Nestor Rodriguez, and Malek Adjouadi. 2011. The Computing Alliance of Hispanic-Serving Institutions: Supporting		other computing courses, research focused program		gender, race. ethnicity		student survey(s)	Section 4.6 starting on pg 16:14 FemProf - research mentorship program, female students partnered with researchers to do research internships/seminars (2 year program) Collaborating between University of Puerto Rico + University of Houston-Downtown 19 female students surveyed

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
Hispanics at Critical Transition Points. ACM TOCE (Oct. 2011). https://doi.org/10.1145/2037276.2037280 [43]							- increased confidence + sense of beloging - increased interest in graduate school/research trajectory -
Danielle Geerling, Jonathan Butner, Tamra Fraughton, Sungchoon Sinclair, Joseph Zachary, and Carol Sansone. 2020. The Dynamic Association of Interest and Confusion: The Potential for Moderation by Utility Value and Gender. The Journal of Experimental Education (May 2020). https://doi.org/10.1080/00220973.2018.1561403 [44]		intro computing courses (CS0, CS1, CS2)		gender		performance indicators (e.g., grades, test scores)	Context: On-line course HTML & JS, CS1 in a US university Intervention: Selected students were given Utility-Value (UV) intervention throughout lab sessions. Specifically, students were provided with a list of nine potential applications of what they would be learning, and were asked to select the three applications that were the most personally useful to them Participants: 16F 47M, 30 intervention, 33 control, 1 institution, 1 class Method: the intervention group received the intervention whenever they accessed material Results. for women in the UV condition, higher levels of confusion were associated with increases in interest over time, while in the control condition, higher levels of confusion were associated with decreases in interest over time. Thus, it appears that the UV intervention switched the direction of the association between women’s, but not men’s, confusion and change in interest over time.1
Mark Guzdial. 2013. Exploring Hypotheses about Media Computation. ICER ’13. https://doi.org/10.1145/2493394.2493397 [47]	USA	intro computing courses (CS0, CS1, CS2)	No details	gender	Media computation	student survey(s)	The two downstream papers should be reviewed rather than this summary paper. But for some reason neither papers are in our sample, perhaps because of dates. Therefore this summary paper has been included. The approach to engage with females was to use motivating contexts such as media, storytelling and robotics. Context: US undergrad introduction to CS called MediaComp for non CS majors or CS majors Population: Data not available in the summary paper. Method: Student surveys Intervention: There is limited information in the summary paper. Results: The summary paper states "We have evidence that the contextualization aspect of MediaComp influences female students perception of relevance, which motivates success, and leads to greater retention." "We did

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							two interview studies to explore why women found the class more motivating than the traditional course. Female students in our first study told us that they liked the content more in the MediaComp class [13]. In a larger study in which we interviewed women in both our MediaComp and our traditional course, students emphasized the sense of relevance. The women we interviewed valued that the content was useful [35]. The downstream papers are [13] A. Forte and M. Guzdial. Computers for communication, not calculation: media as a motivation and context for learning. In Proc. 37th Annual Hawaii International Conference on System Sciences, pages 10 pp.–, 2004. [35] L. Rich, H. Perry, and M. Guzdial. A CS1 course designed to address interests of women. In Proceedings of the ACM SIGCSE Conference, pages 190–194, 2004.
Hussam A. Hallak, Soad Ibrahim, Christy Low, and Ayman El Mesalami. 2019. The Impact of Incorporating Hands-on Raspberry Pi Projects with Undergraduate Education in Boosting Students' Interest in Scientific/Engineering Majors and Encouraging Women and Minorities to Advance their Integration in Practical Fields. (2019). [48]	USA	intro computing courses (CS0, CS1, CS2)	37	gender, race. ethnicity	Physical computing (Raspberry Pi)	student survey(s), performance indicators (e.g., grades, test scores)	Context: Paper evaluates the use of raspberry pi computers in to provide active learning experiences and make the course more attractive to underrepresented minorities Population: 37 North American Distance Learning CS1/CS2 students Method: Students were provided raspberry pi computers (in person) or purchased their own kits (online) to complete interactive "hands-on" assignments in CS1 & CS2 courses. Following the courses surveys were completed asking about the students perceptions of the impact of using the hands on hardware based assignments. Result: Grades improved by 25% and lab attendance increased by 10% for female students (corresponding numbers not given for male students). Overall survey results were positive for both genders, the only instance where there was a difference was " 5. Does the use of Raspberry Pi mini-computer in the lab as a group project help increase your communication level with your classmates?" Most males said not, most females said yes.
Brian Harrington, Jingyiran Li, Mohamed Moustafa, Marzieh Ahmadzadeh, and Nick Cheng. 2019. On the Effect of Question Ordering on Performance and Confidence in Computer Science Examinations. SIGCSE '19. https://doi.org/10.1145/32		intro computing courses (CS0, CS1, CS2)		gender, race. ethnicity		performance indicators (e.g., grades, test scores)	This paper examines the effects of question order on students' perception of their own achievement. The researchers randomly administered students to exams conditions which included questions ordered either from easy-to-hard or hard-to-easy. After the exam, the students were asked to predict their marks on a per-question basis. A total of 243 students participated. The results demonstrate that the question order had a small but not statistically significant effect on the performance, and virtually no effect on predicted marks when treating the entire class as one unstratified sample. However, the effect was significant for certain subgroups created via stratification. In particular, swapping the order of the questions (hard-to-easy) may have hurt the performance of

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
87324.3287398 [49]							international students, but significantly raised both the performance and confidence of female students.
Bryan Hosack, Billy Lim, and W. Paul Vogt. 2012. Increasing Student Performance Through the Use of Web Services in Introductory Programming Classrooms: Results from a Series of Quasi-Experiments. Journal of Information Systems Education (2012). [53]	USA	intro computing courses (CS0, CS1, CS2)		gender	Using data and real-world problems	performance indicators (e.g., grades, test scores) gpa, class rank	CONTEXT: Two introductory programming course (one for IS majors and one for CS majors) were redesigned to use web services in order to incorporate data and real world problems more likely to be of interest to students. The focus of the project was on measuring student learning. POPULATION: large research university in the US; 586 students over two years METHOD: For each semester offering of the course there were two sections - one experimental using web services problems and another control section. Scores on a common final exam were analyzed, taking into account major, gender, GPA, and class rank. RESULTS: Of the four semesters in which the course was offered, if as suggested by the authors, the data from the third semester is dropped, then modest positive gains in scores are seen on the final exam. If the third semester is included, then it cancels out the gains from the other semesters. No significant differences were seen by gender.
Sarah Hug, Heather Thiry, and Phyllis Tedford. 2011. Learning to love computer science: peer leaders gain teaching skill, communicative ability and content knowledge in the CS classroom. SIGCSE '11. https://doi.org/10.1145/1953163.1953225 [54]	USA	intro computing courses (CS0, CS1, CS2),lab	89 of which 73 female (82%) (TAs intervention) no control	gender, race	Peer-led team learning Leaders	student survey(s)	Context: Benefits to peer leaders leading CS related lessons and activities at Computing Alliance for Hispanic Serving Institutions (CAHSI) in the US. PLTL sessions occur in a 45-60 minute block of time during the laboratory session of each course at one of the institutions (not necessarily all six institutions). Intervention: Students led PLTL (Peer Led Team Learning) sessions. Leaders are paid a modest stipend, and are recruited by faculty and by their peers as being knowledgeable in course content. They attend a day-long training and are required to attend all class lectures and weekly meetings with program coordinator. They submit weekly reports and meet regularly with instructors. Participants: 89 peer leaders over five college semesters and six Computing Alliance for Hispanic Serving Institutions (CAHSI) Method: self-report surveys at the end of the semester. 44 of 89 peer leaders were Hispanic, 30 were white, and 82% male. Results: Results show no statistically significant differences by ethnicity or gender. A majority of leaders also state the experience increased their content knowledge (71%) and a smaller fraction said it improved their study skills (56%).
Sandy Irani and Kameryn Denaro. 2020. Incorporating Active Learning Strategies and Instructor Presence into	USA	other computing courses, Discrete	315-381 (intervention) 390 (control)	gender, race. ethnicity, class/SES,	online version vs face to face	student survey(s), performance indicators	US, undergraduate, discrete maths course within CS course. Intervention: Conversion of a face to face (F2F) course to an online version. Participants n = 771 (intervention n=381, control previous year comparison n=390) for course test outcome comparison and general course feedback survey, n=315 for an

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
an Online Discrete Mathematics Class. (SIGCSE '20). https://doi.org/10.1145/3328778.3366904 [55]		maths in a CS course		Other: 1st Generation, transfer students	maths CS course	(e.g., grades, test scores)	additional survey only of the intervention group. The intervention was to transfer to an online version using videos, online quizzes, asynchronous and synchronous discussion. The same course content was delivered across the two versions and using existing tests there was no statistically significant degradation in outcomes for students including looking at this by low income, gender, first-generation, underrepresented minorities. Standard Student surveys similarly were high for both versions of the course. The one-off survey revealed that students were overall very positive and reported they felt they had an opportunity to have questions answered.
Alan C. Jamieson, Lindsay H. Jamieson, and Angela C. Johnson. 2012. Application of non-programming focused treisman-style workshops in introductory computer science. SIGCSE '12). https://doi.org/10.1145/2157136.2157219 [57]	USA	intro computing courses (CS0, CS1, CS2), workshop	145 CS1 students (15 in workshop)	gender, race. ethnicity	1-credit enrichment workshop for CS1 students	performance indicators (e.g., grades, test scores)	Context: This paper evaluates a pilot project on the user of terisman-style workshops in CS1. These workshops attempt to focus more on group work and community building than just technical laboratory style workshops. Population: 145 CS1 North American students (15 in workshop) Method: Students attend 2hr 2x/week workshops that focus on abstract ideas rather than implementation, as well as topics such as oral reports on computer history (with a focus on female innovators). The topics are generally more advanced than material covered in CS1, but the focus is on teamwork and grading is based on participation and effort rather than completion. Attitude surveys are given pre-post participation in workshops. Grades are also tracked. Result: No change was detected in attitudes or grade scores. Higher percentage of female students who attended the course went on to CS2 than those not in the course. (although since program was optional, this may just be due to the fact that more students who attended program went on to CS2 overall)
David S. Janzen, Sara Bahrami, Bruno C. da Silva, and Davide Falessi. 2018. A Reflection on Diversity and Inclusivity Efforts in a Software Engineering Program. In 2018 IEEE Frontiers in Education Conference. https://doi.org/10.1109/FIE.2018.8658677 [58]		other computing courses		gender		student survey(s)	Context:The paper details a number of efforts at improving gender diversity in an undergraduate software engineering program. Several initiatives are described, including an applied focused CS1 course where students could choose their domain (gaming, mobile, music, robotics); a female focused student club (WISH); sponsoring travel to the Grace Hopper celebration; a number of outreach events; and a capstone project course. The paper analyzes the relative perceived benefit of these initiatives, as well as a more in-depth analysis of the capstone course by gender. Population: 118 students (exp1) + 29 teams of 5-6 students (145-174 students) (exp2), undergraduate north american graduating students Method: experiment 1: graduating students were asked to provide likert scale answers on wether each initiative (topic based cs1, grace hopper conference, capstone course, outreach program, student club) positively impacted their retention in the program

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							experiment 2: post capstone course students were asked a series of questions to assess their enjoyment of the course, data aggregated by gender Result: Neither experiment showed any significant effects by gender (though they did note that anecdotally retention of female students did improve)
Lindsay Jarratt, Nicholas A. Bowman, K.C. Culver, and Alberto Maria Segre. 2019. A Large-Scale Experimental Study of Gender and Pair Composition in Pair Programming. ITiCSE '19. https://doi.org/10.1145/3304221.3319782 [59]	USA	intro computing courses (CS0, CS1, CS2)	969 students	gender	having a female partner in pair programming	student survey(s), performance indicators (e.g., grades, test scores)	Context: US Undergraduate CS Introductory modules. Intervention Pair programming composition comparing 3 random pairings per student in a semester. Participants: 969 (41% female, 24% CS majors). Results: Having a female partner increased 1. greater lab attendance 2. greater confidence in the finished product 3. greater confidence in the assignment solution. Advantages of having a female partner were occasionally greater for females than males and greater for CS majors.)
Wei Jin, Cynthia L. Johnson, and Sonal Dekhane. 2020. A Guided Inquiry Approach for Detecting and Developing Problem-solving Strategies for Novice Programming Students. ACM Southeast Conference 2020. https://doi.org/10.1145/3374135.3385289 [60]		intro computing courses (CS0, CS1, CS2)		gender, race. ethnicity		student survey(s), observations of students or classrooms	NO NCWIT engagement practice ticked as the intervention was to introduce a pedagogy that was not particularly collaborative, but was well structured and highly scaffolded. US undergraduate Introduction to programming. Intervention: Small study introducing an interactive highly scaffolded and guided problem-solving strategy (similar to POGIL) measured with a pre-and post-test and survey. The pedagogy is 1. analyse example programs and problem-solving strategies 2. Summarise findings and strategies 3. Apply to solve other similar problems. Participants: Intervention n=33, control n=46. Results: Those using intervention have a statistically significant higher performance on problem-solving using loops, (especially for females) compared to the control group. Intervention students also reported higher motivation and confidence.
Karen A. Kim, Amy J. Fann, and Kimberly O. Misa-Escalante. 2011. Engaging Women in Computer Science and Engineering: Promising Practices for Promoting Gender Equity in Undergraduate Research Experiences. ACM TOCE (July 2011). https://doi.org/10.1145/1993069.1993072 [62]	USA	research experiences	117 (survey) 96 student participants (intervention) 20 educator interview (intervention) no control	gender	Research Experiences for undergraduates(REU). Research Process. Rich Faculty experience. Cross subject, real world.	student survey(s), student focus groups, faculty surveys, faculty interviews	CONTEXT: This study sought to examine elements of REUs (research experiences for undergraduates) that might support women in pursuing next steps such as advanced degrees and careers in computing. The research questions are: 1) Do REU programs in CSE offer specific mentoring, programming, or professional development components addressing gender equity issues? If so, what do these look like? and 2) What can we understand about the specific mechanisms and components of REU programs that benefit female students? POPULATION: US-based; NSF REU program directors (117 survey responses, 20 interviews) and 96 students who participated in an REU at a large research university over 4 years METHOD: Mixed methods approach. Online survey was used to collect data from program directors of NSF-funded REUs in engineering, CS, and related fields. Follow up interviews were conducted with a subset of those who either offered

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							content around gender equity or had a goal of increasing the number of women in PhD programs. Student surveys and focus groups were conducted at a single REU site over four years. RESULTS: Findings from program directors included these strategies: — having a critical mass of women participants in the program (According to the national data, REU participation was higher overall than the percent of women enrolled in undergrad computing programs) — providing role models and mentors, which supports women's participation and long-term interest in the field (the lack of women faculty was noted as a barrier by program directors) — introducing gender equity topics indirectly, in part by discussing "safe" topics (i.e., work/family balance issues, introducing the topics of "implicit bias" and "stereotype threat") that relate to both men and women (This evolved as there was backlash from approaching it directly due in part to gender blindness/consciousness on the part of both male and female undergraduates) — a rich, hands-on research experience under faculty mentorship for both female and male interns — fostering research competence and community building through structured groups of undergraduates working with graduate students and faculty — weekly presentation meetings where individual progress is shared
Alyona Koulanova, Ary Maharaj, Brian Harrington, and Jessica Dere. 2018. Fit-breaks: incorporating physical activity breaks in introductory CS lectures. ITiCSE 2018. https://doi.org/10.1145/3197091.3197115 [64]		lecture		gender, international	incorporation of physical activity into lecture breaks	student survey(s), performance indicators (e.g., grades, test scores)	Context: multiple CS1 sections at a Canadian institution (University of Toronto Scarborough) Intervention: During the ten-minute intermission in lecture, students were led through "Fit-Breaks": short bursts of easy-to-follow exercises and stretches. Music was played during lecture break for the control section. Participants: 750 students in all five CS1 sections, but only three sections had intervention. Method: Students completed weekly online questionnaires Results: Levels of break enjoyment were significantly higher in the Fit-Break group. While the sample size was too small to provide statistical significance, the data indicates that theFit-Break activities had a positive impact on several important areas of course perception, particularly among female students. In particular, we found that female students in the Fit-Break group showed improvements in their perceptions of sense of control, enjoyment and perceived usefulness: three areas that are traditionally problematic for female students in CS. Similarly, females in Fit-Break group had higher but not statistically significant

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							difference in course grades (although there was a statistically significant difference for international students).
Julie Krause, Irene Polycarpou, and Keith Hellman. 2012. Exploring formal learning groups and their impact on recruitment of women in undergraduate CS. SIGCSE '12. https://doi.org/10.1145/2157136.2157192 [65]	USA	intro computing courses (CS0, CS1, CS2)	76 students in survey and 22 in interviews /focus groups	gender	formal learning groups	student survey(s), student interviews	Context: US Undergraduates Introduction to CS. Participants 76 students (15 female) completed pre and post-course surveys and 22 students (12 female) took part in interviews and focus groups. Method Pre and a post-course survey asking about intent to study CS, attitude to formal learning groups, ranking of learning groups compared to other instructional approaches in terms of what they learned most. Plus interviews and focus groups. Only descriptive stats used for comparison. Ranking compared learning groups to reading books, observing lectures, programming in python, and watching video - but reporting of the ranking data is very limited. No qualitative methodology mentioned regarding data analysis. Intervention: Formal Learning Groups Results: Very difficult to assess the results as only descriptive data and no qualitative method discussed with regard to reliability and validity of data analysis. The changes between the pre and post survey questions on how fun students thought learning groups would be/were was minimal and had little gender difference. Similarly little difference in makes and females with regard to impact on intention to Major in CS (numbers are very small). The most important finding perhaps is that most reported that Learning Groups did not impact their intent to pursue CS further, with little gender difference. However, the paper concludes that "The results of our study showed that many female students who took the course reported that they learned from and enjoyed formal learning groups" Reading textbooks was reported as being more effective for learning than Learning Groups by most. And I think Learning Groups was ranked higher than most for other approaches. Positive responses to Learning Groups rather than other approaches is summarised as 1. Feeling more comfortable participating in a small group than in front of a whole class. 2. Experiencing additional motivation from having an impact on others 3. Gaining additional interactive instruction Negative responses 1. Having the potential for learning incorrectly 2 Possibility of having unproductive groups

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
Sophia Krause-Levy, Leo Porter, Beth Simon, and Christine Alvarado. 2020. Investigating the Impact of Employing Multiple Interventions in a CS1 Course. SIGCSE '20. https://doi.org/10.1145/3328778.3366866 [66]	USA	intro computing courses (CS0, CS1, CS2)		gender	Mindset interventions, thinkathons, metacognitive tasks	student survey(s), performance indicators (e.g., grades, test scores)	Context: This study seeks to explore the following research question: Can combining previously identified promising intervention techniques within a standard structure of a CS1 course significantly increase students' performance and what value, if any, do students perceive from these interventions? This model of the course was compared to a model that did not include these interventions. Population: CS1 students in US. 222 students in total - 114 in the experimental group, 108 in the control group. Method: The intervention techniques implemented were: 1) mindset interventions; 2) thinkathons, where on-computer coding exercises were replaced with paper- based scaffolded code comprehension and writing exercises (following the Thinkathon model) targeting the same learning outcomes; 3) metacognitive tasks were used to support study skill development and inclusion and values reflections to foster growth mindset and belonging. In particular, values interventions and study skills interventions were used. Results: Mostly, the results indicated that the interventions did not make a difference, and there weren't sex-based differences. The exception is in exam performance, where the female students showed a small but statistically significant improvement (7%) whilst the male students showed a small and not statistically significant decline (7%). The authors conclude that the paper does not provide sufficient evidence to determine the truth of their assertion that these intervention would improve performance but that it provided a basis for further investigation, especially around how it could potentially improve the performance of female students.
Anagha Kulkarni, Ilmi Yoon, Pleuni S. Pennings, Kazunori Okada, and Carmen Domingo. 2018. Promoting diversity in computing. ITiCSE '2018. https://doi.org/10.1145/3197091. 3197145 [67]	USA	intro computing courses (CS0, CS1, CS2)	60(intervention) ?? (control)	gender, race. ethnicity	(PINC) cohort based, targeted recruitment, mentoring, near peer - peer-led (PLTL and ARG programs), project based, cooperative learning, specific	student survey(s), number of students remaining in the course	The paper reports on a pilot program (PINC) at San Fransisco State University that aimed to promote inclusivity in computing. PINC aimed to bring computer science to students majoring in other fields, instead of expecting them to change their career course to CS. The researchers recruited PINC students from majors such as Biology, and Biochemistry which enjoy fairly high enrolments by underrepresented minority groups and women. The program included cohort-based program structure, near-peer mentoring, project-driven and cooperative learning, to attract, retain, and successfully graduate a highly diverse and interdisciplinary student body. In total 60 students from other disciplines participated in this program. The findings of the administered survey showed that the majority of the students feel less intimidated about computer programming, and understand its utility and necessity. Compared to a traditional CS course in the same university, the composition of PINC program had an entirely different gender distribution with 73% of students being female students compared to the corresponding 16% in the CS traditional course. Moreover, the results demonstrate that while in the traditional CS course the number of female students decreased by 15 percentage points by the end of the

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
					curriculum, cross curricula		program, for the PINC program, the drop in female students (2 percentage points) was fairly small. Similar trends were observed in the numbers of URM students. In the traditional CS bachelors program, the starting population of URMs is already small (34%) and it drops to 30% whereas in the PINC program, the URM students account for 56% of the initial enrollment and this number persists till the final courses of the program.
Celine Latulipe, Audrey Rorrer, and Bruce Long. 2018. Longitudinal Data on Flipped Class Effects on Performance in CS1 and Retention after CS1. SIGCSE '18. https://doi.org/10.1145/3159450.3159518 [69]	USA	intro computing courses (CS0, CS1, CS2)	70M 29F (intervention) 530M 69F (control)	gender, race. ethnicity	Flipped lecture + lab	performance indicators (e.g., grades, test scores)	Context: US Undergraduate, CS1, 2013 to 2016. Intervention: Flipped. Method Comparing male/ female, underrepresented groups retention if they experienced a flipped or traditional lecture/ lab versions. Participants: n=599 traditional/control (female = 69) , n= 99 flipped/intervention(female =29)) Results: Women were less likely to switch majors if they attended the flipped intervention. Males were less likely to switch majors if they attended the traditional lessons. Performance in the flipped class was higher. One year retention in the major for under-represented groups (women and racial minorities) was higher in flipped but not for transfer students.
Celine Latulipe, N. Bruce Long, and Carlos E. Seminario. 2015. Structuring Flipped Classes with Lightweight Teams and Gamification. SIGCSE '15. https: //doi.org/10.1145/2676723.2677240 [68]	USA	intro computing courses (CS0, CS1, CS2)	"N=45 and 45, response rates of 52%) and Spring 2014 (N=59 and 44, response rates of 94% and 70%) semesters ."	gender	Gamification	student survey(s), performance indicators (e.g., grades, test scores)	Context: This work investigates the use of 'lightweight teams', where students work together in class but where the work done together does not much impact students' final grades. They also used gamification and a flipped classroom. They hypothesise that lightweight teams enhance students' ability to learn and make the course more engaging, that gamification elements encourage students to work hard and make the course more engaging, and that this whole approach makes students more socially integrated with their peers. Population: CS1 students in a US university. 92 took the relevant course in the first semester (fall 2013) and 65 in the second (spring 2014) Method: The organisers taught the course using the three above interventions and conducted surveys twice per semester (towards the beginning and towards the end) to get students' views on the approach. The early one had 44 questions; the latter one 50. The first semester had responses from 52% of students each time; the latter had 94% and 70% respectively, because students were given time to do surveys in class. They also put women together in pair programming rather than having mixed-sex groups. Results: The response of all students to the interventions were positive, especially for the lightweight teams. Responses were more positive for most questions in the later surveys, once the students had spent longer doing them. Mostly, there were no statistical difference between male and female students. The exception was a question about collecting stamps (part of the gamification, which was designed to encourage team competition rather than individual competition), for which female students gave a more positive response in the fall survey, though in the spring survey there wasn't a significant difference. Students who had done this course did better in follow-up courses to students

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							who had done more traditional CS1 courses, but this data was not desegregated by sex. Female students generally agreed that gendered pairs for pair programming was a good policy, though some disagreed with this.
Elynn Lee, Victoria Shan, Bradley Beth, and Calvin Lin. 2014. A structured approach to teaching recursion using cargo-bot. ICER '14. https: //doi.org/10.1145/2632320. 2632356 [70]	USA	intro computing courses (CS0, CS1, CS2)		gender	Combining video games with lectures	student survey(s), performance indicators (e.g., grades, test scores)	The paper presents a structured approach to teaching recursion at the University of Texas that combines classroom lectures and self-paced interaction with Cargo-Bot, a video game in which users solve puzzles using a simple visual programming language. The authors devised a lesson plan that uses Cargo-Bot game playing to scaffold key concepts used in writing recursive Java programs. They empirically evaluated their approach using 204 undergraduates enrolled in a CS2 course (88 students in the experimental group and 116 in the control group); The results demonstrate strong statistical evidence on improving student learning of recursion over traditional lecture-based instruction alone but gender and race/ethnicity were not significant factors for gains in total, across the entire group.
Colleen M. Lewis, Nathaniel Titterton, and Michael Clancy. 2012. Using collaboration to overcome disparities in Java experience. ICER '12. https://doi.org/10.1145/2361276.2361292 [71]	USA	intro computing courses (CS0, CS1, CS2)	355 CS2 students across 2 years (49 Female, 306Male)	gender	group projects, pair programming	performance indicators (e.g., grades, test scores)	Context:This paper sets out to overcome the discrepancy caused by previous experience in a CS2 course by encouraging student collaboration. Population: 355 North American CS2 students across 2 years (49Female, 306Male) Method: Course was shifted to highly collaborative model with all else being held similar: individual/paired projects were replaced with group projects (static group of 3-4), emphasis put on pair programming in labs, homework added an option for partnered submission. Students' gender, group work preference, and prior experience were self reported at the beginning of the term, and correlated with course outcomes. Result: Fewer female students dropped (5% vs 37%)the highly collaborative model (not statistically significant, due to small numbers). No claims could be made about grades with respect to gender for the two class models, authors suggest data is clouded by students who would have dropped in non-collaborative model, stayed on but struggled in highly collaborative model
Phil Maguire, Rebecca Maguire, Philip Hyland, and Patrick Marshall. 2014. Enhancing Collaborative Learning Using Pair Programming: Who Benefits? AISHE-J: The All Ireland Journal of Teaching and Learning in Higher	Ireland	intro computing courses (CS0, CS1, CS2),lab	24 (survey) 99 (exam grades)	gender	Pair programming in weekly labs	student survey(s), performance indicators (e.g., grades, test scores)	Context: Ireland introductory programming Intervention: Pair programming, students randomly allocated on a week-by-week basis to work in labs. Rather than sharing a workstation, students had one workstation each - they only worked together on the design and to discuss errors and bugs. Each student did their own implementation. Participants: 19F 80M for exam results, 10F 30M for surveys

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
Education (2014). [72]							Method: within the same cohort the first semester was taught without intervention, second semester with intervention. Differences between the exam scores and survey results pre- and post-intervention were compared. Results. No significant difference between the genders on the main measures but 2-way ANOVA indicated females' exam scored improved more following the intervention ($p=0.14$). Surveys showed a range of significant gender differences, particularly 'I learned programming skills from being paired'
David J. Malan. 2021. Toward an Ungraded CS50. SIGCSE 2021. https://doi.org/10.1145/3408877.3432461 [73]		intro computing courses (CS0, CS1, CS2),lab		gender		student survey(s), number enrolled	I have not assigned a NCWIT Engagement Practices as there was not a matching item. Context: US undergraduate CS50 (Introduction to Computer Science) Intervention: Long term initiative since 2010 to introduce the option for students to be graded pass or fail (Satisfactory/Unsatisfactory) or by letter grades. Participants: 671 students in 2017. Results: Increase in the % of women (up 8% from 36% to 44%) taking the course and more students reported they might take a CS concentration.
Lauren E. Margulieux, Briana B. Morrison, and Adrienne Decker. 2020. Reducing withdrawal and failure rates in introductory programming with subgoal labeled worked examples. International Journal of STEM Education (Dec. 2020). https://doi.org/10.1186/s40594-020-00222-7 [75]	USA	intro computing courses (CS0, CS1, CS2)	0	gender, race. ethnicity	Sub-goal labeled worked examples	performance indicators (e.g., grades, test scores)	Context: Subgoal labelling breaks problems down into smaller steps in a formal way to help novices learn how to tackle smaller problems and build towards larger. Has been shown to help in small controlled experiments, this paper set out to test the idea at scale in a semester long course. Population: 265 North American undergrad CS1 students Method: Course was split into 5 sections, 3 of which used normal worked examples, 2 used subgoal labeled worked examples Result: No correlation between gender and performance across groups
Monica M. McGill. 2012. Learning to Program with Personal Robots: Influences on Student Motivation. ACM TOCE (March 2012). https://doi.org/10.1145/2133797.2133801 [77]	USA	intro computing courses (CS0, CS1, CS2)	35	gender	Robotics in programmi ng	student survey(s)	Context: This paper analyzed the introduction of robotic programming into a CS0 course, and its impact on motivation Population: 35 North American CS0 students (13 female and 22 male) Method: Robotics programming was introduced into a CS0 course, pre + post intervention, Keller's Instructional Materials Motivation Survey was administered to students to measure motivation, the surveys also included specific questions about the robots Result: Female students responded more strongly that robotics was intimidating, and also that they were useful in learning. Female students agreed that working with robots was "a pleasure" while male students mostly disagreed. No statistically significant relationships between gender and motivation factors was found

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
Paola Medel and Vahab Pournaghshband. 2017. Eliminating Gender Bias in Computer Science Education Materials. SIGCSE ’17. https://doi.org/10.1145/3017680.3017794 [78]		other computing courses		gender	Removal of gendered examples in textbooks	student survey(s)	Context: Most CS textbooks use gendered language (e.g., "Alice is working on a problem" or "she is struggling with a problem"). The study looked at a number of textbooks and found that women were more likely to be portrayed as negatively. And then attempted to analyze the impact of replacing human gendered characters with engendered animal characters with they pronouns. The study presented is only a pilot study to explore the feasibility of the material. Population: 69 north american undergraduate students across 2 cryptography courses Method: control group given normal texts, experimental group given gender neutral texts with human characters replaced by animals and he/she pronouns replaced by they. Post course survey studied confidence of understanding Result: Female students showed an improvement in confidence, with no change in male confidence (no statistical information given)
Danaë Metaxa-Kakavouli, KellyWang, James A. Landay, and Jeff Hancock. 2018. Gender-Inclusive Design: Sense of Belonging and Bias in Web Interfaces. CHI ’18. https://doi.org/10.1145/3173574.3174188 [79]		intro computing courses (CS0, CS1, CS2),Other	Comparison of introductory CS course webpages	gender	modification of web resource aesthetics	Crowdworkers aged 18-25 survey	US, Undergraduate, introductory CS course MATERIAL - not the course. Intervention: Comparison of 2 identical content webpages with different aesthetic features such that one was perceived as masculine while the other was gender-neutral comparing impact on attitudes of those who used it. Participants: 60 crowd workers (aged 18-25) pretest, 109 crowd workers (aged 18-25) main study. Results: Women were unconsciously impacted with respect to attitudes and intention negatively in the masculine condition, this was not the case for men. Women were statistically significantly less likely (compared to men) to want to enrol on the course, were less likely to feel they would belong, anticipated lesser success, lower levels of confidence in technical ability, less interest in learning cs or programming in the long term and were more likely about the way their gender would be perceived. For the ambient version, there was not a statistically significant difference between male and female attitudes. Comparing between the masculine and ambient pages. Men did not perceive them as being more or less masculine, whereas women did. Also men, were less positive about CS also using the more masculine version, but this was not a statistically significant difference.
Sathya Narayanan, Kathryn Cunningham, Sonia Arteaga, William J. Welch, Leslie Maxwell, Zechariah Chawinga, and Bude Su. 2018. Upward Mobility for Underrepresented Students: A Model for a Cohort-Based Bachelor’s Degree in Computer Science. SIGCSE ’18.	USA	intro computing courses (CS0, CS1, CS2),other computing courses,homework or other out of class	This is a strategic partnership between a community college and a university and	gender, class/SES, first generation	(CSin3) Transition intervention, accelerated degree, chohort community, predefined module	performance indicators (e.g., grades, test scores)	Country - US. Level - CS Degree program. Intervention - Strategic partnership between a community college and a public university includes degree split as half at community college, half at university, pre-defined transfer pathway, a math-intensive preparatory program and summer bridge program for college readiness, mandatory study time on campus, academic support from two dedicated support coordinators, closely organised tutoring from trained peers, cohort grouping of underrepresented groups, enrichment sessions e.g. on cohort culture, team building, programming fundamentals, test techniques, sample project work, an internship, career development with internship and job preparation workshops, student funding (scholarships) so that students don't need to work during study. *** With relationship to what can a CS lecturer do -

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
https://doi.org/10.1145/3159450.3159551 [82]		activities,Other	includes a whole host of extra support		pathway, just in time academic and admin support, peer-led team learning, tutors, financial aid, focus on skills, internship		pedagogy and content wise there seems to be 2 things 1. Adding in peer wise tutoring (section 4.11) .2. Grouping students with other students from a similar background (section 4.4.) . About peer wise tutoring. Students in later years of the course are trained on tutoring and PCK each week by a coordinator and guide first year students through problem sets for 90 mins to 120 mins each week. This is called a "more robust academic framework of foundational programming skills". However it is impossible to tell whether which of the many interventions or combinations thereof have the most (or any) impact. Number of participants. There were 102 students who started the program. But it is not clear how many graduated it seems to be 46. But is reported as 79%. Results: There are stats related to an increase in representation of Hispanic and female students but this is shown as a ratio change so absolute values are not clear. The ratio change is across the whole of the university CS cohort which includes the intervention group. Change is from 1/34 to 1/20 Hispanics and 1/30 to 1/13 female. There are also stats from standardised exams. 32 intervention students and 26 non intervention took a CLA (general critical thinking) before and after their university course and students in the intervention group made more progress, but it is hard to understand the scale of the difference. Students also sat a standardised CS test and there seems to be little difference in between the intervention and non intervention group, but the sample is small and the protocol for inclusion is not stated. Threats to validity: The population may not be representative as many students had GPA's where they could have gone to universities that were more highly ranked. This is an interesting paper with ideas for interventions such as more contact with faculty and peer tutoring but one cannot be sure of the impact of these interventions compared to the overall program.
Tia Newhall, Lisa Meeden, Andrew Danner, Ameet Soni, Frances Ruiz, and Richard Wicentowski. 2014. A support program for introductory CS courses that improves student performance and retains students from underrepresented groups. SIGCSE '14. https://doi.org/10.1145/2538862.2538923 [84]	USA	intro computing courses (CS0, CS1, CS2)	graphs but no exact numbers	gender, race. ethnicity	attending help sessions with student mentors	student survey(s), performance indicators (e.g., grades, test scores), enrollment	Country: USA Level: Undergraduate, CS1/CS2 Intervention: Mentoring program for CS1 Results: Increase in female and racial URM enrollment, qualitative insights into the impacts of mentoring HH note: CS1 class also changed programming languages, which may have impacted enrollment
Keith J. O'Hara, Kathleen Burke, Diana Ruggiero, and Sven Anderson. 2017. Linking Language /&		other computing courses		gender		student survey(s)	The paper describes the design, implementation and preliminary evaluation of a computing component for a three-week writing-intensive introductory program at a liberal arts college. The primary objective of the component was to expose all first-year students to the power of code as language and computational

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
Thinking with Code: Computing within a Writing-Intensive Introduction to the Liberal Arts. ITiCSE ’17. https://doi.org/10.1145/3059009.3059018 [86]							thinking, and positively change their attitudes about such matters. The design of the course included the following characteristics: non-expert-led instruction, computing unplugged, quick start and rapid feedback, context relevant to students’ interests, close readings consistent with the culture of the program. An analysis of attitude surveys showed that the program positively changed students attitudes about enjoyment solving CS problems, but potentially reinforced the misconception that CS is just learning programming languages. Particularly, female students demonstrated a greater appreciation of computation which suggests a significant shift towards a more positive view of computer science as an enjoyable activity. On the other hand, the results also suggest that the computing program did not help students see a relationship between computer science and their own interests. The authors believe that the increased identification of computer science with programming may have been an unintended consequence.
Jennifer Parham-Mocello, Martin Erwig, and Emily Dominguez. 2019. To Code or Not to Code? Programming in Introductory CS Courses. VL/HCC 2019. https://doi.org/10.1109/VLHCC.2019.8818909 [89]				gender		student survey(s)	Context: The paper details and evaluates "delayed" and "coding free" CS0 where students focus on unplugged activities and story-based algorithm learning approaches for the first half of the term and the entire term respectively Population: 147-191 north american undergraduate CS1 students Method: Students could self select between traditional "coding first" CS0, "delayed coding" CS0 (only non-coding story based algorithms for first 5 weeks) and "coding free" CS0 (only non-coding story based algorithms for entire 10 weeks). Interest level was surveyed before and after the class. Result: Female students are more interested in the non-coding version of the course. No mention of difference in post course survey.
Jennifer Parham-Mocello, Shannon Ernst, Martin Erwig, Lily Shellhammer, and Emily Dominguez. 2019. Story Programming: Explaining Computer Science Before Coding. SIGCSE ’19. https://doi.org/10.1145/3287324.3287397 [88]	USA	intro computing courses (CS0, CS1, CS2)		gender	Story programming	student survey(s), performance indicators (e.g., grades, test scores)	CONTEXT: This study re-designed an introductory orientation course (for students in the major) to utilize story programming as an approach for developing computational thinking skills before introducing students to programming. A comparison of different programming languages in each section was also conducted. POPULATION: CS0 students in USA at a state university. 277 students total: 107 students in control group, 65 students in experimental group using Haskell as the programming language and 105 students in experimental group using Python. 110 consented to the survey. METHOD: Experimental design using random assignment of students to three sections of the course taught by the same instructor. Course sections included: traditional, story programming with python, and story programming with Haskell. Student DFW rates and grades were collected for each section and surveys on attitudes towards the computational thinking and coding activities

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							RESULTS: The traditional section had a higher percentage of male students. No differences by gender in DFW rates, grade distribution, or in the ratings of the CT or coding activities is seen by course section. However, when females are considered across all sections, there is a significant difference (pvalue=.01) in engagement of coding activities, rating them higher than male peers. (p.384) Overall, the choice of language had greater impact on student outcomes than the story programming vs traditional approach; students using Haskell had stronger positive reactions to the CT activities than students using python.
Nayna Patel, Willem-Paul Brinkman, and Jane Coughlan. 2012. Work placements and academic achievement: undergraduate computing students. Education & Training (2012). https://doi.org/10.1108/00400911211254299 [90]	UK	Work placements		gender, class/SES	Work placements	performance indicators (e.g., grades, test scores)	This is not an intervention this is a strategic approach of including work placements in CS courses. Therefore it is hard to say that a specific NCWIT practice is engaged other than the placement is a relevant and meaningful context. Context: UK university CS undergraduate data analysis. Participants: 290 student profiles analysed. Method: Data analysis comparing students who do or do not do a placement and academic performance. Intervention: Doing a placement. Results: Students who did a placement achieved a higher grade to their overall degree than those who did not. The results show that the increase was not because students with higher academic performance take up placements. Gender had no impact on the findings.
Jon K. Piper and Dwight Krehbiel. 2015. Increasing STEM Enrollment Using Targeted Scholarships and an Interdisciplinary Seminar for First- and Second-Year College Students. Journal of STEM Education: Innovations and Research (Dec. 2015). http://www.proquest.com/docview/1767144450/abstract/5BB8B973D9046FEPQ/1 [91]	USA	interdisciplinary learning community/ course	150 STEM students but it is not clear how many are CS (intervention) no control	gender, race. ethnicity	STEM learning community	student survey(s)	Note: some of the outcome data (surveys and upper-level seminar course enrollments) is disaggregated by gender and/or minority status but is NOT disaggregated by major, and so includes non-CS majors of biology, chemistry, math, physics, and psychology. CONTEXT: The authors studied the impact of a STEM scholarship program for first and second year undergraduates POPULATION: small private liberal arts college in the midwest (US), about ~150 students METHOD: Developed al scholarship program and a year-long course -- an interdisciplinary STEM Learning Community. The course is a requirement for freshman and sophomore students receiving a scholarship and includes readings on ethics, conferences, exploration of STEM careers and internships, and a collaborative research project. RESULTS:

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							Student end-of-course evaluations suggest that students found the activities useful and that it improved their understanding of STEM skills, careers, and other areas needed to be successful in a STEM major and career. The enrollment numbers in the upper-level (junior and senior) courses of female and minority students increased.
Leo Porter and Beth Simon. 2013. Retaining nearly one-third more majors with a trio of instructional best practices in CS1. SIGCSE '13. https://doi.org/10. 1145/2445196.2445248 [92]	USA	intro computing courses (CS0, CS1, CS2)	2067 - during interventi on, 1231 - after interventi on	gender	pair programmi ng, peer instruction	performance indicators (e.g., grades, test scores)	Porter's and Simon's (2013) study describes the effects of an intervention conducted in a US university CS1 course aiming to contribute to computer science students' retention in computing. Their intervention included an instructional design informed by three practices: Media Computation, Pair Programming, and Peer Instruction. The authors collected and compared students' data from before and after the intervention. In total, 2067 student data were collected from the years before the intervention (2001-spring 2008), and 1231 student data were collected from the years after the intervention (fall 2008-2011). The results of their study demonstrate that when only those students who passed the course are considered, retention was increased for both genders, but it was only statistically significant for males. However, retention of CS majors originally enrolled in CS1 at week 4 (the deadline to withdraw) was statistically and significantly improved for both genders.
Cyndi Rader, Doug Hakkarinen, Barbara M. Moskal, and Keith Hellman. 2011. Exploring the appeal of socially relevant computing: are students interested in socially relevant problems?. SIGCSE '11. https://doi.org/10.1145/19 53163.1953288 [93]		intro computing courses (CS0, CS1, CS2), other computing courses		gender		student survey(s)	Context: Embed humanitarian assignments into CS curricula into CS1 and Software Engineering courses at US institution. Population: 71 CS1 students (first year students)and 26 software engineering (fourth year students) Method: Student surveys to assess which assignments students found appealing. Result: Women rate the humanitarian projects higher than male students, but the highest ratings went to the projects they found to be personally relevant (e.g. calorie counter) and the fun ones (e.g. Tic-tac-toe)
Sagar Raina, Blair Taylor, and Siddharth Kaza. 2015. Security Injections 2.0: Increasing Engagement and Faculty Adoption Using Enhanced Secure Coding Modules for Lower-Level Programming Courses. In Information Security Education Across the Curriculum (IFIP Advances in Information and Communication Technology).	USA	intro computing courses (CS0, CS1, CS2)	80 students	gender, race. ethnicity	0	student survey(s)	Please note for NCWIT it was not collaborative learning but it was well structured. Also, the assessment included automated feedback and hints but it is not wholly student centred. The main change is more carefully scaffolded teaching, breaking down the material and forcing active learning through questioning. "Context: A CS0 web-based cybersecurity course using Security Injections. Population: 80 CS0 students in a US university, 42 in the treatment group 38 in the control group. Method: Post-intervention student engagement survey comparing the original (control) and new version (treatment) of three modules of a course. The survey was an adapted well-tested instrument (User Engagement Scale). Instructor

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
https://doi.org/10.1007/978-3-319-18500-2_6 [95]							feedback was also gathered, but the protocol for collection and analysis is not clear. Intervention: Enhancement of 3 modules (integer error, input validation, buffer overflow) of a web-based cybersecurity course incorporating e-learning design principles of segmentation (to reduce skipping content) and interactivity (to improve outcomes and engagement). Techniques used included 1. breaking up content 2. adding auto-marked questions (true/false, MCQ or constructed response) 3. questions can be attempted many times by the students 4. students only progress if the answer is correct 5. if the answer is incorrect, knowledge-based feedback (hint) is provided until the third attempt and "elaboration" feedback after. 6. a new instructor dashboard of question attempts Results: An analysis of the the student engagement surveys revealed that students, irrespective of gender and ethnicity, reported being more engaged with the new version of the course. Females showed a greater increase in reported engagement than males. Instructor feedback is cited as positive about the changes and the approach has been adopted for both in-class and online classes for approx 180 students by 5 instructors in the past 5 years (since 2015).
Yolanda A. Rankin, Jakita O. Thomas, and India Irish. 2019. Food for Thought: Supporting African American Women's Computational Algorithmic Thinking in an Intro CS Course. SIGCSE '19. https://doi.org/10.1145/3287324.3287484 [97]	USA	intro computing courses (CS0, CS1, CS2)	35 African-American students over 2 cohorts	gender, race. ethnicity	Introduction of food-focused activities	student journals	Context: Introductory CS course at HBCU. The use of food-focused activities to improve Computational Algorithmic Thinking (CAT). What are the perceptions of student about the activity (Dessert Wars Challenge) and how does participation support developing CAT capabilities Population: 35 African American women. 24 in 2015 and 11 in 2016. Method: Self-reflection journals with guided prompts. Content analysis was performed on the full set of responses and themes were then categorized into easy, difficult and suggestions. Students in the 2016 section were also given a test to evaluate CAT skills. Result: The intervention provided motivation and situated context for students to develop their CAT skills. High retention rate of students 96 and 100%. In the course taught the following year 2017 without the food-related activities, decrease retention rate of 79%. They posit that this was because the first part of the course focused on learning Python and not understanding algorithms.
Anja Remshagen and Christine Rolka. 2014. Contextualized learning tools: animations and robots. ACM Southeast Regional Conference '14.	USA	intro computing courses (CS0, CS1, CS2)	Not specified	gender	Animation and robotics programming	student survey(s), performance indicators (e.g., grades, test scores),	Context: Much effort had been put into making the course more accessible, including the use of animation to teach programming, but 50% of students were still failing. They therefore tried using robotics in addition to the animation, as a way to motivate students, especially female students. Population: CS0 students at a US university, mostly with low interest in CS. Not

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
https://doi.org/10.1145/2638404.2638458 [99]						faculty interviews	clear how many students were involved, but the data covers several years. Method: The robots were used for programming exercises throughout the semester. They introduced the robots in 2011 throughout all sections, and in 2013 switched to having two sections where only animations were used and two were both animations and robots were used (not entirely clear what sections are). Results: Survey feedback from students indicated they liked the robots but preferred the animation (no gender-based differences). Female students slightly more likely to find robot exercises difficult (52% to 45.5%; not clear if this is statistically significant). Prior to the introduction of robotics, there was no statistically significant difference between male and female students, but after robots were introduced, there was, in favour of female students. The semester grade in 2013, where two of four sections had no robotics, showed that female students did slightly worse with robots + animation than with just animation, whilst male students did slightly better, but these were not statistically significant. Overall, the results suggest that the use of robots don't influence student performance overall, but they they do lead to the performance gap between male and female students widening in favour of women.
Penny Rheingans, Anne Brodsky, Jill Scheibler, and Anne Spence. 2011. The Role of Majority Groups in Diversity Programs. ACM TOCE (July 2011). https://doi.org/10.1145/1993069.1993075 [101]	USA	scholarship program	480 (survey)	gender	Community builidng, Mentoring, academic support, leadership training LIVING LEARNING	student survey(s), student focus groups, student interviews	CONTEXT: The CWIT Scholar Program provides scholarships and support resources for undergraduates (both women and men) in computing and other technical programs who are committed to increasing women in technical fields. Program elements include community-building, mentoring, academic support, leadership training, and a living-learning community. Scholars take two gender and women's studies courses as well. A second program on campus (SITE) is similar to CWIT but does not require the GWS courses. Data was reported for both programs. POPULATION: large research university in the US; 480 students completing survey METHOD: Focus groups, interviews, and survey on computing/technical program climate of students both in and not in the CWIT/SITE programs RESULTS: Differences in perceptions of various aspects of climate were reported: -- general climate: men took an individualistic approach to utilizing program resources but also had stronger non-academic community ties to other students in the major) -- predictors of academic outcome: program elements designed to support female students may also support men participating in the scholar program -- gender climate: men in the program were more aware of gender issues such as

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							climate, retention, confidence, perceptions of women's preparation for the major, etc.
Christine Rolka and Anja Remshagen. 2015. Showing Up is Half the Battle: Assessing Different Contextualized Learning Tools to Increase the Performance in Introductory Computer Science Courses. Georgia Educational Researcher (Jan. 2015). https://doi.org/10.20429/ijsotl.2015.090110 [102]		intro computing courses (CS0, CS1, CS2)		gender, race. ethnicity		performance indicators (e.g., grades, test scores)	Context: The main focus of this paper was assessing robots vs animations in the classroom, but they also reported general statistics on gender and attendance Population: 87 North American CS1 students Method: The main study assessed the relative impacts of robotics vs animations, but the result were not reported by gender. However a secondary study correlated gender vs attendance and course grade Result: Female students had overall high grades in the course, but this was found to be due to them having higher attendance [NOTE: It could be argued that the result here would be "female students show up more than male students to an animation or a robot based CS1", but they never really identified that as a conclusion explicitly in the paper]
Audrey Smith Rorrer, Joseph Allen, and Huifang Zuo. 2018. A National Study of Undergraduate Research Experiences in Computing: Implications for Culturally Relevant Pedagogy. SIGCSE '18. https: //doi.org/10.1145/3159450.3159510 [103]	USA	Research skills course	439 (intervention) no control	gender, race. ethnicity	Research Experiences for undergraduates(REU). Research Process. Rich Faculty experience. Cross subject, real world.	student survey(s)	[Note it is not clear what culturally relevant engagement practices were included in each of the institutions that took part, but the paper implies that relevant and meaningful contexts and collaboration are included. There may be others but one needs to look at the source material of the toolkit and discover how it was applied.) Context: US CS undergraduates undertaking a research experience summer course within a large scale long-term study evaluating a Culturally Relevant Pedagogy and toolkit. The study was conducted in the 2 summers of 2015 and 2016. Participants: Although 924 students over 58 sites engaged with the programme , not all responded to the survey therefore the study is of 169 students in 2015 and 270 students in 2016 who completed both the pre and post course surveys for each year Method: Pre and Post course student surveys measuring variables related to research and how this changes over the intervention. The variables were computing self-efficacy, intent toward graduate school, computing attitudes, academic helpseeking/coping skills, research skills, leadership skills, scientific identity, and mentoring. Intervention: The summer schools followed guidance in he Toolkit which contains: evaluation instructional materials, an inventory of assessment tools and techniques employed within the community, a standardized program applicant management tool called the Common Application, and an instrument to measure student outcomes known as the A la

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							Carte Survey. Results: Between the pre and post-course surveys for both years, all participants increased in computing self-efficacy, research skills and scientific leadership. Intent to go to Graduate school, computing attitude and scientific identity only increased for 2016 cohorts. 2015 female participants reported a higher degree of professional identity as a scientist than males and 2016 females reported a higher degree of scientific leadership than males. Ethnic groups also had increased self-efficacy, academic help-seeking/ coping, research skills in 2015 and self-efficacy in 2016 to non-ethnic groups. Effect sizes were small for all increases. How culturally relevant pedagogy was implemented in each site are not stated.
Lauren Russell. 2017. Can learning communities boost success of women and minorities in STEM? Evidence from the Massachusetts Institute of Technology. Economics of Education Review (Dec. 2017). https://doi.org/10.1016/j.econedurev.2017.10.008 [104]	USA	Major/instit ution wide	2011 to 2015 students at MIT various groups for stats	gender, race. ethnicity	Learning community	performance indicators (e.g., grades, test scores), Registrar data	*Note: this is an institutional level intervention, and even then at a very specialized place like MIT. Not saying this isn't important, but I think it is important to emphasize the context of the study and note potential limitations in generalizability. Context: This study investigates the impact of a freshman learning community at MIT (called the Experimental Study Group, or ESG). ESG has a lottery-based admission system. The author uses this to analyze causal effects of enrollment in the ESG on broadening participation across STEM fields. Population: All MIT undergraduates enrolled in 1997, 1999, 2003, 2004, and 2006–2015 Methods: The author uses data from the MIT registrar to obtain academic data as well as whether a student applied for ESG, received admission based on the lottery system, and if they eventually enrolled in ESG. The author uses this data in regression analyzes to estimate the effect of ESG on academic outcomes. Results: While the data and results are MIT-wide, and thus cover all STEM majors, some results are broken down by major (with a note that a student cannot major in CS alone at MIT, but must also major in Electrical Engineering or Math). The overall findings indicate positive outcomes for female and underrepresented minority students enrolled in ESG, indicating learning communities can support these students. It's also important to note, as the author does, that a learning community is a confluence of variables (small class sizes, unique curricula, advising/mentoring, and exposure to predominantly female instructors), so it's difficult to point to just one treatment and gauge it's effectiveness.
Adrian Salguero, Julian McAuley, Beth Simon, and Leo Porter. 2020. A Longitudinal Evaluation of a	USA	intro computing courses (CS0, CS1, CS2)	1510+1365 = 2875 unique students	gender, race. ethnicity	peer instruction, media computatio	performance indicators (e.g., grades, test scores)	Context: This is a longitudinal study examining the impact of an alteration in the first year course for non-programmers to make it more interactive. Previous research had demonstrated that this had improved retention and grades within the course, but it was not known whether these benefits would carry on

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
Best Practices CS1. ICER '20. https: //doi.org/10.1145/3372782. 3406274 [105]			before interventi on 1868 + 1179 = 3047 after interventi on		n and pair programmi ng		throughout the full degree. The impact on female students and on under-represented minority groups was considered. Population: CS1 students at a US university over a period of 17 years. Method: Over a decade ago, a university changed it's intro programme for non-programmers to incorporate three interventions: peer instruction, media computation and peer programming. This paper looks at the performance of students throughout their degree courses over multiple years to determine what impact these interventions had over the long term. Results are compared with the intro programme for students with experience, in which these changes were not made, to determine whether there are differences between the two courses (though the students on the courses differ in key ways) Results: From a gender point of view, the results show that the change in course improved female fail rates, time to complete degree and upper-division GPA, and these improvements were not seen for male students. Both male and female students improved switch rates (likelihood of switching into a CS major).
Benjamin J. Schreiber and John P. Dougherty. 2017. Assessment of Introducing Algorithms with Video Lectures and Pseudocode Rhymed to a Melody. SIGCSE '17. https://doi.org/10.1145/30 17680.3017789 [106]	USA	intro computing courses (CS0, CS1, CS2)		gender	Using creative videos to teach algorithms	student survey(s), performance indicators (e.g., grades, test scores)	The paper reports on an intervention aimed to increase CS students' conceptual understandings and change in their confidence regarding CS1 topics relevant to Binary search and Selection sort algorithms. For each algorithm, the authors created a video series: each narrated series begins with an overview of the corresponding algorithm and a step-by-step simulation on an interactive blackboard. After that, it proceeds with a video that demonstrates how to perform a complexity analysis with guided examples, and then applies that to the corresponding algorithm. The series concludes with a video showcasing a song with algorithm pseudocode as lyrics, which are utilised line by line to implement the algorithm in code. The researchers were particularly interested in investigating the effects of these videos on students' technical grasp and self-confidence about understanding and applying the algorithms. To this end, they used a pre-survey and post-survey in which students answered questions about their familiarity with programming, music, and the given algorithm, and how confident they were with their understanding of the algorithm. In total 35 students participated in the Binary search video series and 53 in the Selection sort series. The results of their study demonstrate that students showed substantial improvement in their knowledge of the complexity of each algorithm but the authors do not report gender differences. Regarding students' perceived confidence, there was a significantly positive difference between pre-survey and post-survey questions for the entire sample, evident in both genders.
Amber Settle, James Doyle, and Theresa Steinbach. 2017. The Effect of a Computing-focused Linked-courses Learning	USA	intro computing courses (CS0, CS1, CS2)	51 (intervent ion) 403 (control)	gender, race. ethnicity	Living learning community	student survey(s)	The paper reports the effects of a linked-course learning community on improving women and underrepresented students' retention. In a linked-course learning community, students simultaneously enrol in courses from different disciplines that are connected in content, purpose and organisation. Extracurricular activities that support the goals of the courses can be included in

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
Community on Minority and Female Students. SIGITE '17. https://doi.org/10.1145/3125659.3125679 [107]							order to further connect students with each other and their instructors. The community is designed to provide students with a learning environment that enhances student achievement, improve their feeling of belonging and confidence, improve their study habits, and ultimately, improve their retention. To this end, the researchers created a linked-course learning community as an effort to improve female and minority students' retention at their institution. To measure changes in students' experiences and attitudes, the authors developed and administered a survey to students participating in the learning community and those attending the traditional CS1 course and had not participated in the community. A total of 454 students responded in both the pre and post-surveys. Of these, 403 students took the traditional CS1 course while 51 took part in the learning community. The results demonstrate that learning community students were more likely to agree that they were part of a group of programmers and more likely to say that they had friends who are interested in computing. Also, they were significantly different from other introductory programming students in their reported utilisation of the Internet and in consultation with friends or peers post quarter, with most learning community students less likely to use the internet and more likely to work with friends after their course experience. The authors conclude that since a sense of belonging and the support of like-minded friends are important for the retention of underrepresented groups, their study results are encouraging.
Amber Settle, John Lalor, and Theresa Steinbach. 2015. Evaluating a Linked courses Learning Community for Development Majors. SIGITE '15. https://doi.org/10.1145/2808006.2808031 [108]	USA	intro computing courses (CS0, CS1, CS2), homework or other out of class activities	323 CS 1 students and 17 students in the learning communit y	gender, race. ethnicity	Learning Communiti es	student survey(s)	Context: North American University, linked courses learning community for 1st year students majoring in a degree that requires CS1 (Python) were also enrolled in the same general education course and had co-curricular and extra-curricula activities Population: 323 CS 1 students and 17 students in the learning community Man of color of Women in CS Method: All students at DePaul have to take a GE course, these students were placed in the same one, in addition to their CS1 course. Survey to measure attitudes towards computing and programming was administered at the beginning and end of the term for 3 or 3 terms (not sure what a quarter is) Result: No statistically significant changes in student attitudes or study habits over time. So no meaningful different between pre and post term attitudes. However Students in the learning community were more like to agree with the 3 questions that may signify a growing sense of belonging and support (Q17: "Once I start trying to work on a problem, I find it hard to stop," Q25: "I am a part of a community of programmers," and Q29: "I have a lot of support that will help me to succeed in computer science courses.")
Amber Settle and Theresa Steinbach. 2018. Retention Rates for the First Three Years of a Linked-courses Learning Community. SIGITE '18.	USA	intro computing courses (CS0, CS1, CS2), Learning Community	67 students	gender, race. ethnicity, class/SES	learning community	performance indicators (e.g., grades, test scores), student	Context: three years of learning communities targeting females and minority males enrolled in CS 1 (several computing majors in their first year) Population: 19 students (18 in data set) in 2014-15 cohort, 27 students in 2015-16 cohort, and 22 students in 2016-17 cohort at a private Midwestern institution in the United States. Demographics are given for first generation, Pell grant

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
https://doi.org/10.1145/3241815.3241854 [109]						retention rates	eligible (SES) and whether students came from a Chicago Public School high school but not gender. Method: First year students are required to enroll in a Chicago Quarter class; these students chose to take a learning community linking CS 1 and an Explore Chicago class focusing on the digital divide and specifically on the social issues surrounding access to information and communications technology. Retention rates and GPAs for each cohort is compared to the entire class as well as underrepresented CS1 students and underrepresented School of Computing students. Result: The first cohort initially had higher retention rates and higher academic performance, but their retention rates were lower than the general population by the third year. The second and third cohorts no longer had higher retention rates that other populations. [NOTE: this 2018 paper follows up on at least three other papers on the same Learning Community intervention by the same authors.]
Sadia Sharmin, Daniel Zingaro, Lisa Zhang, and Clare Brett. 2019. Impact of Open-Ended Assignments on Student Self-Efficacy in CS1. Global CompEd. https://doi.org/10.1145/3300115.3309532 [110]	USA	intro computing courses (CS0, CS1, CS2)		gender	Open-ended assignment	student survey(s), performance indicators (e.g., grades, test scores)	Context: CS1 course at large North American research university. They evaluated the impact of open-ended assignments on student self-efficacy in CS1 Population: 245 students (116 in section A and 129 in section b) Method: Used a switching replications quasi-experimental design, in which students in 1 section first got the control assignment and the other section got the open-ended variation, and then they switched assignment types for the next assignment. Data collected: Pre-test survey administered at the beginning of the term, 2. Surveys administered at the end of assignment 1 and 2. The surveys include questions from Computer Programming Self-Efficacy scale (CPSES). The pre-test survey also included relevant info. Relating to gender, prior experience with programming. Result: Students with higher self-efficacy earned higher grades on the assignments. The open-ended assignment had higher self-efficacy scores but this was not statistically significant. Gender had a significant effect on assignment as female students had a higher average on the assignments, but this is not because of the open-ended intervention. So I guess we can conclude that the open-ended intervention did not result in a different result for female students
Max O. Smith, Andrew Giugliano, and Andrew DeOrio. 2018. Long Term Effects of Pair Programming. IEEE Transactions on Education (Aug. 2018). https://doi.org/10.1109/TE.	USA	intro computing courses (CS0, CS1, CS2), other computing courses	152F 593M experimental, 46F 212M control	gender	pair programming	performance indicators (e.g., grades, test scores)	Context: US public university CS1-3 Intervention: pair programming was introduced in a CS2 class Participants: intervention F152 M593, control F46 M212, 1 institution Method: Observational study: exam and project scores in CS3 module were analysed and compared with CS2 student partnerships, GPA and gender. Students had the option to partner in CS2, in which case they had to use pair

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
2017.2773024 [111]							programming (using the same computer). In the CS3 course all students worked individually Results. CS2 exam Z-scores showed a statistically significant larger difference between women who worked alone and men who worked alone. Specifically, women working alone averaged a 0.32 higher exam Z-score than partnered women. In CS3 those who partnered in their earlier CS2 course averaged a 0.14 higher CS3 project Z-score and no statistically significant difference in exam Z-score. In CS2, those who partnered averaged a 0.12 lower CS2 exam Z-score and 0.21 higher CS2 project Z-score (N=2,468). The net impact of partnerships on CS2 Z-scores was positive.
Jeffrey A. Stone and Tricia K. Clark. 2011. The impact of problem-oriented animated learning modules in a CS1-style course. SIGCSE ’11. https://doi.org/10.1145/1953163.1953182 [113]	USA	intro computing courses (CS0, CS1, CS2)	90 students	gender, race. ethnicity	0	student survey(s), performance indicators (e.g., grades, test scores)	CONTEXT: Introduced animated module content to a CS1 course that authors indicate includes interdisciplinary connections though concrete information on this was not provided. Student engagement, course retention, and student success were studied. POPULATION: 90 participants from a US research university, most of whom (90%) were STEM majors; 81 completed surveys including 13 females (16.0%) METHOD: Data was collected from a pre/post module survey, course retention (i.e. DFW rates), and course grades. In year 1, an online module using multimedia was used in the first four weeks before programming was introduced. In year 2, a second module to introduce the C++ programming language was also included. RESULTS: Male students (p<0.05) and white students (p<0.01) were significantly more likely to pass the class than other students. Gender was not significant in course withdrawal and overall course average. Non-white students had a significantly lower course average (p<=0.01). Survey data was not disaggregated by gender or race/ethnicity. DFW rates were lower in year 2 though data on demographic differences was not provided by year.
Elizabeth Sweedyk. 2011. Women build games, seriously. SIGCSE ’11. https: //doi.org/10.1145/1953163.1953218 [114]	USA	other computing courses	208 students	gender	game development	student survey(s), observations of students or classrooms, performance indicators (e.g., grades, test scores)	Context: This paper addresses the issue of using game development projects in CS1 from the issue of potential stereotype threat Population: Study 1: 30 North American undergraduate students over 2 semesters (13Female 18Male) Study 2: 56 North American Undergraduates applying for game-focused independent study projects Method: After receiving complaints of stereotype threat in a game focused software development course, the course was changed to focus on "serious" games (client focus, real purpose instead of student led development). Pilot study: 1 female team, 3 male teams, time tracked, peer evaluation, student outcomes post course tracked. Students developed separate projects for

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							separate clients Study 2: Analysis of student applications for independent study projects. Number/gender of applicants tracked for game related & non game related projects Result: Study 1: - Female team spent more time on project (27hrs vs 15-20hrs/week), continued to work on project as independent study, peer evaluations rated female team's game as best overall - Female students more likely than male students to self report that serious games made good course projects Study 2: - Female students were more likely to apply for game based independent study projects, with big upswing after serious game focused course implemented - Many female students ranked serious game projects as 5 (highest rank), more than male students
Andrea Tartaro and Haley Cottingham. 2014. A problem-based, survey introduction to computer science for majors and non-majors. Journal of Computing Sciences in Colleges (Dec. 2014). [115]		intro computing courses (CS0, CS1, CS2)		gender		number of students' enrollment	This paper describes a new course in an undergraduate computer science curriculum that uses an interdisciplinary problem-based approach to introduce the computer science discipline to both majors and non-majors. Each section of the course focuses on one interdisciplinary problem, which is chosen by the instructor and varies between sections. Although the authors have not formally evaluated the effects of their intervention on specific students outcomes, the initial enrolment data indicate an increase in female students' enrolment as a result of their approach.
Dilan Ustek, Erik Opavsky, Henry M. Walker, and David Cowden. 2014. Course development through student-faculty collaboration: a case study. - ITiCSE '14. https://doi.org/10.1145/2591708.2591723 [116]		intro computing courses (CS0, CS1, CS2)		gender		student survey(s), performance indicators (e.g., grades, test scores)	The paper describes a robot-based CS2 course designed and developed at Grinnel College with the aim to increase students' interest in introductory computer science courses. The course development involved the collaboration of four students with a faculty member; students were responsible for designing, testing and refining the course materials by drawing on their learning experiences with CS courses. The evaluation of the new course was based on students' performance, surveys and feedback through end-of-course evaluation. The evaluation of the new course suggests that even though students' performance did not demonstrate any significant changes compared to the previous years, 65 percent of the students replied positively to the question regarding how useful the robot-design course was to their understandings. Regarding students' interest, the authors reported that the number of enrolments depicts a moderate increase, although no statistical measures are reported. The authors highlight that a part of this increase is due to the enrolment of more females in the whole introductory sequence courses (CS1, C2, CS3), while across a semester, female numbers show a moderate variation (again, no statistical measures are provided).

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
Becky Wai-Ling Packard, Jaemarie Solyst, Anisha Pai, and Lu Yu. 2020. Peer-Designed Active Learning Modules as a Strategy to Improve Confidence and Comprehension Within Introductory Computer Science. Journal of College Science Teaching (June 2020). [119]		intro computing courses (CS0, CS1, CS2)	45 students (36 reported female, 1 male, 1 non binary, 7 not reported)	gender	Active learning	student survey(s)	*Note: I have some qualms about including this, but I wrote it up anyways :) I'm mostly concerned that gender is not mentioned in the reporting. Yes, the intervention happened at a women's college, but even from the participants that did report their gender, it's clear not all participants were women or female. Gender was not a focus of the study, but more a side-effect of the environment. Context: This study investigated the impact of implementing active learning in a CS classroom via peer mentors, who designed and led active learning modules. Population: 45 introductory students in a CS2 course. 36 of the participants identified as female, 1 as male, 1 as non-binary, and the rest did not report their gender. Methods: Students took pre- and post-tests on confidence, interest, and experience, as well as rated their perception of individual modules (with some including qualitative explanations on why certain modules were helpful). Students also completed surveys mid-semester (Weeks 4 and 7) to rate their peer mentor and indicate if lessons were contributing to their confidence or understanding. The authors describe the 6 active learning modules. These active learning modules took place during one hour of a required three-hour lab. Results: The perceptions of the peer mentors were generally rated favorably as were the perceptions of the peer-led active learning sessions. Students generally agreed that the active learning modules were helpful for their learning.
Dana LinnellWanzer, Tom McKlin, Jason Freeman, Brian Magerko, and Taneisha Lee. 2020. Promoting intentions to persist in computing: an examination of six years of the EarSketch program. Computer Science Education (Oct. 2020). https://doi.org/10.1080/08993408.2020.1714313 [120]		other computing courses		gender, race. ethnicity		student survey(s)	Context: Meta-analysis of several US studies, some with high school students, some with college students in 2016/17 Intervention: Students learn the basic elements of computing (i.e., Python or JavaScript code for fundamental computing concepts such as loops and lists) to algorithmically create music in popular genres through sample-based music composition (i.e., composition using musical beats, samples, and effects) using EarSketch Participants: intervention 206, control 163 for the college-related activities Method: pre- and post- "Intent to persist" questionnaires were used Results: the intervention had a positive impact on intent to persist, but this was independent of gender and URM (under-represented minority) status
Hayden Wimmer (2019). Hour of Code: A study of gender differences in computing. Information		intro computing courses (CS0, CS1, CS2), Hour of code		gender		student survey(s)	Context: US Undergraduates in an introductory CS course (not CS majors) undertook a one-off activity - hour of code within the course Population: 99 completed the surveys and are reported on (61% male,

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
Systems Education Journal. [122]		activity with CS undergraduates					Method: Pre and post-intervention programming experience, attitude and knowledge survey. Intervention: An hour of code activity. Results: Before the intervention: 1. males report having taken more programming courses than do females. 2. males are more likely to take a programming course than are females. 3. more males believed programming is important than females. After the intervention: 1. males are more likely to take a programming course than females. 2. females outperformed males on one comprehension question on loops. 3. male participants appreciate more the importance of learning programming than females. Comparing pre and post 1. With no gender difference, respondents are more positive about programming after the intervention. 2.male participants appreciate more the importance of learning programming than females. 3. Males are more willing to take programming courses than females.
Rebecca N. Wright, Sally J. Nadler, Thu D. Nguyen, Cynthia N. Sanchez Gomez, and Heather M. Wright. 2019. Living-Learning Community for Women in Computer Science at Rutgers. SIGCSE ’19. https://doi.org/10.1145/3287324.3287449 [123]	USA	intro computing courses (CS0, CS1, CS2)	37 (intervention) 66 (control)	gender	Living learning community, specific curriculum, targeted recruitment, mentors, admin support, faculty advisor, industry activities	student survey(s), student focus groups, student interviews, performance indicators (e.g., grades, test scores)	The paper describes the experiences of developing and running a Computer Science Living-Learning Community (LLC) for first-year women at Rutgers University. Living-learning communities are programs based in college residence halls that link students’ living environments with a particular academic theme or topic, supported by access to additional academic resources and support. Through an immersive educational and community-building experience, the program is designed to promote student engagement and success in CS. Specifically, the LLC incorporates mentoring, academic and professional development activities, a community of peers for friendship and academic support, and exposure to the issues and applications of CS. LLC students (61 in the current study) were recruited from a pool of women intending to major in CS, as indicated in their admission applications, who have been accepted into the university. To evaluate the program, participants and a similar comparison group were surveyed at the beginning and end of the academic year and a focus group was conducted with program participants. Particularly, a repeated measures quasi experimental design was employed that included both qualitative and quantitative methods to evaluate the efficacy of the program on participants’ expected outcomes based on the goals of the program. LLC participants and a comparison group were surveyed at both the beginning and end of the academic year. At the end of the academic year, university staff held focus groups with LLC

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
							participants. Follow-up interviews were also carried out with some members of the cohort. Overall, evaluation findings suggest that while the program did not make dramatic impacts on students' outcomes over time, LLC participants describe the program as being important to them, and tend to report more mentor support than other women who are not in the LLC program. LLC participants also seem to have stronger intentions to persist in CS than the comparison group. However, both in the LLC participants and in the comparison group, psychological and aspirational outcomes decreased over time which implies that the LLC program did not act as a "buffer" to environmental or cultural threats to the extent it was expected. On the positive front, LLC women were more likely than the comparison group to report engagement in extracurricular activities related to women in computing. Engagement in these activities may help women's persistence in the long-term, because activities like conferences and groups for women expose students to real-world contexts, research innovations, and opportunities to broaden their professional networks.
Xin Xu and Wei Jin. 2021. Game development workshops designed and delivered by peer mentors to increase student curiosity and interest in an introductory programming course. ACM Southeast Conference '21. https://doi.org/10.1145/3409334.3452046 [124]	USA	intro computing courses (CS0, CS1, CS2)	22-34(intervention) no control		Peer design and delivery of courses, comparison of a face to face to online version	student survey(s)	US undergraduate CS1 single college small pilot study (post-class survey) of a course re-design using highly scaffolded, guided peer mentoring by TAs in teaching introductory programming through Game Development Workshops. The study was affected by covid, so that half of the intervention was face to face (F2F) workshops and half online workshops. The intervention was measured by simple student surveys on curiosity about programming and IT and enjoyment of each class. Number of participants was small - 22 to 34 participants of which 7 to 12 female, underrepresented 12-16. 91% of all students prefer F2F Workshops to the original standard class format. 56% preferred Online workshop to the original standard format. Benefits greatly reduced when online. Compared to counterparts, females and underrepresented groups preferred F2F workshops to online workshops and were thought to benefit from peer modelling and personal interaction
Kimberly Michelle Ying, Fernando J. Rodríguez, Alexandra Lauren Dibble, Alexia Charis Martin, Kristy Elizabeth Boyer, Sanethia V. Thomas, and Juan E. Gilbert. 2021. Confidence, Connection, and Comfort: Reports from an All-Women's CS1 Class. SIGCSE '21. https://doi.org/10.1145/3408877.3432548 [126]	USA	intro computing courses (CS0, CS1, CS2)	27 women (intervention) 89 women (control)	gender	Female taught small all women class	student survey(s)	Context: US Undergraduate CS course, 2019. Intervention: An alternative small all-women's class (35 women students) in addition to the traditional lecture class (601 students; including 149 women), covering the same concepts but led by different instructors. Methods: Survey and test results. Results: All-women's class reported, compared to women in the traditional class, 1. significantly greater social connections and comfort collaborating with their peers, 2. greater feelings of support within their class, 3. more confidence in their CS knowledge, 4. a more welcoming classroom environment, 5. significantly lower dropout rate (5.7% compared to 24.8%,

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
Kimberly Michelle Ying, Lydia G. Pezzullo, Mohona Ahmed, Kassandra Crompton, Jeremiah Blanchard, and Kristy Elizabeth Boyer. 2019. In Their Own Words: Gender Differences in Student Perceptions of Pair Programming. SIGCSE ’19. https://doi.org/10.1145/3287324.3287380 [125]	USA	intro computing courses (CS0, CS1, CS2)	104	gender	pair programming	student survey(s)	Context: The intervention looks at the impact of pair programming. It uses a survey to answer the question “what differences do we observe in men’s and women’s perceptions of pair programming?” by examining the students’ views on perceived advantages/disadvantages of pair programming. Population: Study conducted with an introductory CS class in a large US public university. 411 students were enrolled in the course (29.7% female), 104 students submitted usable responses to the survey (36.5% female). CS was the most common major (34.6%), but other engineering and some non-engineering students were also included. 59/1% percent of the male students had prior experience, whilst only 18.4% of the female had. Method: In the third lab session, students were put into pairs, which they would stay in for the next three lab sessions. F/f, m/f and m/m pairs were included. Occasional trios. After the fifth lab, students were given short-answer prompts which included questions about their thoughts on pair programming and their partners. They were given credit for completing these, but not for the content. Results: Overall, positive comments by students fell into seven categories: 1) improved learning experience - most common positive feedback by both sexes - women slightly more likely to mention; 2) career skills - similar in both sexes; 3) positive atmosphere - more likely to be reported by women; 4) networking - more likely to be reported by women, mentioning the mitigation of the isolation women often feel; 5) efficient/productive - more likely to be mentioned by men; 6) more engaged - not frequently mentioned by either group, but more likely to be mentioned by women; 7) personal gain - three women mentioned that this helped their confidence - no men reported the same. Some students also had negative comments about pair programming. Both men and women reported incompatibility with their partners, with two women stating they felt they were burdening their partner (no men did). A few men viewed their partners as freeloading - no women did.
Leila Zahedi, Jasmine Batten, Monique Ross, Geoff Potvin, Stephanie Damas, Peter Clarke, and Debra Davis. 2021. Gamification in education: a mixed methods study of gender on computer science students’ academic performance and identity development. Journal of Computing in Higher Education (March	USA	intro computing courses (CS0, CS1, CS2)		gender	Gamification	student interviews, performance indicators (e.g., grades, test scores)	Context: The paper is about the use of gamification to specifically determine whether this will engage women. Other research on this topic has been inconsistent. Despite specific trends in aptitudes and interests being determinable from a sex-based point of view, this approach was careful to avoid stereotyping. SEP-CyLE, an online gasified tool that rewards adult students’ successful activities with virtual points, was used. Population: 181 UG CS students in a programming I course, 34 of whom (19%) were female. Method: A mixed-method design was used to allow broad understanding of trends at the population level, as well as more detailed insight at the individual level. Two independent strands of qualitative and quantitative data were

Reference	Coun-try	Context / Course	Sample	Demo-graphic groups	Interven-tion	Type of Data	Annotation
2021). https://doi.org/10.1007/s12528-021-09271-5 [127]							collected and analysed in a single phase. Interviews were conducted with participants both before and after they encountered the intervention using semi-structured interviews. Quantitative data included info on time on modules, times on quizzes and quiz results as well as demographic info. Results: There was a marked increase in CS identity development and self-efficacy for both genders, but it is not clear that this was impacted by the gamification. The surmise that gamification has little to no impact on women's interest and engagement in computing, but that it did increase their perceptions of the field and of themselves in the field.
Daniel Zingaro. 2015. Examining Interest and Grades in Computer Science 1: A Study of Pedagogy and Achievement Goals. ACM TOCE (July 2015). https://doi.org/10.1145/2802752 [129]	Canada	intro computing courses (CS0, CS1, CS2)	119 students	gender	peer instruction in CS1 lecture	student survey(s), performance indicators (e.g., grades, test scores)	Context: two sections (one Peer Instruction and one lecture) of CS1 taught in fall 2012 at a large Canadian research-intensive university. The classes used the same labs, assignments, midterm, and final exams Population: 119 students (of 221 taking the CS 1 final exam) took part in the study. 66% were male and 44% were female. 62% were Asian, 32% were Caucasian. Method: Students completed an achievement goal survey early in the semester, interest and enjoyment survey at the end of the survey. Final exams scores were also collected. Result: Students who adopted mastery goals scored significantly higher on the final exam and were more interested in CS than those students who did not adopt mastery goals. Mastery goals in the PI section are more potent for males, and mastery goals in the traditional section are more potent for females. In addition, mastery goals are related to enjoyment for females but not males. Peer Instruction students' interest increase as performance goals increase; for the traditional section, however, interest decreases as performance goals increase. PI may protect against some negative influences of performance goals, but that performance goals may be undesirable overall. There were no gender or pedagogy effects on final exam grades.
Daniel Zingaro. 2014. Peer instruction contributes to self-efficacy in CS1. SIGCSE '14. https://doi.org/10.1145/2538862.2538878 [128]	Canada	intro computing courses (CS0, CS1, CS2)	109	gender	peer instruction	student survey(s), performance indicators (e.g., grades, test scores)	The paper describes the effects of peer instruction (PI) on students' self-efficacy and performance in CS1 course at a Canadian university. The authors compared the outcomes of two CS1 classes, one taught using PI and the other was taught traditionally. In total, the authors compared 109 students' performance in the final exam and their responses on a questionnaire regarding their self-efficacy. The results of their study suggest that students' post self-efficacy was higher than that of students in the traditional section but not significant differences were observed regarding the performance of the two groups in the final exam. Regarding gender differences, males performed better than females and thus, the use of PI did not moderate this relationship.