“Regular” CS x Inclusive Design = Smarter Students and Greater Diversity

Rosalinda Garcia

Oregon State University, garciros@oregonstate.edu

Patricia Morreale

Kean University, pmorreal@kean.edu

Lara Letaw

Oregon State University, letawl@oregonstate.edu

Amreeta Chatterjee

Oregon State University, chattera@oregonstate.edu

Pakati Patel

Kean University, patpanka@kean.edu

Sarah Yang

Oregon State University, sly.yuzh@gmail.com

Isaac Tijerina Escobar

Oregon State University, escobarisaac89@gmail.com

Geraldine Jimena Noa

Oregon State University, noaguevg@oregonstate.edu

Margaret Burnett

Oregon State University, burnett@eecs.oregonstate.edu

What if “regular” CS faculty each taught elements of inclusive design in “regular” CS courses across an undergraduate curriculum? Would it affect the CS program’s climate and inclusiveness to diverse students? Would it improve retention? Would students learn less CS? Would they actually learn any inclusive design? To answer these questions, we conducted a year-long Action Research investigation, in which 13 CS faculty integrated elements of inclusive design into 44 CS/IT offerings across a 4-year curriculum. The 613 affected students’ educational work products, grades, and/or climate questionnaire responses revealed significant improvements in students’ course outcomes (higher course grades and fewer course fails/incompletes/withdrawals), especially for marginalized groups; revealed that most students did learn and apply inclusive design concepts to their CS activities; and revealed that inclusion and teamwork in the courses significantly improved. These results suggest a new pathway for significantly improving students’ retention, their knowledge and usage of inclusive design, and their experiences across CS education—for marginalized groups and for all students.

CCS CONCEPTS •Human-centered computing•Applied computing~Education

Additional Keywords and Phrases: GenderMag, diversity, inclusion, broadening participation, CS education, Inclusive Design, HCI education

1 INTRODUCTION

1946-6226/2023/1-ART1 $15.00
http://dx.doi.org/10.1145/3603535

Suppose a Computer Science department's faculty, most of whose specialties are not Human-Computer Interaction (HCI), decided to start teaching elements of HCI—specifically, inclusive design—throughout most CS/IT major courses. This is what occurred at University X.

This paper presents a year-long Action Research investigation into how this update to traditional undergraduate CS courses would impact University X students. University X is an urban, Hispanic-serving institution (HSI) in the Eastern U.S. (Being HSI in the U.S. is defined as the university's overall enrollment being at least 25% Hispanic.) 78% of University X students receive financial aid. The university's CS department offers majors in both Computer Science (CS) and Information Technology (IT). 52% of its undergraduate CS/IT students identify with marginalized races/ethnicities that are underrepresented in the computing profession (30% Hispanic, 22% Black) and 17% of the CS/IT students are female.

Action Research is an iterative form of field research that involves developing scholarly knowledge through engaging with a community that wants to bring about a change [68, 93]. Because of the connection between problem solving and knowledge development in Action Research investigations, the line between researchers and participants is blurred, such that some participants may act as researchers and vice versa [39, 62, 68, 93]. This is because, in Action Research, research is done "with" the participants instead of just "to" them. This is one way Action Research differs from controlled studies: another is that the latter achieve isolation of variables, whereas Action Research instead achieves real-world applicability and seeks lasting impact.

Action Research is based on multiple tenets, and we summarize some central tenets [68] here. One central tenet defines its process, which is to follow a highly iterative planning-acting-observing-reflecting spiral, the outcomes of which then feed into the next planning-acting-observing-reflecting spiral, and so on. It is also collaborative: those responsible for the actions are involved in deciding how to improve upon them. In our investigation, the participating faculty were those responsible for the changes to their courses, and thus were participant-researchers. For this collaborative tenet to be viable, participant-researchers must have control over the changes they make, and participant-researchers should remain open to surprise and responsive to opportunities, and be able to change their interventions as a result. Sections 3 and 4 will provide several examples of how these tenets played out in our investigation. Finally, toward the aim of developing scholarly knowledge, Action Research is strongly evidence-based, and requires systematic record-keeping and data; we describe our approach to data collection in Section 3 and summarize that data in Sections 4-5.

Our investigation began when 13 CS faculty, inspired by ideas from inclusive design research, collectively decided to bring about change by integrating elements of inclusive design across their 4-year curriculum. The faculty's mission was to incrementally expose students to more and more inclusive design skills, in ways that would reinforce and build over the 4-year degree program. Through this change, they sought to teach students not only how to *recognize and understand* when software is inclusive and when it is not, but also how to *use* this understanding when they build software. The faculty also hoped that innovating their course offerings in this way would positively change students' experiences and success with CS. For example, they hoped that the students' retention in CS would improve, and further that students would start becoming purveyors of inclusivity themselves in their interactions with each other and in their understanding of technology's inclusivity bugs and biases. They also hoped these changes would impact the students' work and attitudes, thereby spreading even beyond the particular courses they were innovating.

The faculty's hopes were not unfounded. One pertinent theory is the Integrated Interest Development for CS Education (IID/CS) framework, which argues that tying CS Education to students' personally relevant interests can foster increases in both retention and performance in CS Education [70]. The rising interest by today's teens and young adults in diversity and equity [23, 80] suggests a potential tie to today's college students' interests. Another aspect of the IID/CS framework suggests that expanding what CS includes has an effect of expanding the number of interests that the curriculum can then match. According to IID/CS, interest matches contribute to increasing diverse students' sense of belonging in CS, which also can improve retention [70]. The notion of belonging, in turn, brings its own theoretical foundations (e.g., [34, 71]). We consider how theories such as these relate to our results in the pertinent results sections.

The inclusive design method that University X used to implement this vision was the GenderMag (Gender Inclusiveness Magnifier) method for finding, fixing, and/or averting inclusivity bugs [14]. The GenderMag method provided a reasonable starting place because it has been used in practice by technologists across the world (e.g., [3, 29, 49, 53, 87, 98]) and because of its accuracy (low false-positives rate): at least 95% of inclusivity bugs detected with GenderMag in other works also arose with actual users in lab studies or in the field (e.g., [14, 77, 98]). Perhaps most critically, several of the faculty members had heard about GenderMag before, and one faculty member had previously tried teaching it, with good results.

GenderMag was a basis, but not an atomic unit. Instead, we splintered it into individual "elements" of inclusive design. For this investigation, an element was any portion of GenderMag. Examples of how faculty used these elements were emphasizing that students were designing for someone else, including a few personas, covering a few different ways users might problem solve, and so on. (We define the elements more precisely in Section 2.1, and show how the faculty built upon them in Section 4.) The 13 CS faculty chose any of the elements they wanted and integrated them however they wanted into their own CS/IT courses. We helped them get started by working with them during a summer workshop and then iteratively over the summer months as they updated their courses.

Most of these courses were not HCI courses—they were "core CS/IT" courses; i.e., non-HCI courses required for the CS/IT degrees such as CS0–CS2, Software Engineering, Capstone, etc. There were three reasons for including most courses, not just HCI courses. (1) The first reason was to integrate inclusivity as a critical ingredient in "*regular CS work*" these future CS practitioners do, so that they begin to understand that inclusivity is a desired attribute of all their CS processes and products, not a separate activity. (2) The second reason was to help students *improve their work products*. Today's CS practitioners need to able to create software artifacts that fit together with other software artifacts. Further, these artifacts need to be highly reusable and highly maintainable. Hence, the APIs, libraries, frameworks, documentation, design documents, etc., that CS practitioners create are the "user interfaces" to other CS practitioners who will need to use these artifacts. Following inclusive design practices as they create such artifacts provides concrete practices by which they can safeguard the reusability, maintainability, and practical usefulness of their work products. (3) The third reason was to *gradually* build students' understanding of inclusive design with a small addition/application in each course, to avoid repetition across courses.

However, the faculty did not change much of any *one* course, because the courses were already full of other content that needed to be taught. Instead, they usually inserted an element as a tiny nuance. For example, an assignment that used to say "write a program to <x>" might be updated to "write a program to <x> and explain whether it will work well for <persona>." The faculty began teaching their courses with such updates starting in Fall 2021. Our study focuses on that academic year.

To find out how this intervention would impact students' experiences, we gathered extensive data over that year, and analyzed it qualitatively and quantitively, to answer the following research questions:

- RQ1: How and to what extent did the approach improve student success, as defined by better grades and better retention in the form of reduced IFW (Incomplete/Fail/Withdraw) rates?
    - RQ1a: ...especially for marginalized students?

- RQ2: How and to what extent did students learn the new inclusive design material?
- RQ3: How and to what extent did the approach change the education climate (climate in the courses, project teams, and CS/IT majors/minors) for students in the computing programs?

Positionality statement: We are of multiple races (Asian, Latinx, White), with national/ethnic backgrounds from Asian, South American, and North American nations. Several of us also have the intersectional identity of women of color. As such, a number of us have experienced lack of representation in computing courses firsthand. However, we recognize that, as academic researchers and people with access to higher education, we are in positions of privilege. One author's position as an inventor of the GenderMag inclusive design method motivated our interest in investigating GenderMag as a vehicle for broadening participation in CS Education. Two of us have inclusivity leadership positions, which brought us credibility with the participating faculty. Two authors are also members of NCWIT, which added the privilege of being well connected in the space of broadening participation. These privileges eased our access to people and institutions with an interest in this work. We are committed to using these privileges to contribute to CS-Education's inclusivity, so as to not only broaden participation in CS, but also to make CS-Education a better experience for everyone.

2 BACKGROUND AND RELATED WORK

2.1 Background: GenderMag

As already mentioned, the inclusive design elements incorporated into the 4-year curriculum are elements of the Gender Inclusiveness Magnifier (GenderMag) method. GenderMag [14] is an evidence-based technique for identifying, avoiding, and fixing inclusivity "bugs" in software, and has been used in practice in a wide variety of settings (e.g., [15, 29, 41, 74, 77, 87, 98]). It is also well suited to classroom use because it is an analytical method that does not require user involvement, and because it is flexible enough to use across a wide range of the software lifecycle, from student-created work in progress to full-fledged products and websites that people use regularly.

GenderMag draws from foundational research on how users of different genders tend to interact differently with user-facing technology, due to users' diversity of cognitive styles (known as "facets" in GenderMag). The five facets used in GenderMag capture diversity of 1) motivations for using tech; 2) information processing style; 3) computer self-efficacy; 4) technology learning style[1] (by process or by tinkering); and 5) attitude toward risk. If a technology feature does not support the full spectrum of a cognitive style, GenderMag terms this feature to have an "inclusivity bug." An example of an inclusivity bug involving the learning styles facet would be a feature supporting tinkerers but not individuals who prefer understanding a process first before fiddling with details. Such bugs are cognitive inclusivity bugs because they disproportionately affect those with particular cognitive styles. Such bugs are also gender inclusivity bugs, because the facets statistically cluster around genders [2, 14, 19, 20, 94, 98].

GenderMag uses three personas: Abi (Abigail/Abishek), Pat (Patricia/Patrick), and Tim (Timara/Timothy), shown in Figure 1. Abi's and Tim's values for each of the facets lie at opposite ends of a wide spectrum. Pat provides a third set of values within the spectra [14]. The underlying principle is that technology that

[1] Here and throughout, "learning style" refers to the GenderMag facet about learning new technologies via process (a top-down style) versus via tinkering (a bottom-up style). This is different from the use of the term to denote learning through different formats (auditory, visual, kinesthetic, etc.).

simultaneously supports all three personas also supports people with different mixes of facet values, as long as all facet values lie somewhere on the spectra.

The GenderMag method sets the personas and their facets into a specialized cognitive walkthrough [14, 65]. As with other cognitive walkthroughs [65], a GenderMag walkthrough involves walking through every step of a use-case/scenario and answering questions about each subgoal/action a user should do to succeed at the use-case. The GenderMag walkthrough also refers specifically to the persona and facets in each question, such as:

Before taking any actions: Will <persona> have this subgoal/take this action? Why/what facets?

After taking the "should take" action: If <persona> does the right thing, will they know that they did the right thing and are making progress toward their goal? Why/what facets?

Throughout the remainder of the paper, the term "inclusive design elements" will refer to anything that uses any portion of GenderMag. The faculty in this paper used: the GenderMag facets (also known as cognitive styles in this paper), the personas, the GenderMag walkthrough, the GenderMag Heuristics (a set of heuristics based on these facets), and the full GenderMag method. We detail ways the students experienced these elements in Section 4.

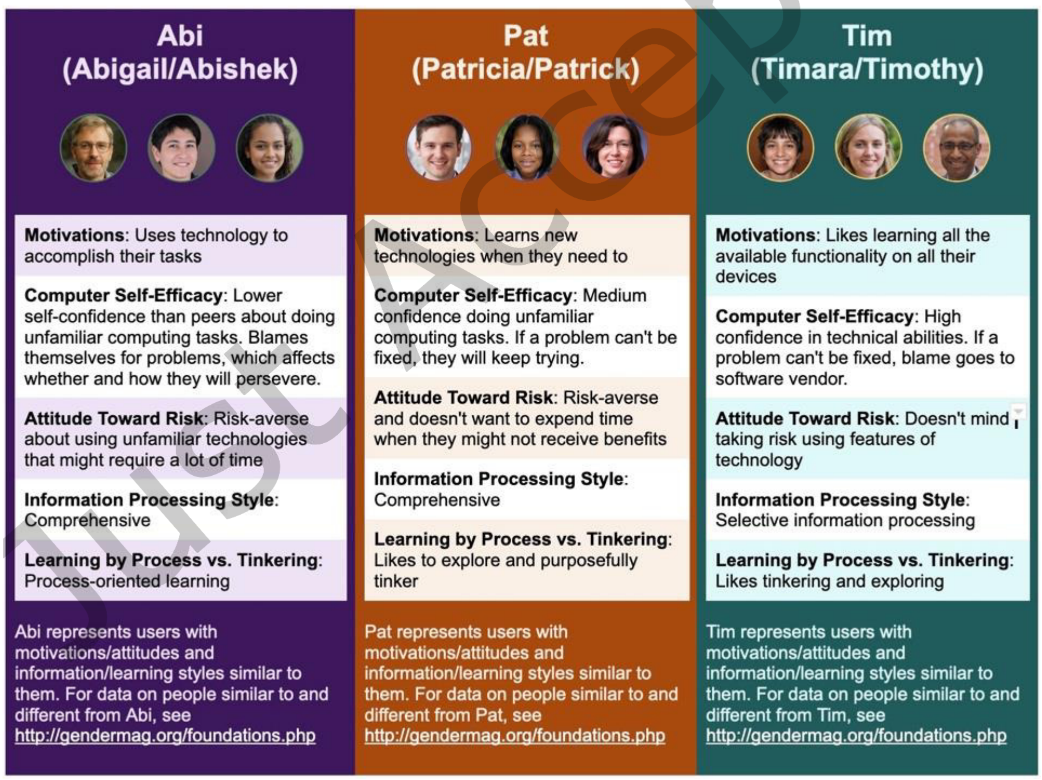


Figure 1: The abbreviated form of the 3 GenderMag personas with short definitions of their facet values [60].

2.2 Related Work

This work relates to efforts to improve retention with and without forms of inclusive design, to Responsible CS approaches, and to multi-year approaches to teaching ethics and inclusive design. In this section we discuss related work in each of these areas.

We begin with work on marginalized populations. As Pournaghshband and Medel point out, much of society embraces a widely accepted "fallacious archetype" of what a successful CS student looks like: a young, White male with at least mid-level socioeconomic status [81]. One area of particular study has been marginalized genders. For example, in online CS communities, LGBTQ+ programmers anticipated that, because of the heterosexist climate, few women and LGBTQ+ people would join [31]. Similarly, in CS education, women often face "othering" [78], have lower retention in online programs [46, 51] and have reported feeling less motivated and less technologically capable [105]. Moreover, when universities shifted to online learning due to COVID-19, first-generation and women students felt a lack of sense of belonging [57].

To address problems like these, a significant body of work has investigated practices for increasing retention across marginalized students in CS education. Among the especially well-known practices are pair programming (e.g., [102, 106]) and leveling the playing field with mechanisms such as having everyone start with a language new to all or eschewing programming entirely to instead focus on problem solving (e.g., [54]). Other research has also shown benefits from including gender-inclusive elements in course presentation, more representation, and using neutral visual designs [52]. Beyond making changes to teaching techniques or course presentation, existing research indicates that including meaningful or socially relevant assignments (e.g. [66]) can positively impact computing students. For example, studies have found positive impacts on students' retention [4] and motivations to complete a project [12].

A number of studies further suggest that teaching aspects of accessibility or inclusive design can increase inclusivity in CS classes. For example, Blaser et al. proposed including universal design principles in engineering courses to increase feelings of inclusion for women students and students with disabilities [8]. The idea that these integrations might benefit women rests on prior investigations that reported women in engineering often particularly value contribution to society, suggesting that women may be drawn to inclusive and universal design (e.g., [38]), which comes back to the socially relevant assignments point mentioned earlier. Additionally, Izzo et al. found that teaching universal design and using universal design-informed teaching practices to include people with disabilities helped students and instructors to improve accessibility, awareness, and instructional flexibility [59]. Another recent work by Oleson et al. in introductory design courses using the CIDER (Critique, Imagine, Design, Expand, Repeat) technique demonstrated that CIDER not only enabled students to understand and respond to many different types of design bias, but also had long-term beneficial effects on their design methods, encouraging them to value inclusivity in their subsequent design work [76].

Some researchers have begun to consider how GenderMag in particular might influence inclusivity in CS education. The GenderMag study that especially motivated University X faculty to embark on their mission was Letaw et al.'s research from two online courses which found that learning GenderMag-based material helped students feel included and improved their interest in finishing the CS major [60]. Chatterjee et al. have also done research with GenderMag in online courses [21], investigating how online instructors can find inclusivity bugs in their courseware using the GenderMag method manually or with an automated version called AID/Courseware. Turning to in-person settings, Oleson et al. did an Action Research investigation into how HCI-oriented faculty teach GenderMag in face-to-face university CS classes [74]. That work produced 11 elements of inclusive PCK (Pedagogic Content Knowledge) that can be leveraged to teach inclusive design in courses. However, no prior work has investigated the impact of embedding elements of GenderMag into "core CS" classes, and no prior work has investigated integrating GenderMag across a 4-year CS degree program.

In addition to impacting diverse students, our approach also aims to enable CS students, as tomorrow's computing professionals, to create inclusive software. This effort aligns with the Responsible CS effort to teach students mindfulness of ethical problems, respect for stakeholders, and how to make accessible design decisions [25, 72]. "Responsible CS" has similarities to "Critical CS" [55] in that both schools of thought aim to encourage CS students to think critically about ethical behavior and be mindful of the societal implications of their work (e.g., [25, 33, 55, 107]). As with a body of work from CS ethics outside of education (e.g., [27, 103]), many Responsible and Critical CS approaches emphasize building awareness through reflection and reasoning activities. For example, one technique for teaching ethical reasoning is critiquing algorithmic design biases [83]. At one university, classes discussed targeted advertising, bias, disinformation, and other ethics-related topics with the aim of learning and reflecting upon these issues [25]. Nevertheless, a recent study suggests that such approaches may not be enough to make a difference in CS students' ethical behaviors [35]. That study found that, even when students were aware of ethical issues, they often did not act upon them ethically. This result suggests that CS faculty also need to teach students new skills for acting upon such issues.

A few approaches go beyond teaching awareness to teaching skills. For example, to teach CS students skills for enacting positive change, some studies have investigated teaching accessibility, for example through including accessibility skill-building in programming classes [47] or in standalone design classes [64]. Oleson et al.'s study attested to the need for such approaches: they found that HCI faculty and students reported lacking the ability to design for diverse populations or to guard against bias [75].

One key aspect of our approach is that it is multi-year, which is a format that has been investigated before, although usually for purposes other than inclusive design, such as ethics. For example, Burge et al. integrated two pedagogical practices into a program-wide framework to teach strong communication skills [13] and Deb et al. integrated big data and cloud computing in three core 2nd, 3rd, and 4th year undergraduate CS/IT courses to prepare students for technological advancements in the industry [30]. Grosz et al. integrated ethical reasoning into courses throughout the CS curriculum through Harvard's Embedded EthiCS program to habituate students to thinking ethically as they develop and design systems [36]. This effort demonstrated that it is possible to teach ethics content in core CS courses, with positive student reactions. Another example is the Responsible CS initiative at Brown University which integrated ethics into multiple assignments in eight courses in introductory and upper-level CS courses with the help of undergraduate teaching assistants [25]. The effort at Brown resulted in 5 takeaways for teaching ethics across the curriculums, including two that are relevant to our approach and the faculty involved: the need for leveraging expertise on how to teach the content and coordination across courses to reduce repetition. Shapiro et al. also began to investigate how to use a role-playing activity as a pedagogical tool in first-year, fourth-year, and online graduate level CS courses to broaden student perspectives and meaningfully incorporate ethics into these courses [86]. Early results of this approach indicated that roleplaying was thought provoking and could help students identify multiple perspectives. Besides improving the interventions' targeted skills, some multi-year approaches have at the same time produced positive results on students' learning outcomes. For instance, Peteranetz et al.'s intervention to include computational creativity exercises in lower- and upper-division CS courses enhanced student grades, their learning of core CS knowledge, and their self-efficacy for creatively applying CS knowledge [100]. Another example is Coleman et al.'s use of collaboration across the curriculum to teach team skills, which resulted in students being more successful at teamwork in the senior capstone [26]. Our approach also has improving students' successes as a goal.

Closest to our approach of enabling faculty to teach inclusive design in core CS classes across the curriculum is work from Waller et al. and Putnam et al., which both integrated accessibility concepts into the curriculum and did so across multiple courses in the major. Waller et al. experimented with integrating accessibility throughout the 4-year curriculum [99] and Putnam et al. suggest guidelines for achieving this goal in the accessibility domain [82]. Our approach builds upon many of the Putnam and Waller recommendations but instead of aiming for teaching accessibility, aims to both integrate *inclusive design skill-*

building and to do so in a cross-curriculum way that increases student successes (including retention), student learning, and the education climate that students experience in CS courses.

3 METHODOLOGY

The educational context of our Action Research investigation at University X was 44 sections of 12 unique CS courses. The investigation involved 13 faculty and 613 students and spanned a full academic year, using the methodology we detail next.

3.1 Procedures

Action Research begins with an appetite for change [107]. To see whether this appetite for change was present at University X, the CS department chair began by canvassing the CS faculty as to whether they were collectively interested in a coordinated adjustment to the 4-year CS/IT degrees involving the use of GenderMag elements, with the goals of improving the (1) students' academic success (especially marginalized students'); (2) their students' knowledge and usage of inclusive design; and (3) education climate of the CS programs (specifically the climate in the courses, project groups, and CS/IT majors/minors). The chair also explained that, as part of the process, the researchers would conduct a summer workshop to help the faculty get started, would collaboratively work with the faculty as they figured out how they wanted to proceed, and would offer a modest $500 summer stipend. University X has a longstanding interest in broadening participation in computing (BPC). The CS department is a member of several BPC-oriented organizations, and regularly runs BPC initiatives; the primary ones prior to this investigation have aimed to (1) promote growth mindset among students, and (2) actively encourage marginalized students to engage in research experiences and/or internships. Perhaps because of these attributes, and because University X faculty are rewarded for emphasizing quality teaching, the faculty was collectively enthusiastic.

13 faculty members ultimately decided to bring about this change and integrate inclusive design elements into their courses during Fall 2021 and Spring 2022. We refer to these terms as F21-intervention and S22-intervention throughout the rest of the paper.

To facilitate faculty's integration efforts, we offered a workshop for interested faculty members during Summer 2021. The workshop was 12 hours in total over 3 days and enabled faculty to incorporate whatever inclusive design elements they wanted into their courses. To accomplish this, the workshop used Letaw et al.'s educate-the-educators curriculum [61] which uses techniques from Community of Practice [101], Training of Trainers [10, 17, 79], and Pedagogic Content Knowledge [88]. We used this curriculum to help faculty gain the following skills: (1: motivational) analyze the costs/benefits of integrating elements of inclusive design into their own courses; (2: content) know how to use GenderMag inclusive design; (3: design) integrate whatever inclusive design elements they wanted into their own courses; and (4: teach) know how to teach the inclusive design elements they had selected. During the workshop, faculty attendees began to leverage what they learned to integrate inclusive design elements into their courses. They then iterated through two rounds of feedback that we provided over the summer. They continued to iteratively develop and improve their new integrations over subsequent terms.

In total, 18 faculty members attended the workshop, 16 of whom were from University X's CS department. Of the 16 CS educators, three decided over the summer not to integrate inclusive design after all and one retired. The remaining 12 of the 16 CS faculty followed through to teach GenderMag concepts in F21-intervention and/or S22-intervention. One more CS faculty member joined the effort in Spring 2022, after attending the archived workshop online, bringing the total number of faculty to 13.

3.2 Participants: the faculty and the students

As mentioned previously, in Action Research some participants may also act as researchers. In this case, the 13 faculty members *created and administered* the intervention (researcher role) in their Fall 2021 (F21-intervention) and Spring 2022 (S22-intervention) classes that the students experienced; and *experienced* (participant role) the intervention in the classes when the students' class experiences affected the faculty's class experiences. The "regular" participants were, of course, the students who took these courses, and the (baseline) students who took the corresponding courses during Fall 2020 (F20-base) and Spring 2021 (S21-base). For reasons of space limitations, this paper reports results for only the student participants, reporting on what faculty did only from the perspective of interventions they created in their researcher role.

Both faculty and students were given IRB-approved informed consent forms; this paper includes data from only those who opted to participate. All 13 of the faculty and 88% of the students enrolled in these courses opted to participate.

Table 1 lists the 13 CS faculty and the 12 courses they chose for integrating elements of inclusive design. Some of those courses were for CS majors, some for IT, and some for both, but the CS and IT majors at this university were part of the same department. We refer to these courses as the targeted courses.

Nine faculty taught the first 6 targeted courses (15 sections) during F21-intervention; this was the first time students experienced faculty's new integrations (course content that contained elements of inclusive design). Then, in S22-intervention, 11 faculty taught 9 targeted courses (29 sections). Three of these courses (CS0, CS1, CS2) had also been F21-intervention courses. 7 of the 11 Spring faculty had taught a targeted course in the Fall term but not necessarily the same course that they were teaching in the Spring. Table 1 details which courses faculty taught during which term. In total, 613 unique students enrolled in 44 sections of the targeted courses (with some students taking multiple of these courses and enrolling in one or both terms); 540 of these students opted in to allow research access to their individual assignments.

Table 1: (Left) CS faculty participants and (Right) the targeted courses they taught. (Column does not add up to 13 faculty because some faculty were involved in more than 1 course.) In total, they taught 12 unique courses, some multiple times, with inclusive design integrated during F21-intervention and S22-intervention. Note: CS0, CS1, and CS2 are ACM's abbreviations for three initial courses in the CS major.

				Undergraduate CS/IT targeted courses	Taken by major(s)	# of faculty	When taught
Faculty teaching load	Men	Women		CS0 (Intro to Programming)	CS, IT	3	F, S
CS Full time (3-5 sections/semester)	4	5		CS1 (OOP)	CS, IT	3	F, S
CS Part-time (Adjunct, 1-2 sec./sem.)	3	1		CS2 (Data Structures, for CS majors)	CS	1	F, S
Total CS faculty	7	6		OOD (Object Oriented Design)	CS	1	S
				Web-CS (Programming the WWW, CS)	CS	1	S
				Web-IT (Programming the WWW, IT)	IT	1	F
				Mobile (Mobile App Development)	IT	1	F
				HCI (Human Computer Interaction)	CS, IT	2	S
				SE (Software Engineering)	CS	1	F

				ProjMgt (Project Management)	IT	1	S
				Cap-CS (Senior Capstone for CS)	CS	1	S
				Cap-IT (Senior Capstone for IT)	IT	1	S
				Total: 12 unique courses	2 majors	13 unique faculty	2 semesters

In addition to students enrolled in targeted courses, our study includes aggregate data on all students enrolled in undergraduate CS/IT courses required for the CS/IT degrees in F20-base, S21-base, F21-intervention, and S22-intervention. Demographics for students enrolled in the CS/IT courses during these terms are summarized in Table 2.

Table 2: Base vs. intervention terms' CS/IT student demographics in all courses. (CS/IT enrollment was provided by the university, which capture biological sex but not gender identity.)

	Total enrolled	Female enrolled	Marginalized race/ethnicities enrolled
F20-base	1401	177	728
S21-base	1479	249	670
F21-intervention	1279	204	583
S22-intervention	1375	200	650

3.3 Data

We collected the data shown in Figure 2 from four terms: F20-base, S21-base, F21-intervention, and S22-intervention. Because triangulation can help safeguard the credibility of Action Research results [16], we collected a variety of data to see if the results manifested across multiple sources, as we will discuss later in Section 6.2.

Our Action Research investigation began in S21-base, at the end of which University X distributed a questionnaire about the education climate to all undergraduate University X CS/IT students in targeted courses only (RQ3). They distributed the same climate questionnaire after F21-intervention and S22-intervention (to targeted courses) and we analyzed data from all three terms. In total, 133 students (from 8 targeted courses) responded to the climate questionnaire during S21-base and 285 responded in F21-intervention and S22-intervention combined (from 6 targeted courses in F21-intervention and 9 in S22-intervention). Climate questionnaire results were anonymous, and were handled by researchers who were not faculty members to avoid faculty bias.

The education climate questionnaire measured student feelings of inclusion in their major/minor, course, and their group project teams through nine Likert scale questions. Students answered questions about teams if they had worked on a team-based project or assignment. They also self-identified their gender and race/ethnicity. The full questionnaire can be found in the Supplemental Documents.

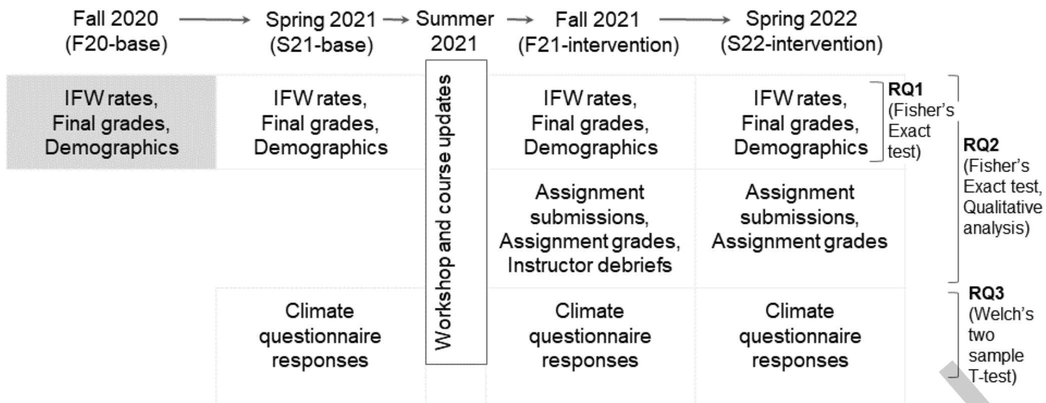


Figure 2: Timeline of data collection and analysis by RQ. Grey indicates the term before our study's initial questionnaire. Row 1's data (IFW rates, final grades, demographics) were collected for all CS/IT courses. Row 2-3's data (assignment submissions, assignment grades, instructor debriefs, and questionnaire responses) were collected for targeted courses only.

For students' academic success, University X also provided us student IFW (Incomplete/Fail/Withdraw) rates, demographics, and final grades for all courses required for a CS or IT major, the aggregates of which we used to measure students' baseline and post-intervention success (RQ1, RQ2). We examined these data for all four terms (F20-base, S21-base, F21-intervention, S22-intervention). During the intervention terms, F21-intervention and S22-intervention, faculty participants also provided us consenting students' grades for assignments containing inclusive design elements. Across the 12 targeted courses in these terms, faculty integrated inclusive design into 1-2 assignments/activities per course section, so the total data of this type spanned 21 unique assignments/activities across these 12 courses' 44 course sections.

Additionally, to measure student's knowledge and use of inclusive design, faculty provided student submissions for these assignments. The specific assignments that contributed to this measure are detailed in Section 3.4.

3.4 Qualitative Analysis

The inclusive design assignment submissions provided three types of data: data type 1 (early takeaways): students' qualitative (free-text) reactions to the new integrations; data type 2 (usage responses): students' ruminations after they had used inclusive design elements; and data type 3 (grades): students' grades on the assignments. Four of the 21 updated assignments were able to provide qualitative data of these types: a CS1 homework assignment prompting students for feedback provided information for data type 1 (early takeaways); an SE homework assignment, a Cap-IT report assignment, and a Web-IT feedback assignment provided information for data type 2 (usage responses). Data type 3 (grades) were quantitative.

We used qualitative coding to analyze the submissions we collected for data type 2 (usage responses). We began with 2 codes from Letaw et al.'s codeset for inclusive design assignments from online courses [60]: "implied users are different" and "explicit users not like me." However, these codes initially captured mentions of only cognitive style diversity/inclusion, so we expanded the definitions of these two codes to include a wider range of diversity (e.g., including mentions of diverse races/genders or mentions of varying familiarity with technology) so as to capture more impacts of the intervention on climate. Since we were using only the inclusive design part of the assignments, we then segmented the dataset such that each students' submission was a single segment. Two coders independently coded 21% of the data and reached

100% agreement (Jaccard method) [89]. Given this high agreement, one coder coded the remaining data. The final code set can be found in the Supplemental Documents.

We also coded these data leveraging Bloom's Taxonomy as a code set. Similar to the above process, two coders independently coded 35% of the data with 80% Jaccard agreement. Given this acceptable level of consensus [90], one coder coded the remaining data. This code set is detailed in the pertinent Results section later in this paper.

3.5 Quantitative Analysis

We statistically analyzed students' final grades and their questionnaire data using inferential statistics. When the hypothesis could be tested with counts of participants, as with the final grades, we used Fisher's Exact Test (contingency tables can be found in the Supplemental Documents). When the question required actual scores or values, as with the questionnaire data, we used an unpaired t-test (details are in the Supplemental Documents). Because our datasets were large enough, we assumed a Normal distribution by the Central Limit Theorem [58]. Figure 2 details the statistical test we used for each RQ.

3.6 A methodological conundrum of the times: What to compare with?

For purposes of comparison against our intervention period, what should we consider to be a "normal" period for University X students, with the only important independent variable being the intervention? We needed to decide which was the most appropriate basis of comparison: the academic terms immediately before the intervention, i.e., F20-S21-base, or the most recent (mostly) pre-COVID period, i.e., F19-S20. (In the U.S., COVID became a major event starting mid-March of 2020, when much of the country shut down.)

We quickly realized that COVID vs. pre-COVID was not the only question. According to news summaries of 2020 and 2021 [42, 43, 56, 67], four major news stories were in play in the U.S. with direct "we are different from them" themes (Table 3). These were (1) COVID (vast differences in different socioeconomic groups' jobs, risk factors, death rates); (2) Jan. 6 violence and Proud Boys (White Nationalists) resulting in perceptions of being excluded and/or practices of violence and exclusion; (3) George Floyd and related violence against Blacks events; and (4) the Anti-Asian hate that ensued, instigated by blaming the pandemic on Asians.

Indeed, a number of studies and reports have found negative effects of all four events on diverse young adults and students. Impacts of COVID are well-documented: COVID disproportionately impacted the lives, financial wellbeing, academic wellbeing, and mental health of Black, Latino, lower socioeconomic status, women, and LGBTQ+ students [5, 11, 22, 32, 48, 57, 69, 73, 84, 85, 95]. CS and STEM university students alike experienced problems with academic motivation [48, 69] as well as changes to the academic landscape that impacted classroom content, tools, and communication with both negative and positive impacts on students [89]. Beyond COVID, studies on Anti-Asian hate in the U.S. reported increased mental disorders for Asians (both Asian-Americans and Asian immigrants) compared to Whites [104]. Other studies have also reported that Asians in the US experienced discrimination [37], increased suicidal ideation [50], and fear [95] due to COVID-related anti-Asian racism. George Floyd, COVID, and the January 6 violence all contributed to reduced trust in public institutions by the Black community [44]. Studies following George Floyd's death also reported increased distress about police brutality for LGBTQ+ and lower socioeconomic young adults [45]. Any of these traumatic events could have impacted groups of University X students' feelings of inclusion as well as their success and ability/willingness to persevere in their studies.

As Table 3 shows, F19 was a "pre-Everything" term, meaning it differed from F21-intervention in five important inclusivity-related independent variables: COVID, George Floyd and related events, Anti-Asian violence, Jan. 6 events, and the intervention itself. Likewise, S22-intervention differed to almost the same extent with S20-pre(almost)Everything. However, F20-base and S21-base had the disadvantage of being primarily remote terms, whereas F21-intervention was fully on-campus and S22-intervention was a mix. Still, the terms most similar to F21-intervention and S22-intervention in terms of all of these variables *except* the intervention are F20-base and S21-base. Thus, we chose F20-base and S21-base as baselines for comparison. However, we recognize the possibility that not all readers will agree with this choice, so also include in the Supplemental Documents full comparisons with F19-preEverything and S20-pre(almost)Everything (collectively referred to as the pre-Everything period), and allude to those results as well throughout the main paper.

Table 3: Timespans of four independent variables (gold) directly related to different student populations' success and feelings of inclusion:
COVID: varying degrees of COVID coping & anxiety (c = University X students were on-campus, r = they were learning remotely; when both, the letters are ordered chronologically).
George Floyd: George Floyd event's aftermath and similar events. (George Floyd initial date: May 25, 2020, after S20 ended).
Anti-Asian hate: Anti-Asian hate spiking due to blaming for pandemic. (First Twitter spike March 2020 [28].)
Jan 6 riot & ensuing investigations (Initial date: Jan. 6, 2021).
We ultimately chose F20-S21 as the baseline because they have the most similar independent variables as F21-S22.

	COVID	George Floyd	Anti-Asian hate	Jan 6 riot & investigations	
F 19					
S 20	√: c+r				
F 20	√: r	√	√		Baseline
S 21	√: r	√	√	√	
F 21	√: c	√	√	√	Post-inter-vention
S 22	√: r+c	√	√	√	

4 THE INTERVENTION: WHAT THE FACULTY CREATED FOR THE STUDENTS

This work investigates effects on the *student* participants. From this empirical perspective, the faculty members' key role was their researcher role to create and administer the intervention—the academic context the student participants experienced. This section describes what the faculty created for the students to experience.

Prior to Summer 2021, the state of the academic context was the way the targeted courses had been taught before. Starting from that state, the faculty began during the Summer 2021 workshop to integrate whichever elements of inclusive design they wanted into their own courses' content in 12 courses—11 non-HCI courses and one HCI course.

Recall from Section 2.1 that the inclusive design elements the faculty drew upon were: the GenderMag facets (cognitive styles), the GenderMag personas, the GenderMag variant of the cognitive walkthrough, the GenderMag Heuristics, and the full GenderMag method. To help with actionability, we provided the faculty with a small starter set of (modifiable) course materials that draw upon these elements, such as lecture slides and videos, a few suggested homeworks and some class activities incorporating these elements, and so on. Over time, the faculty members then built more materials based on these elements and shared them[2].

Recall that one of the tenets of Action Research is that it "starts small", meaning that individuals have control over their changes [68]. Consistent with this point, the faculty members made all decisions about which of these elements to integrate into what courses and in what ways. For example, some faculty members added/changed portions of homework assignments they had used before, some added or changed in-class activities, and some made additions to their lectures; more examples are shown in the upcoming subsections. Faculty members continued to work iteratively on their changes over the remainder of the summer in preparation for F21-intervention.

When fall arrived and they tried out their changes with their classes, their observations and reflections led some faculty members to adapt their changes during the course of the investigation. This too is consistent with Action Research: the faculty members were able to react to their own experiences and incoming data to iterate on their changes over the F21-intervention and S22-intervention semesters, learning lessons along the way about how their integrations were working.

Overarching these efforts was one of the faculty's collective goals for their inclusive design integrations: to gradually build students' inclusive design knowledge over the 4-year curriculum while avoiding overlap between courses. They intended that their integrations into their individual courses would continually and gradually build students' inclusive design awareness and skills. Thus, earlier courses would, one by one, *introduce* a few elements of inclusive design with later courses gradually adding elements with *more breadth and depth* and students *applying* the concepts they were learning (Table 4). Of course, building upon earlier terms wasn't possible during the first term. Thus, some integrations in the first term had overlapping elements of inclusive design (to the irritation of some students); in one case, elements covered in some sections of CS1 were also in a section of CS2.

Table 4: Highest level of inclusive design learning covered during each targeted course per intervention term, with number of assignments per course (parenthesized). The big-picture intent was to build and reinforce students' inclusive design skills gradually over the 4-year curriculum.

	Early-level: *Introduce* 1 to 2 elements of inclusive design	**Mid-level**: *Apply* elements of inclusive design (more concepts/more depth)	**Upper-level**: *Use hands-on* elements of inclusive design when *building* software
F21-intervention	CS0 (1) CS1 (3 sections @1 each) CS2 (1)	Web-IT (1 in one section; 2 in the other)	SE (1) Mobile App Dev (1)

[2] https://www.oercommons.org/groups/gendermag-teach-inclusivemag/10149/

S22-intervention	CS0 (1) CS1 (3 sections @1 each, 1 reused)	CS2 (1) Web-CS (2, 1 reused from Web-IT) OOD (1)	HCI (1) Cap-CS/Cap-IT (2, used in both) ProjMgt (2)

4.1 The Early Courses: Introducing 1-2 Inclusive Design Elements to Students

At the core of inclusive design, as with all of HCI, is that the user is not like the person creating the software. This concept was the first one introduced to the students. The faculty teaching early courses leveraged two elements to do this: the cognitive styles and the personas. In CS0, faculty began by introducing the students to GenderMag's "Abi" persona (the left persona of Section 2.1's Figure 1). Some faculty introduced Abi in more detail than others, but the point was to start students off from the very beginning of their studies thinking about some user not necessarily like them—in this case, the user was like Abi. For example, students needed to consider Abi when writing their first program (Figure 3(a)).

Introductory-level courses after CS0 built upon this concept in different ways. For example, in one section of CS1, students received a handout summarizing the cognitive styles from Section 2.1 (Figure 3(b)) to expand students' awareness of how diverse users think in ways the students may not have considered before. The faculty member then asked them to reflect on these as in Figure 3(c). Another section of CS1 (taught by a different faculty member) introduced two more personas (shown in Figure 1) before asking students to reflect and consider their code from the point of view of all three personas (Figure 3(d)). CS2 also used a similar exercise.

4.2 The Mid-Level Courses: More Elements + Applying Inclusive Design to Technology

In the mid-level courses that follow the introductory sequence described above, students began to apply inclusive design more in-depth. Instead of only reflecting on the content or answering simple questions, faculty began to incorporate the GenderMag cognitive walkthrough element alongside the persona and/or cognitive style elements. For example, the Web-IT and Web-CS courses had students evaluate use-cases or scenarios for entire web sites, using one or more of the personas (e.g., Figure 4(a)). One faculty member also included a prompt for students to suggest fixes for any bugs found during the evaluation (shown in Figure 4(b)). Similar topics were covered within S22-intervention CS2, which had previously contained only early-level elements (Table 4). The OOD (Object-Oriented Design) course used a similar activity to evaluate the different personas' approaches to use-case diagramming. (In that course, use-case diagramming is a standard topic.)

Another instructor for the Web-IT course did a more interactive version of the assignment in Figure 4(a). That faculty member introduced inclusive design to the students by first lecturing on the GenderMag method, the goals, and the importance of inclusive design, and then moved into an interactive think-pair-share activity:

> Web-IT-Instructor: "Then we did two hands-on activities where students broke off into teams of two or three students. And then they did the role-play facet persona activity to evaluate a couple of different websites that I had identified … ahead of time"

This faculty member described the activity as:

> Web-IT-Instructor: "…[not] some random add-on to the course... it was still like a fluid part of what we've been teaching [on building websites]..."

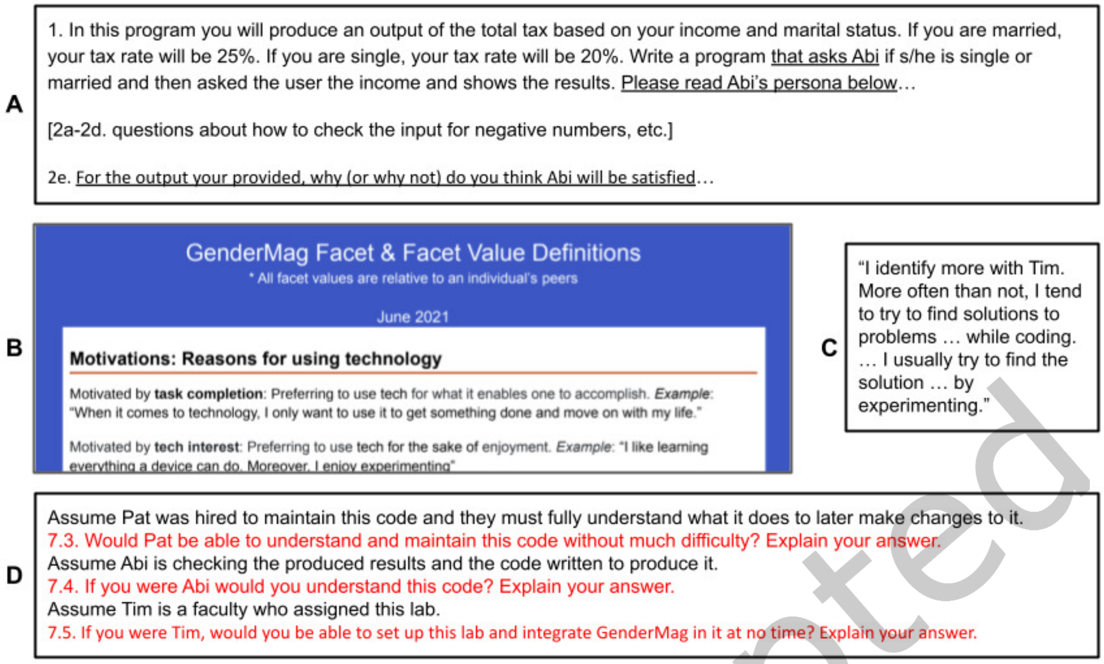


Figure 3: Examples of early courses' incorporation of inclusive design elements: (a) Snippet of one CS0 assignment. The non-underlined text was asked this way in prior terms; we added underlines to show the instructor's small updates to incorporate the early inclusive design concept of designing for someone else, using the Abi persona. (b) Part of a handout used in a CS1 section, emphasizing a wide spectrum of cognitive styles. (c) Snippet of student response to a CS1 lab question asking which persona they identified with most and why. (d) Part of a CS1 assignment from another course asking students to consider how all three GenderMag personas would approach the coding portion of the assignment. (Red text is as per instructor's formatting.)

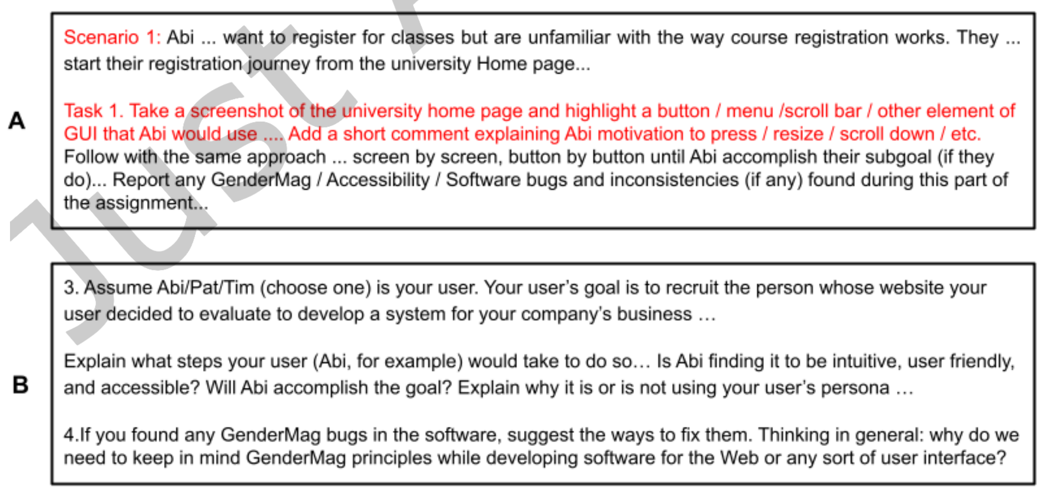


Figure 4: Examples of mid-level courses' uses of inclusive design elements: (a) Snippets from a Web-CS assignment, in which students worked in pairs to do an inclusive design evaluation of website with a pre-set use-case/scenario. (Red text is as per instructor's formatting.) (b) Snippets from a Web-IT assignment to informally evaluate a given use-case/scenario and find solutions to any identified bugs.

4.3 The Upper-Level Courses: Using Inclusive Design Hands-On to Build Better Technology

By the time students reached the most advanced CS/IT courses, they had learned enough elements to begin using the full GenderMag method on their *own* software projects, as per their own use-cases, with their own customizations of the personas. Figure 5 (a) and (b) show examples.

Then, in the final capstones, faculty encouraged students in ProjMgt, Cap-CS, and Cap-IT to use all the elements that they had learned, fully autonomously. For example, in ProjMgt, the faculty member led students through a team building exercise with their project groups utilizing the cognitive styles element before letting the teams determine how best to apply elements of inclusive design such as the full GenderMag method to their projects. The faculty for all three of these courses captured the outcomes through project reports such as those in Figure 5 (c) and (d).

A

1. Personas for your project: Using the personas discussed in class (Abi Tim, Pat):
   a. Select 1-3 personas and edit them to be your project users OR
   b. Develop 1-3 personas and define them in the manner of Abi-Tim-Pat to be your project users.
2. Cognitive walkthrough for your project: Using the worksheet provided, develop 1-4 scenarios, subgoals, etc. Using the persona and taking careful notes, conduct a cognitive walkthrough of your project.

B

1. (25) Create two personas, keeping in mind the five problem-solving facets: *Motivation and Attitude, Self-Efficacy, Attitude toward Risk, Information Processing Style* and *Learning: by Process vs by Tinkering*. Either use the form (https://gendermag.org/customPersona.html) or make your own form…

The personas could be used as part of your team project later, if you decided to do so.

C

Weekly Status Report

Name:

Date:

…

Inclusivity (GenderMag) Issues – related to your project and/or team (complete this section at least once per month):

D

Final Section of Written IT Project Reports

Additional Notes

This section is for any auxiliary information that could serve to enhance the document. Include anything that doesn't necessarily fit well into the preceding sections. In this section, you can discuss any inclusivity (GenderMag) issues as related to your project and/or team.

Figure 5: Examples of upper-level courses' uses of inclusive design elements: Students in SE (a) and HCI (b) developed or customized personas using GenderMag facets to provide appropriate background/education for their project users. Once completed, the students used the personas in full GenderMag evaluations of their projects. At the highest level, Cap-IT and Cap-CS students reported on their inclusive design element usage both in weekly check-ins (c) and as part of their final project report (d).

5 RESULTS: IMPACTS ON THE STUDENTS

As Section 4 described, the intervention was variable and necessarily "messy"—the intervention was not only new to the students, it was also new to the faculty. Thus, consistent with Action Research methodology, the intervention evolved as faculty grappled with problems that arose in situ, experimented with different ideas in different sections, overlapped with each other in unintended and sometimes unavoidable ways, and learned as they went along.

5.1 RQ1: Student successes (and lack thereof)

How did this intervention, despite its startup bumps and hiccups, affect students' successes? To answer our first research question, we compared post-intervention with baseline student success measures for targeted courses. Specifically, the targeted courses were the 6 courses (15 sections) the faculty participants taught in F21-intervention after having integrated elements of inclusive design as per Section 4, and the 9 courses (29 sections) they taught in this way in S22-intervention (Table 1). In the baseline period, the targeted courses were only the courses/sections that corresponded to the intervention's targeted courses.

5.1.1 RQ1's Results for all students

We first consider students' success by looking at their *lack* of success: their IFW (Incomplete/Fail/Withdrawal) rates. In addition to comparing the baseline with the post-intervention data, we examined IFW differences between targeted and non-targeted courses. In the baseline data, significantly more students in the targeted courses had IFW'd than students did in the non-targeted courses (Fisher's Exact Test, p=.002). This suggests the targeted courses may have particularly needed an intervention. Indeed, three of the targeted courses were introductory courses (CS0, CS1 and CS2) that have historically had high attrition rates [6, 91].

After the intervention, these differences disappeared. The improvement in students' IFW rates in targeted courses was so large that the IFW rate in the intervention period's targeted courses landed *below* the IFW rate in the non-targeted courses (brackets in Figure 6(Left)). In fact, students' IFWs from the targeted courses dropped to less than half of their previous rates. Statistically, significantly fewer students IFW'd from targeted courses (Fisher's Exact Test, p<.0001) than in the baseline targeted courses.

Interestingly, fewer students IFW'd (Fisher's Exact Test, p<.0001) than in the baseline non-targeted courses in *non-targeted* courses as well. This outcome may or may not have been influenced by the intervention; it was in line with faculty anticipations of the potential for their mission to impact even non-targeted courses (recall Section 1).

Figure 6: IFW and grade outcomes in targeted (dark green), non-targeted (gold), and all (gray) courses, in the before-intervention periods ("Baseline") vs. the post-intervention terms ("Post"). N= 995 (targeted, Baseline), 1885 (non-targeted, Baseline), 786 (targeted, Post), 1868 (non-targeted, Post). (Left): Percent of students enrolled in the courses who received an incomplete grade, failed, or withdrew (IFWs) from the course. Brackets show gaps between targeted and non-targeted; trend lines show targeted-baseline vs. targeted-post. Low = good. (Right): Percent of students who received an "A" or "B" in the course. High = good.

Since IFW rates measure only lack of success, we also measured students' positive success by the excellence of their grades, i.e., whether they achieved A's or B's in their courses (the two letter grades that denote above average performance). As Figure 6(Right) shows, this measure corroborated the IFW improvement in the post-intervention courses. This difference was significant, too: after the intervention, significantly more of the students received A's and B's in targeted courses than they had before the intervention (Fisher's Exact Test, p<.0001).

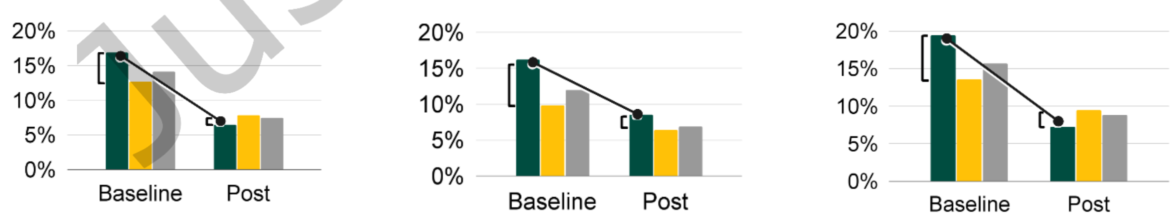
A final source of corroboration came from comparing IFW rates and A/B grades against the pre-Everything period (which includes F19-PreEverything and F20-pre(almost)Everything as described in Section 3.6). In contrast to the comparison with the baseline, the intervention's *non*-targeted course IFWs and A/B grades did not improve compared to pre-Everything; also the intervention's targeted course IFWs improvement did not reach significance (Fisher's Exact Test, p=.1). However, a significant improvement did occur in the pre-Everything period vs. the post-intervention terms: the A/B grades in targeted courses did significantly improve compared to those in the pre-Everything period (Fisher's Exact Test, p=.02) (details in the Supplemental Documents).

5.1.2 RQ1a: The inclusivity emphasis: Effects on females? On racially/ethnically marginalized groups?

GenderMag targets *gender* inclusivity, so one hypothesis is that the intervention might particularly benefit students of marginalized genders that are underrepresented in computing. The university was able to provide students' final grades by binary biological sex only (not gender identity), so we instead investigated a statistically related—but not equivalent—group, namely female students (the student's biological sex as per university records). A second hypothesis, which to some extent competes with the first, is that emphasizing *any* form of inclusivity might benefit students in *any* marginalized group. To investigate this possibility, we investigated students of marginalized races and ethnicities (as per U.S. NSF criteria).

For both of these marginalized categories, the data showed that the intervention was particularly needed for the targeted courses. Comparing Figure 7(a) vs. (b) Baseline columns show that before the intervention, females in the targeted courses had been particularly prone to taking an incomplete/failure/withdrawal grade from those courses in comparison to the non-targeted courses; the Baseline columns in Figure 7(c) show the same phenomenon for marginalized races/ethnicities. These differences were statistically significant: for the baseline, significantly more females IFW'd in targeted courses than in non-targeted courses (Fisher's Exact Test, p=.002), and the same was true for racially/ethnically marginalized students (Fisher's Exact Test, p=.001).

As the Post columns show, both marginalized groups benefited from the intervention. Post-intervention, fewer female students IFW'd from the targeted courses compared to the baseline, although this difference did not reach significance (Fisher's Exact Test, p=.09). Significantly fewer racially/ethnically marginalized students IFW'd from the targeted courses post-intervention than during the baseline (Fisher's Exact Test, p<.0001). These outcomes were again corroborated by the final grades, with significantly more female students (Fisher's Exact Test, p=.02) and racially/ethnically marginalized students (Fisher's Exact Test, p<.0001) earning A's and B's in the targeted courses post-intervention than during the baseline. Even comparing the post-intervention period with the pre-Everything period revealed one significant improvement: racially/ethnically marginalized students' grades were significantly higher in the post-intervention period (Fisher's Exact Test, p=.002).



(a) All students' IFWs baseline vs. post-intervention. (Repeats Figure 6 for ease of comparison.) N=995 (targeted, Baseline), 1885 (non-targeted, Baseline), 786 (targeted, Post), 1868 (non-targeted, Post)

(b) Females' IFWs baseline vs. post-intervention. N=142 (targeted, Baseline), 284 (non-targeted, Baseline), 106 (targeted, Post), 298 (non-targeted, Post)

(c) Marginalized students by race and/or ethnicity IFWs baseline vs. post-intervention. N=503 (targeted, Baseline), 855 (non-targeted, Baseline), 359 (targeted, Post), 874 (non-targeted, Post)

Figure 7: Percent of students who received an incomplete, failed, or withdrew (IFW), during the baseline and post-intervention. Colors: same as Figure 6. Brackets show gaps between targeted and non-targeted; trend lines show targeted-baseline vs. targeted-post. Low=good.

5.1.3 A theory perspective

One way of understanding these results is through the lens of the Integrated Interest Development for CS Education (IID/CS) framework [70]. In essence, IID/CS focuses on the educational effects of connecting students' *interests* with CS Education *content*.

IID/CS has three dimensions. The first dimension is the value students ascribe to the content—due to their pre-existing interests, or through faculty's communications of what is valuable about the content, or through realizations the students have of the value after engaging in the content. All three of these examples of student interest were present. Regarding pre-existing interest, we have already mentioned the rising interest in diversity and inclusion among teens/young adults [23, 80]. Regarding communicating value, several of the faculty explicitly motivated the value of the inclusive design content during lecture, and several others implicitly motivated it by tying it directly to design/maintenance activities (e.g., to web design as in Figure 4(a, b), to software maintenance as in Figure 3(d)). Regarding students' emergent realizations of value, students' quotes in the next section show several instances.

In IID/CS, the second and third dimensions interact with each other and with the value dimension. For example, for the belonging dimension, the framework points out that students' sense of belonging can be increased by engaging them in tasks that they see as valuable. IID/CS proposes that one path for doing so is to expand what is included in CS, and in this work, that expansion was to thread inclusive design into other CS content. For the knowledge dimension, developing knowledge both draws from and adds to first two dimensions. For example, adding to students' skills and expertise in topics the student values (e.g., inclusive design, which they may now see as being part of CS), can further contribute to their sense of belonging. IID/CS also proposes that a way to increase sense of belonging is to expand notions of *who* does computing, which also directly relates to the inclusive design content.

Evidence suggests that aligning education content with student interests via these dimensions can increase performance and persistence (retention). For example, Michaelis and Weintrop point to evidence from neuroscientific research on how interest circuitry in our brains ties to learning and feelings of reward, thereby increasing an individual's attention and performance [70, 40]. Another example is Lewis et al., who have also shown sense of belonging to be strongly tied to performance and retention [63]. Thus, the University X students' improvements in performance (grades) and retention (IFW) align with those predicted by IID/CS.

5.2 RQ2: Did students actually learn any inclusive design?

The students' higher grades in the post-intervention period have another positive implication: they suggest that the intervention did not interfere with the students' ability to learn the CS/IT courses' core content. However, the grades alone did not answer whether these CS/IT students gained understanding or changed behavior *about improving the inclusivity of the technology* they build. To analyze this question, we turned to the three types of assignment data described in Section 3.4: data type 1 (early takeaways), data type 2 (usage responses), and data type 3 (inclusive design grades).

5.2.1 Motivating the Students: Their earliest reactions to new inclusive design integrations

Part of teaching is motivating the content, and data type 1 (early takeaways) from the students helped to reveal whether the faculty's interventions were succeeding at motivating the students. Specifically, we obtained written feedback from students in one section of CS1. These students had used the facet reflection exercise (similar to Section 4's Figures 3(a,b)) in their homework assignment. At the end of the assignment, they responded with open-ended feedback about the element of inclusive design they had just covered.

Students' responses were mixed: Some did not see importance in what they had learned:

> CS1-3: "Well to be honest the usage <of the inclusive design elements> in our class…does not offer much purpose or help to me…"

Some offered suggestions to improve the content:

> CS1-22: "My only feedback would be to include maybe another persona to cover more bases…"

Others were enthusiastic about how it could help them in the future:

> CS1-18: "… it is a good idea to include the GenderMag portion … to further develop things we will use at our future jobs."
> CS1-9: "I think this is a great idea to use GenderMag because it's not around too much … the more that we implement GenderMag, the more understanding people will have."

5.2.2 Learning by doing: Students' outcomes and reactions after applying inclusive design elements to software/websites

Despite the mixed early motivation results, when students began actually using the elements, learning outcomes were positive. To analyze students' learning outcomes after they applied inclusive design elements, we gathered data from the mid- and upper-level courses in which students applied elements of inclusive design hands-on to existing software/websites (Web-IT course, similar to Section 4's Figure 4) or to software they were actually building (SE and Capstone courses; e.g., Figure 5). From these courses, we obtained 42 written responses from students as part of course activities in which they made use of inclusive design elements (data type 2: usage responses). In addition, from the Web-IT course we collected an instructor's comments on how students were making use of the inclusive design elements.

One measure of students' learning outcomes is the Bloom's Taxonomy level [9] they achieve, so to assess how much the students learned from the interventions, we used Bloom's as a codeset to analyze the students' responses above. Bloom's taxonomy, a framework used by educators to categorize different levels of learning, has six levels [1]. Each level requires learners to engage with a higher level of abstraction than the last. The first three levels are:

Level 1: Knowledge/Remember: Remember previously learned information. *In our domain*: list/name/recall attribute(s) of inclusive design elements they have been taught.

Level 2: Comprehension/Understand: Demonstrate an understanding of the facts. *In our domain*: express or describe something they learned from inclusive design elements.

Level 3: Application/Apply: Apply knowledge to actual situations. *In our domain*: use elements of inclusive design to design/build/evaluate a technology product.

Levels 4–6 are considered more advanced, ranging from Analyzing to Creating a new whole. For our case, these levels are hard to distinguish from Level 3 (applying what was learned), because applying inclusive design often necessarily involves analyzing, evaluating, and/or creating. Thus, we differentiate only Level 1 from Level 2 from Level >=3. The results are shown in Table 5. Based on this coding, the majority of students demonstrated Level >=3, indicating that most of the students in these data successfully learned to use elements of inclusive design.

Table 5: Maximum Bloom's Level of learning some element of inclusive design, as demonstrated by 42 responses from 3 classes (1 section of each class). Total is 101% due to rounding.

Level, applied to our domain	#	Example	Evidence of this level
No evidence of any level	5/42 (12%)	Web-IT-9a (from written feedback): "Not really important for me, so I don't have any Question."	N/A
Bloom's Level 1 (Remember): Name/recall some aspect(s) of inclusive design elements.	7/42 (17%)	Cap-1 (from their team's final report): "When it comes to inclusivity… GenderMag was still a pretty new concept / topic to us. We have a general but rough idea as to what GenderMag is and how we can incorporate it into our website."	Is aware that GenderMag is about inclusivity.
Bloom's Level 2 (Understand): Express or describe something they learned about inclusive design.	7/42 (17%)	Web-IT-10b (from written feedback): "It was a great ... reminder that there are different learning styles."	Knows that the approach's core is a set of different styles, one of which is learning style.
Bloom's Level >=3 (Apply): Use it to design/build/evaluate a product.	23/42 (55%)	SE-1 (describing how it changed their team's product): "...We were very much <oriented toward> a target audience of teenagers and young adults who are very tech savvy... when reality is, there are many individuals who require different methods of learning in order to understand/navigate..."	An excerpt of their remarks on applying it to the analysis and redesign of their product's features.

In addition, the Web-IT instructor gave this Bloom's Level >=3 assessment of their students' collective mastery of the inclusive design elements taught in that course:

Web-IT-Instructor: "... students were able to not only adopt the personas, ... but use it to evaluate those websites ..."

Beyond an understanding of the inclusive design concepts, students also described a shift in their own inclusivity awareness. As shown in Table 6, our qualitative coding revealed that 15 of the 23 students in SE and 4 of the 7 groups in Cap-IT reported that the concepts helped them to consider diverse users or users not like them:

> SE-6 (Homework): "It was important … I wouldn't have realized how the world is made up of different people (different ages, cultures, professions and behaviors). I would have excluded a lot if it wasn't for GenderMag."
> Web-IT-2a (Written Feedback): "I thought it was helpful because it made me think of other peoples pov <point of view> when doing something."
> Cap-4 (Team report): "Our entire group … incorporating that with what we know about GenderMag into our website."
> Cap-3 (Team report): "… factored into the project has allowed us to unlock an insight of understanding for how different people have their own individual needs."

Table 6: Qualitative coding results from data type 2: (usage responses). The Web-IT feedback assignment and the SE Homework assignments were completed by individual students while the Cap reports were completed by teams of 1-4 students (most students worked in teams of 4). The full code book with code descriptions can be found in the Supplemental Documents.

Code	Web-IT: Feedback	SE: Homework (Students)	Cap: Report

	(Students)		(Teams of 1-4)
Implied users are different	-	5/23	4/7
Explicit users not like me	2/11	8/23	-
Total Unique Students/Teams	2/11	13/23	4/7

These changing attitudes and understandings also impacted the course itself. From the Web-IT instructor's point of view, by evaluating existing software/websites using inclusive design elements like the one in Figure 4, the students not only learned elements of inclusive design, but also gained perspectives that reinforced the "core" mission of the Web-IT course:

> Web-IT-Instructor: "I was pleasantly surprised … [what they were learning here was] also reinforcing some of their web development knowledge that I've been trying to impart…"

In fact, 37 of the 42 students' remarks also indicated that they thought the inclusive design elements improved their work:

> Web-IT-3a (Written Feedback): "... actually helps me use a website differently."
> SE-15 (Homework): "... putting myself in the shoes of others really helped me to understand the shortcomings of our software."
> SE-20 (Homework): "... after applying the gendermag method we realized that these assumptions could potentially be problems."
> SE-4 (Homework): "... <it provided us> an efficient way to detect these issues on our application and features so that we can come up with <fixes>".
> SE-3 (Homework): "GenderMag ... helped to improve the inclusiveness of the ... website. ... the ecommerce website relies on every gender within the population ... GenderMag was crucial in identifying even the subtle issues."

5.2.3 Students' inclusive design grades: As good as their other grades?

Another way to measure students' learning of the inclusive design elements is by comparing their graded inclusive design assignments against their final course grades. Toward that end, faculty provided us final course grades and assignment grades for assignments into which they had integrated inclusive design elements (faculty's new integrations). The grades came from 24 targeted course sections (10 in F21-intervention and 14 in S22-intervention). Because we wanted to compare the quality of students' inclusive design work against their "usual" quality of work in that class (for which we used their final grades) we excluded any such assignments that faculty had graded using simply pass/fail. We then converted each student's letter grades to numeric scores (i.e., A = 8, A- = 7, B+ = 6, etc.) and compared those assignment grades with students' final course grades.

The results showed that at least 2/3 of the students—73% of the F21-intervention students and 67% of the S22-intervention students—earned at least as high a grade on their inclusive design work as they did on their final grade. This suggests that at least 2/3 of the students showed at least as good an understanding of the inclusive design elements as their understanding of the rest of the course material.

5.3 RQ3: The Education Climate

Beyond impacts on learning and success, we hypothesized there would also be positive outcomes for the education climate (i.e., climate in the courses, project teams, CS/IT majors/minors), especially for students of marginalized races/ethnicities and genders. To investigate this hypothesis, we compared baseline with post-intervention responses to the climate questionnaire discussed in Section 3.3. We defined students' climate ratings as the combination of responses to all Likert scale questions. To do this, we normalized the responses to each question to a 0-100 scale to account for questions with different numbers of options (because one question had 3 options, two questions had 6 options, the rest had 5 options), added these up for each participant, and normalized the total for students who had left some questions unanswered. Students who answered no questions were not considered. Given that there were 9 questions, the total rating possible was 900. We analyzed the raw totals, but also present them as a percentage of the 900 possible rating.

Post-intervention, students reported significantly improved climate ratings, with a mean increase of 31 points out of the 900 possible (from 736 to 767, i.e., 82% to 85%; Welch's Two Sample T-Test, p=.03). In fact, students reported improvements in all areas, as shown in Figure 8. In particular, the largest improvements reported by students were in instructor ability to create an inclusive classroom environment and inclusion in teams compared to past terms. Detailed breakdowns of the responses to these two questions are shown in Figure 9 and breakdowns for all questions can be found in the Supplemental Documents.

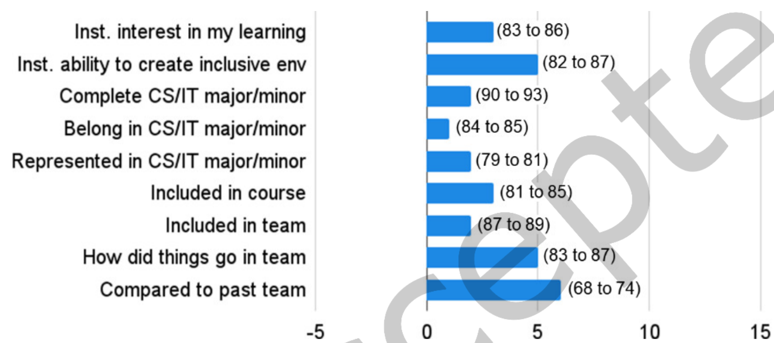
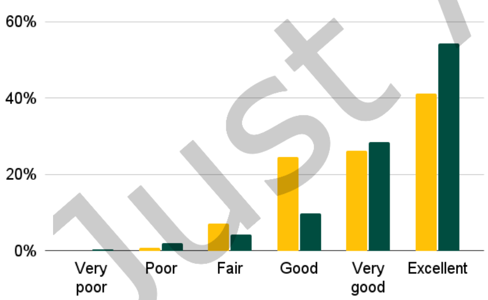
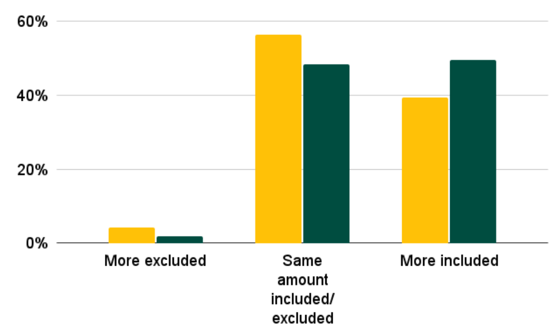


Figure 8: Improvements in all students' climate ratings in the targeted courses baseline vs. post-intervention, for all nine questions. N= 132 (baseline), 284 (post) Bars' labels: baseline to post-intervention climate score, with difference plotted on an x-axis with max of 100% difference. >0% = improvement. Climate ratings improved for every question.



"Instructor ability to create inclusive classroom environment" N=126 (Baseline), 284 (Post)



"Compared to past CS/IT teams, in this course I felt…" N=117 (Baseline), 252 (Post)

Figure 9: Climate comparisons: Average inclusive classroom environment climate ratings by all students in the targeted courses, SP21-base (gold) vs. post-intervention (dark green). Baseline ratings were already good, but post-intervention ratings were better.

In their assignments collected for qualitative data types (1) and (2) (as per Section 3.4), students corroborated inclusiveness outcomes through feedback on the elements of inclusive design they had experienced:

> Web-IT-11a: "... interesting and valuable activity … teachers get a better understanding of how [different students] think"
> Web-IT-10b: "... great way to learn about my classmate. It was a great reminder [of differences]..."
> CS1-13: "…GenderMag in our class is very useful … professors could see who their student relates [to]."
> CS1-11: "...it is very important to be inclusive … GenderMag brings us so much closer to that goal..."

Intertwined with education climate is sense of belonging, which was discussed earlier for its importance to the Integrated Interest Development for Computing Education (IID/CS) framework (Section 5.1.3). Good et al. defined sense of belonging as "one's personal belief that one is an accepted member of an academic community whose presence and contributions are valued" [34, 71]. Belongingness research shows that, without feeling like they belong, individuals' motivation and persistence decrease, with an accompanying decrease in performance; conversely, increasing sense of belonging can increase motivation, persistence, and performance [7, 18, 92].

Since climate is intertwined with belongingness, then belongingness theory would predict that University X students' climate ratings increases such as in Figure 8, should show accompanying increases in student performance and retention [71], which indeed they did (Section 5.1.1). IID/CS would also predict that the expansion of what and who are included in computing should increase belongingness and/or climate outcomes [70]. Thus, both IID/CS's and belongingness theory's predictions are consistent with these students' outcomes.

However, for marginalized students, the climate results are more nuanced (Figure 10). Marginalized students' ratings did increase over the baseline period, but not by a statistically significant amount—from 79% to 83% for marginalized genders and from 82% to 84% for marginalized races/ethnicities, versus 82% to 85% for all students.

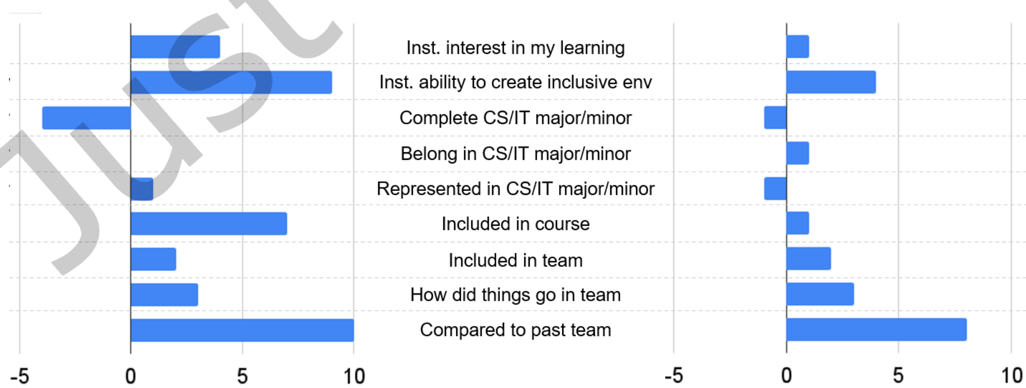


Figure 10: Improvements in marginalized students' climate ratings in the targeted courses baseline vs. post-intervention. (Left) Marginalized genders (non-men) students. (Right) Marginalized race/ethnicity students. Baseline N = 29 (non-men), 69 (marginalized race/ethnicity); Post-intervention N = 59 (non-men), 165 (marginalized race/ethnicity). x-axis: Same as Figure 8, with >0%=improvement. For both marginalized groups, 7/9 questions showed improvement.

The only two climate questions that decreased for marginalized students were completing the CS degree and feeling represented. These are two of the only three questions with strong dependencies on beyond-education factors (e.g., finances), and one possibility is that the improvements in climate the marginalized students noted for the other questions may have been outweighed by such factors. For example, completing a major depends in part on the beyond-education factor of spendable income, which particularly impacts marginalized genders and marginalized races/ethnicities (e.g., [97]). Feeling represented in CS depends on who students see in their classes and in computing professions, which is disproportionately low for marginalized races/ethnicities [108]. It is disproportionately low for marginalized genders as well [108], but those students' ratings of representation increased; one possibility is that the inclusion of *Gender*Mag was particularly impactful for students of marginalized *genders* feeling better represented.

For the remaining questions, the marginalized students' climate ratings increased. Particularly notable are the magnitude of improvements in these students' team experiences—improving by 10% for those of marginalized genders and by 8% for students of marginalized races/ethnicities. Perhaps more important, the overall improvements in climate for the marginalized students—4% for marginalized genders and 2% for marginalized races/ethnicities—were roughly the same as for all students (3%). This outcome is very encouraging and is consistent with the goal of inclusive design itself: to improve experiences not only for underserved populations, but for everyone [24].

6 DISCUSSION

6.1 Limitations

No empirical study is perfect. One reason is the inherent trade-off among the kinds of findings different types of studies can produce. Controlled lab studies can isolate controls but cannot measure real-world applicability. Field studies, including Action Research studies, can capture real-world applicability, but at the cost of not being able to isolate different variables.

In fact, Action Research is even less viable than other field studies for isolating variables. It not only tries to capture real-world applicability, but also at the same time allows the "treatment" to change, so as to not only study real-world situations, but also improve them [39, 62, 68, 93]. Thus, consistent with other Action Research investigations (e.g., [14, 74]), in our investigation the intervention varied from one course to the next and from one faculty member to the next. Each faculty member devised their own ways of implementing the integration in their own course section. Even within a single course section, interventions could vary over time, because faculty used students' responses to finetune their intervention for that class. Given variability like this, no Action Research investigation can "prove" causation—and ours does not do so.

Also, as a real-world field study, an Action Research investigation does not control the numerous independent variables involved. Even beyond the intervention's variability, other uncontrolled factors may have influenced the outcomes. For example, students' grades can be influenced by a myriad of factors: a student's mood, a faculty member providing extra support to students because of the difficult climate of the time, a faculty member grading an assignment with inclusive design elements less strictly than coding assignments, etc. Any of these factors could influence outcomes.

Another limitation of our investigation is that it took place at a single institution. It spanned one year, which yielded extensive data, but the intervention goal is to be an across-the-major integration into the four-year CS/IT curriculum, with hoped-for cumulative effects. Given the timespan of only one calendar year, we were not able to investigate how students might respond to experiencing integration over a four-year period. For example, it could turn out not to fit together in ways that appropriately build from year-to-year. Another limitation is that participant recruitment was subject to (self-)selection bias: both the faculty and the students who consented to participate in our investigation may have different characteristics than faculty and students who did not participate in the investigation. Also, the intervention took place at a single institution in multiple courses, with a variety of students being taught by a variety of faculty. Finally, as noted earlier, choosing which baseline period to use for comparisons was a judgment call.

To mitigate all of these limitations, we triangulated our results extensively, as discussed in Section 6.2, and the triangulation results generally supported our results' robustness against these many variations. That said, with so many limitations, we do not view our results as being generalizable beyond the particular context of our investigation, but rather as encouraging evidence of the approach's promise.

6.2 Triangulation: What it says about the validity of our results

As in any Action Research investigation, in the absence of controls, confounds abound in our investigation. To guard the reliability of the results despite such confounds, triangulation often plays a central role in Action Research investigations in education settings [16]. Triangulation can provide confidence in whether an investigation's measurements and interpretations of the measurements validly reflect what really happened. (In contrast, triangulation is not a way of providing confidence in the generality of results or in showing causation, as already pointed out in Section 6.1.)

As Capobianco and Feldman explain, the essence of triangulation is viewing a situation from multiple angles and perspectives to see whether they corroborate one another [16]. As they point out, one recommendation for education settings is to not only bring multiple data sources, but also involve multiple points of view: those of teachers, of students, and of observers/investigators. In line with these recommendations, we made extensive use of triangulation from both data source and point-of-view perspectives.

From the data source perspective, each RQ section of Table 7 shows how each RQ's result was cross-confirmed by multiple data sources of evidence. Details on each source of evidence and when it was collected were given in Section 3.3, particularly Figure 2. As Table 7 shows, every result was cross-confirmed in multiple ways—RQ1's outcomes (including RQ1a) were evidenced in 10 different sources of evidence, RQ2's outcome was evidenced in 3 different sources of evidence, and RQ3's outcome was evidenced in 5 different sources of evidence.

Turning to the points-of-view perspective, every data gathering method depends on different points of view. The data thus depend in part on both the actors who define the method and the actors who create data under that method. Our triangulation involved all three recommended points of view: teachers, students, and investigators, as we explain column-by-column. The grades data ("Gr" columns in Table 7) were created by the 13 *teachers* who taught the 44 course sections, and the IFW data ("IFW" columns) by the *teachers* and the 613 *students* who took their courses. The climate gathering questionnaire ("Q'aire" column) was created by the *investigators*, and the resulting data came directly from the *students*. The submissions and instructor debriefs ("Subm"and "Debrf" columns) were contributed directly from the *students* and *teachers*, respectively, and the results we drew from them came from the *observers*/*investigators*. All results that break out marginalized groups also involve separate points of view from *different subpopulations of students*.

Thus, as Table 7 shows, RQ1, RQ2, and RQ3 were each supported in total from multiple data perspectives and from multiple (sub)populations' points of view.

Table 7: Results for each RQ, triangulated across multiple sources of evidence. (Comparisons against pre-Everything are detailed in the Supplemental Documents.) Gr=Grades, IFW=IFW rates, Debrf=Instructor debriefs, Q'aire=Questionnaire, Subm=Students' submission content. Black portions are N/A for that row.
Significance annotations: ***: p< .001; **: p< .01; *: p< .05; ~: p< .1; †: raw data but not signif; √: evidence (but comparisons N/A).

	Course outcomes vs. Baseline		Course outcomes vs. pre-Everything		Climate vs. Baseline	Inclu. design learning		
	Gr	IFW	Gr	IFW	Q'aire	Gr	Subm	Debrf
RQ1 & RQ1a: students' success improved:								
All students	***	***	*	~				
Marginalized sex (Female)	*	~						
Marginalized race/ethnicity	***	***	**	†				
RQ2: learned inclusive design:								
All students						√	√	√
RQ3 education climate improved:								
All students					*		√	
Marginalized genders (Non-men)					†		√	
Marginalized race/ethnicity					†			

6.3 Choosing the "Right" Baseline

We pointed out in Section 3.6 that choosing the baseline period for comparison with our intervention period was challenging. In fact, we posit that challenges in doing so not only arose for us and not only in this particular study, but will arise for most researchers doing CS-education field investigations involving 2020-onward data.

In 2020, CS-education climates changed dramatically. Here we focus on U.S.-based CS-education, which is our setting, but CS-education in many countries also faced similar phenomena. In pre-2020 investigations in U.S. settings, comparing a real-world intervention against a baseline was a fairly straightforward design decision: compare a year with the intervention against the year before, in the same courses with the same set of instructors. However, two categories of events beginning in 2016 rendered this formerly straightforward choice no longer straightforward.

First, in late spring of 2020, COVID suddenly appeared and exploded in the U.S., changing everyone's lives in a matter of days. Universities (including the one in this paper) abruptly closed their physical doors, and careened into zoom-based classrooms. Students and faculty were confined to their homes. Most medical care of any sort was suddenly unavailable. Even visiting the grocery was frightening. Second, hate-based violence exploded in both frequency and in public attention—from George Floyd (2020) to Trump's escalation of Anti-

Asian hate (2020) to the January 6 violence (early 2021) and more. Suddenly in the midst of coping with these factors, faculty and students across belief systems, cultures, races, and genders found themselves in a life populated by emotional, financial, infrastructural, and/or physical traumas.

Our investigation occurred in the 2021/22 academic year (Fall 2021-Spring 2022), when all four of these factors were still in play to varying extents (recall Table 3). In deciding what to compare against as a baseline, we had to decide between two possibilities. The first possibility was that the fairest comparison was against the "most similar" year to the intervention period, namely 2020/21 (Fall 2020-Spring 2021). The second possibility was "most normal", namely 2019/20, under the assumption that the U.S. will soon revert to its "normal" (2019) status.

We ultimately decided on "most similar", namely the 2020/21 year (labeled F20-base + S21-base in this paper), but also summarize throughout this paper additional comparisons against "most normal" (F19-preEverything + S20-pre(Almost)Everything) with full details in the Supplemental Documents. We have provided both of these comparisons so that we can leave it to the reader to attend to whichever comparison they believe is the most valid. Note as per Section 6.2 (RQ1), that two of the three results that involved such comparisons are supported by both the comparisons against Baseline and the comparisons against pre-Everything.

We recommend following this strategy, allowing readers to choose for themselves, to other researchers engaged in field investigations involving 2020-onwards data, at least until new norms emerge for empirically investigating 2020-onwards education environments.

7 CONCLUDING REMARKS

This paper's Action Research investigation, on how integrating inclusive design into a 4-year curriculum changed students' experiences, produced results that we consider to be remarkable—because at least four factors were working against the intervention's ability to make improvements. First, the U.S. ecosystem surrounding the endeavor was fraught with challenges to many students' ability to thrive in educational endeavors (recall Table 3) [5, 11, 22, 32, 37, 44, 45, 48, 50, 57, 69, 73, 85, 95, 96, 104]. These challenges could easily have interfered with students' IFWs, their grades, and their feelings of being included, especially for members of marginalized groups in CS.

Second, most of the faculty specialized in non-HCI areas of computer science, such as computer networks, internet of things, mobile computing, operating systems, database systems, and security. Yet, the elements of inclusive design they were teaching to the students was HCI content. Third, consistent with the iterative nature of Action Research, the faculty continually refined their integrations of inclusive design in a trial-and-error fashion, with the students' experiences showing them what was working and what was not.

Fourth, in some ways the students' successes had nowhere to go but down. University X is a primarily teaching-oriented institution, and faculty members' reward systems are based on quality teaching. Thus, their classrooms were already fairly highly rated by the students. For example, the Baseline numbers in Figure 9 showed that 40% of the Baseline students had already rated the inclusivity environment as "Excellent"—a high bar for the intervention period to outperform.

And yet, the interventions these faculty created did outperform the baseline period, and in numerous ways. Each result adds more evidence suggesting that the results are real and robust. The intervention outperformed the baseline in not only students' IFW rates, but also their grades. The students succeeded with the core CS content while also actually learning inclusive design skills. The intervention outperformed the baseline for not just marginalized students but for all students. In addition, most comparisons against pre-Everything courses yielded results consistent with comparisons against Baseline courses.

Among the findings were:

- RQ1 (Successes and Retention): Even though retention (measured here with IFW rates) in the targeted courses had been particularly problematic in the Baseline period—particularly with introductory courses that generally tend to have high attrition rates [6, 91]—the Post-intervention students' IFWs from the targeted courses dropped to less than half of their previous rates, and the differences in IFW rates between targeted vs. non-targeted courses disappeared. Students' grades in the targeted courses also improved significantly.
    - RQ1a: For marginalized students, significantly more female and racially/ethnically marginalized students earned A's and B's in the Post-intervention targeted courses compared to Baseline targeted courses, and IFWs also improved for both marginalized groups.
- RQ2 (Learning Inclusive Design): Many students who learned elements of inclusive design in their courses demonstrated new understanding and even were able to apply what they had learned to technology they were creating.
- RQ3 (Climate): Students' post-intervention climate results were significantly higher than in the Baseline period; for example, students not only felt more included in the course and their teams, but also gave higher ratings to the instructor's ability to create an inclusive learning environment.

This investigation offers encouraging evidence that, if educators choose to incorporate elements of inclusive design into most courses in a computing major, students can benefit with improved academic performance, a new-found understanding of how their software products can support diversity, and feelings of being more included in their education environments. These results suggest a powerful new pathway for significantly improving students' experiences across CS education.

ACKNOWLEDGMENTS

We are grateful to the editor and reviewers for their insightful suggestions and tips which greatly improved this paper, and to the participating faculty and students who made this work possible. This work was supported in part by the OSU Center for Research in Engineering Education Online, the Center for Inclusive Computing at Northeastern University, NSF 1901031, 2042324, and 2042341, and joint support by NSF and USDA-NIFA under 2021-67021-35344. Any opinions, findings and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the sponsors.

REFERENCES

[1] Lorin W. Anderson, David R. Krathwohl, Peter W. Airasian, Kathleen A. Cruikshank, Richard E. Mayer, Paul R. Pintrich, James Raths, and Merlin C. Wittrock. 2001. A Taxonomy for Learning, Teaching, and Assessing: A revision of Bloom's Taxonomy of Educational Objectives. Pearson, New York, NY, USA.

[2] Manon Arcand and Jacques Nantel. 2012. Uncovering the nature of information processing of men and women online: The comparison of two models using the think-aloud method. Journal of Theoretical and Applied Electronic Commerce Research 7, 2 (2012), 106–120.

[3] Moisés Barbosa, Isabelle Brilhante, Artur Andrade, Júlia Luiza Conceição, Genildo Gomes, Tayana Conte, and Bruno Gadelha. 2021. Diversidade de gênero and elicitação de requisitos: uso do GenderMag como estratégia de identificação de requisitos de usabilidade. In WER.

[4] Lecia J. Barker, Charlie McDowell,and Kimberly Kalahar. 2009. Exploring factors that influence Computer Science introductory course students to persist in the major. In Proceedings of the 40th ACM Technical Symposium on Computer Science Education (Chattanooga, TN, USA) (SIGCSE '09). Association for Computing Machinery, New York, NY, USA, 153–157. DOI: https://doi.org/10.1145/1508865.1508923.

[5] Alexandra Barringer, Lauren M. Papp, and Pamela Gu. 2022. College students' sense of belonging in times of disruption: prospective changes from before to during the COVID-19 pandemic. Higher Education Research & Development, 1-14.

[6] Theresa Beaubouef, and John Mason. 2005. Why the high attrition rate for Computer Science students: Some thoughts and observations. Inroads: The SIGCSE Bulletin, Vol. 37, No. 2, June 2005, pp. 103-106. DOI: https://doi.org/10.1145/1083431.1083474

[7] Olga Den Besten. 2010. Local belonging and 'geographies of emotions': Immigrant children's experience of their neighbourhoods in Paris and Berlin. Childhood 17, 2 (2010), 181–195.

[8] Brianna Blaser, Katherine M. Steele, and Sheryl E. Burgstahler. 2015. Including universal design in engineering courses to attract diverse students. In 2015 ASEE Annual Conference & Exposition. ASEE Conferences, Seattle, Washington. https://www.jee.org/24272.

[9] Benjamin S. Bloom, Max D. Engelhart, Edward J. Furst, Walker H. Hill, and David R. Krathwohl. 1956. Taxonomy of Educational Objectives. Longmans, Green and Co LTD, London, England.

[10] Corinne Brion and Paula A. Cordeiro. 2018. Lessons learned from a training of trainers model in Africa. Journal of Educational Leadership and Policy Studies. 2, 1. Retrieved from https://files.eric.ed.gov/fulltext/EJ1227037.pdf.

[11] Matthew H. E. M. Browning, Lincoln R. Larson, Iryna Sharaievska, Alessandro Rigolon, Olivia McAnirlin, Lauren Mullenbach, Scott Cloutier, Tue M. Vu, Jennifer Thomsen, Nathan Reigner, Elizabeth Covelli Metcalf, Ashley D'Antonio, Marco Helbich, Gregory N. Bratman, Hector Olvera Alvarez. 2021. Psychological impacts from COVID-19 among university students: Risk factors across seven states in the United States. PloS One 16(1), e0245327.

[12] Michael Buckley, Helene Kershner, Kris Schindler, Carl Alphonce, and Jennifer Braswell. 2004. Benefits of using socially-relevant projects in computer science and engineering education. In Proceedings of the 35th SIGCSE Technical Symposium on Computer Science Education. 482–486.

[13] Janet E. Burge, Gerald Gannod, Mike Carter, Alanna Howard, Brian Schultz, Mladen Vouk, David Wright & Paul Anderson. 2014. Developing CS/SE students' communication abilities through a program-wide framework. In Proceedings of the 45th ACM Technical Symposium on Computer Science Education (pp. 579-584).

[14] Margaret Burnett, Simone Stumpf, Jamie Macbeth, Stephann Makri, Laura Beckwith, Irwin Kwan, Anicia Peters, and William Jernigan. 2016. GenderMag: A method for evaluating software's gender inclusiveness. Interacting with Computers 28, 6 (10 2016), 760–787. DOI: https://doi.org/10.1093/iwc/iwv046 http://arXiv:https://academic.oup.com/iwc/article-pdf/28/6/760/7919992/iwv046.pdf

[15] Margaret Burnett, Robin Counts, Ronette Lawrence, and Hannah Hanson. 2017. Gender HCl and Microsoft: Highlights from a longitudinal study. In 2017 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC). 139–143. DOI: https://doi.org/10.1109/VLHCC.2017.8103461

[16] Brenda M. Capobianco and Allan Feldman. 2006. Promoting quality for teacher action research: Lessons learned from science teachers' action research, Educational Action Research Vol. 14, No. 4, December 2006, 497–512.

[17] Center for Disease Control and Prevention. 2021. Understanding the Training of Trainers Model. Retrieved August 24, 2021 from https://www.cdc.gov/healthyschools/tths/train_trainers_model.htm

[18] Rachel Challen. 2019. Creating a sense of belonging and connectedness for the student arrival experience in a school of arts and humanities. Social Media in Higher Education: Case Studies, Reflections and Analysis (2019), 15.

[19] Shuo Chang, Vikas Kumar, Eric Gilbert, and Loren G Terveen. 2014. Specialization, homophily, and gender in a social curation site: Findings from Pinterest. In Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing. 674–686.

[20] Gary Charness and Uri Gneezy. 2012. Strong evidence for gender differences in risk taking. Journal of Economic Behavior & Organization 83, 1 (2012), 50–58.

[21] Amreeta Chatterjee, Lara Letaw, Rosalinda Garcia, Doshna Umma Reddy, Rudrajit Choudhuri, Sabyatha Satish Kumar, Patricia Morreale, Anita Sarma, and Margaret Burnett. 2022. Inclusivity bugs in online courseware: A field study. In Proceedings of the 2022 ACM Conference on International Computing Education Research V.1 (ICER 2022), August 7–11, 2022, Lugano and Virtual Event, Switzerland. 17 pages. DOI: https://doi.org/10.1145/3501385.3543973

[22] Alison Clabaugh, Juan F. Duque, and Logan J. Fields. 2021. Academic stress and emotional well-being in United States college students following onset of the COVID-19 pandemic. Frontiers in Psychology: 765.

[23] Sydney Clark. 2019. Why students want more from college diversity initiatives, Best Colleges, https://www.bestcolleges.com/blog/student-diversity-intitiatives/ (retrieved April 12, 2023).

[24] John Clarkson, Roger Coleman, Simeon Keates, and Cherie Lebbon. 2013. Inclusive Design: Design for the Whole Population, Springer.

[25] Lena Cohen, Heila Precel, Harold Triedman, and Kathi Fisler. 2021. A new model for weaving responsible computing into courses across the CS curriculum. In Proceedings of the 52nd ACM Technical Symposium on Computer Science Education. 858–864.

[26] Ben Coleman and Matthew Lang. 2012. Collaboration across the curriculum: a disciplined approach to developing team skills. In Proceedings of the 43rd ACM Technical Symposium on Computer Science Education (SIGCSE '12). 277–282. DOI: https://doi.org/10.1145/2157136.2157220[20]

[27] Sarah Cooney. 2021. Riding the bus in Los Angeles: Creating cultural micro-exposures via technology. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. Number 13. Association for Computing Machinery, 1-9. DOI: https://doi.org/10.1145/3411763.3450392

[28] CSUSB (Univ. California Center for the Study of Hate & Extremism), Fact Sheet: Anti-Asian Prejudice March 2020. March 2021. https://www.csusb.edu/sites/default/files/FACT%20SHEET-%20Anti-Asian%20Hate%202020%203.2.21.pdf Date of last access: Sept. 1, 2022.

[29] Sally Jo Cunningham, Annika Hinze, and David M. Nichols. 2016. Supporting gender-neutral digital library creation: A case study using the GenderMag Toolkit. In Digital Libraries: Knowledge, Information, and Data in an Open Access Society, Atsuyuki Morishima, Andreas Rauber, and Chern Li Liew (Eds.). Springer International Publishing, Cham, 45–50

[30] Debzani Deb, Muztaba Fuad, & Keith Irwin 2019. A module-based approach to teaching big data and cloud computing topics at CS undergraduate level. In Proceedings of the 50th ACM Technical Symposium on Computer Science Education (pp. 2-8).

[31] Brianna Dym, Namita Pasupuleti, Cole Rockwood, and Casey Fiesler. 2021. "You don't do your hobby as a job": Stereotypes of computational labor and their implications for CS education. In Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCE '21). 823–829. https://doi.org/10.1145/3408877.3432396

[32] L. E. Enriquez, A. E., Rodriguez, V. E Morales, K. Chavarria, & A. Ro. 2023. Mental health and COVID-19 pandemic stressors among Latina/o/x college students with varying self and parental immigration status. Journal of Racial and Ethnic Health Disparities, 10(1), 282-295.

[33] Casey Fiesler, Natalie Garrett, and Nathan Beard. 2020. What do we teach when we teach tech ethics? A syllabi analysis. In Proceedings of the 51st ACM Technical Symposium on Computer Science Education. 289–295.

[34] Catherine Good, Aneeta Rattan, and Carol S Dweck. 2012. Why do women opt out? Sense of belonging and women's representation in mathematics. Journal of Personality and Social Psychology 102, 4.

[35] Colin M. Gray, Shruthi Sai Chivukula, Kassandra Melkey, and Rhea Manocha. 2021. Understanding "Dark" design roles in computing education. In Proceedings of the 17th ACM Conference on International Computing Education Research (ICER 2021), August 16–19, 2021, Virtual Event, USA. ACM, New York, NY, USA, 14 pages. DOI: https://doi.org/10.1145/3446871.3469754

[36] Barbara J. Grosz, David G. Grant, Kate Vredenburgh, Jeff Behrends, Lily Hu, Alison Simmons, and Jim Waldo. 2019. Embedded EthiCS: Integrating ethics across CS education. Communications of the ACM, 62, 8, 54-61. Retrieved from https://cacm.acm.org/magazines/2019/8/238345-embedded-ethics/fulltext

[37] Hyeouk Chris Hahm, Yoonsook Ha, Judith C. Scott, Venissala Wongchai, Justin A. Chen, and Cindy H. Liu. 2021. Perceived COVID-19-related anti-Asian discrimination predicts post traumatic stress disorder symptoms among Asian and Asian American young adults. Psychiatry Research 303: 114084.

[38] Thomas A Harris. 1967. I'm OK, you're OK. Random House.

[39] Gillian R Hayes. 2014. Knowing by doing: action research as an approach to HCI. In Ways of Knowing in HCI. Springer, 49–68.

[40] Suzanne Hidi. 2016. Revisiting the role of rewards in motivation and learning: Implications of neuroscientific research. Educ. Psychol. Rev. 28, 1, 61–93. DOI:https://doi.org/10.1007/s10648-015-9307-5

[41] Claudia Hilderbrand, Christopher Perdriau, Lara Letaw, Jillian Emard, Zoe Steine-Hanson, Margaret Burnett, and Anita Sarma. 2020. Engineering gender-inclusivity into software: Ten teams' tales from the trenches. In 2020 IEEE/ACM 42nd International Conference on Software Engineering (ICSE). 433–444.

[42] History.com. 2020. A&E Television Networks. History: 2020 Events. Retrieved August 31, 2022, from https://www.history.com/topics/21st-century/2020-events#race-and-social-justice

[43] History.com. 2021. A&E Television Networks. History: 2021 Events. Retrieved August 31, 2022, from https://www.history.com/topics/21st-century/2021-events#disasters-violence

[44] Sonya Douglass Horsford, Leana Cabral, Cami Touloukian, Siettah Parks, Phillip A. Smith, Chy McGhee, Fawziah Qadir, Domonique Lester, & Jade Jacobs. 2021. Black education in the wake of Covid-19 & systemic racism. Black Education Research Collective: New York, NY, USA, 30.

[45] Lauren C. Howard, Evan A. Krueger, Joshua O. Barker, Tess Boley Cruz, Sam N. Cwalina, Jennifer B. Unger, Jessica L. Barrington-Trimis, & Adam M. Leventhal. 2022. Young adults' distress about police brutality following the death of George Floyd. Youth & Society. DOI: https://doi.org/1 0.1177/0044118X221087282.

[46] Hui-Ching Kayla Hsu and Nasir Memo. 2021. Crossing the bridge to STEM: Retaining women students in an online CS conversion program. ACM Transactions on Computing Education (TOCE) 21.2: 1-16.

[47] Lin Jia, Yasmine N. Elglaly, Catherine M. Baker, and Kristen Shinohara. 2021. Infusing accessibility into programming courses. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. Number 231. Association for Computing Machinery, New York, NY, USA, 1ś6. DOI: https://doi.org/10.1145/3411763.3451625

[48] Sami Kahn, Janet Vertesi, Sigrid Adriaenssens, Julia Byeon, Mona Fixdal, Kelly Godfrey, Jérémie Lumbroso, and Kasey Wagoner. 2022. The impact of online STEM teaching and learning during COVID-19 on underrepresented students' self-efficacy and motivation breadcrumb. Journal of College Science Teaching 51, no. 6.

[49] Tanjila Kanij, John Grundy, Jennifer McIntosh, Anita Sarma, and Gayatri Aniruddh. 2022. A new approach towards ensuring gender inclusive SE job advertisements. In 2022 IEEE/ACM 44th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS), pp. 1-11. IEEE.

[50] Brian TaeHyuk Keum, and Michele J. Wong. 2023. COVID-19 anti-Asian racism, perceived burdensomeness, thwarted belongingness, and suicidal ideation among Asian American Emerging Adults. International Review of Psychiatry, 1-8.

[51] René F. Kizilcec, & Sherif Halawa. 2015. Attrition and achievement gaps in online learning. In Proceedings of the Second (2015) ACM Conference on Learning @ Scale. 57-66.

[52] René F. Kizilcec and Andrew J. Saltarell. 2019. Psychologically inclusive design: cues impact women's participation in STEM education. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

[53] Ю. С. Кириллова and Д. А. Малы. 2017. Gender accessories recognition of user web-applications by classifiers. Аллея науки 4, no. 9: 854-857.

[54] Maria Klawe. 2013. Increasing female participation in computing: The Harvey Mudd College story. Computer 46, 3 (2013), 56–58.

[55] Amy J. Ko, Alannah Oleson, Neil Ryan, Yim Register, Benjamin Xie, Mina Tari, Matthew Davidson, Stefania Druga, and Dastyni Loksa. 2020. It is time for more critical CS education. Commun. ACM 63, 11, 31–33.

[56] Julie Kracov, Chad Cardin. 2020. CBSNews.com. Retrieved August 31, 2022, from https://www.cbsnews.com/news/2020-the-year-in-review-top-news-stories-month-by-month/

[57] Sophia Krause-Levy et al. 2021. "The relationship between sense of belonging and student outcomes in CS1 and beyond." Proceedings of the 17th ACM Conference on International Computing Education Research.

[58] Sang Gyu Kwak and Jong Hae Kim. 2017. Central limit theorem: the cornerstone of modern statistics. Korean Journal of Anesthesiology Vol. 70(2) 144-156. doi:10.4097/kjae.2017.70.2.144.

[59] Margaretha Vreeburg Izzo and William M Bauer. 2015. Universal design for learning: Enhancing achievement and employment of STEM students with disabilities. Universal Access in the Information Society 14, 1 (2015), 17–27

[60] Lara Letaw, Rosalinda Garcia, Heather Garcia, Christopher Perdriau, and Margaret Burnett. 2021. Changing the online climate via the online students: Effects of three curricular interventions on online CS students' inclusivity. In Proceedings of the 14th ACM Conference on International Computing Education Research (ICER 2021), August 16–19, 2021. 18 pages. DOI: https://doi.org/10.1145/3446871.3469742

[61] Lara Letaw, Rosalinda Garcia, Patricia Morreale, Gail Verdi, Heather Garcia, Geraldine Jimena Noa, Spencer P. Madsen, Maria Jesus Alzugaray-Orellana, and Margaret Burnett. Educating educators to integrate inclusive design across a 4-year CS degree program. 2022. arXiv preprint arXiv:2209.02748.

[62] Kurt Lewin. 1952. Group decision and social change. Readings in social psychology. Newcombe and Hartley (Eds.), Henry Holt, New York (1952).

[63] Karyn Lewis, Jane G. Stout, Noah D. Finkelstein, Steven J. Pollock, Akira Miyake, Geoff L. Cohen, and Tiffany A. Ito. 2017. Fitting in to move forward: Belonging, gender, and persistence in the physical sciences, technology, engineering, and mathematics (pSTEM). Psychology of Women Quarterly 41, no. 4: 420-436.

[64] Stephanie Ludi. 2007. Introducing accessibility requirements through external stakeholder utilization in an undergraduate requirements engineering course. In 29th International Conference on Software Engineering (ICSE'07). 736ś743. DOI: https://doi.org/10.1109/ICSE.2007.46 ISSN: 1558-1225.

[65] Thomas Mahatody, Mouldi Sagar, Christophe Kolski. 2010. State of the art on the Cognitive Walkthrough method, its variants and evolutions, International Journal of Human-Computer Interaction, 26 (8), pp. 741-785.

[66] Jennifer Mankoff. 2006. Practical service learning issues in HCI. In CHI'06 Extended Abstracts on Human Factors in Computing Systems. 201–206.

[67] Robert Marston, Chad Cardin. 2021. CBSNews.com. Retrieved August 31, 2022, from https://www.cbsnews.com/news/the-year-in-review-top-news-stories-of-2021-month-by-month/

[68] Robin McTaggert. 1997. Guiding principles for participatory action research, in Participatory Action Research: International Contexts and Consequences, (Robin McTaggert, editor), State Univ. New York Press, 25-44.

[69] Barbara Means and Julie Neisler. 2021. "Teaching and learning in the time of COVID: The student perspective." Online Learning 25.1.

[70] Joseph E. Michaelis and David Weintrop. 2022. Interest development theory in computing education: A framework and toolkit for researchers and designers. ACM Trans. Comput. Educ. 22, 4, Article 43 (December 2022), 27 pages. https://doi.org/10.1145/3487054

[71] Catherine Mooney and Brett A. Becker. 2021. Investigating the impact of the COVID-19 pandemic on computing students' sense of belonging. ACM Inroads 12, no. 2, 38-45.

[72] Mozilla Foundation. Responsible Computer Science Challenge. Accessed 9/5/2019. http://responsiblecs.org

[73] Sarah Nosek. 2023. "The impact of the Covid-19 pandemic on college students with marginalized identities." Equity in Education & Society: 27526461221149035.

[74] Alannah Oleson, Christopher Mendez, Zoe Steine-Hanson, Claudia Hilderbrand, Christopher Perdriau, Margaret Burnett, and Amy J. Ko. 2018. Pedagogical content knowledge for teaching inclusive design. In Proceedings of the 2018 ACM Conference on International Computing Education Research (Espoo, Finland) (ICER '18). Association for Computing Machinery, New York, NY, USA, 69–77. DOI: https://doi.org/10.1145/3230977.3230998

[75] Alannah Oleson, Meron Solomon, and Amy J. Ko. Computing students' learning difficulties in HCI education. 2020. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1-14.

[76] Alannah Oleson, Meron Solomon, Christopher Perdriau, and Amy J. Ko. 2022. Teaching inclusive design skills with the CIDER assumption elicitation technique. ACM Trans. Comput.-Hum. Interact. Just Accepted (May 2022). DOI: https://doi.org/10.1145/3549074

[77] Susmita Hema Padala, Christopher John Mendez, Luiz Felipe Dias, Igor Steinmacher, Zoe Steine Hanson, Claudia Hilderbrand, Amber Horvath, Charles Hill, Logan Dale Simpson, Margaret Burnett, Marco Gerosa, Anita Sarma. 2020. How gender-biased tools shape newcomer experiences in OSS projects. IEEE Transactions on Software Engineering (2020).

[78] Krystle Phirangee and Alesia Malec. 2017. Othering in online learning: An examination of social presence, identity, and sense of community. Distance Education 38, 2 (2017), 160–172

[79] Marshalita S. Peterson. 2014. Public deliberation and practical application of civic engagement through a “train the trainer” process at a historically black college. The Journal of Negro Education, 83, 1, 77-92. Retrieved from https://www.jstor.org/stable/10.7709/jnegroeducation.83.1.0077

[80] Pew Research Center, 2020. On the Cusp of Adulthood and Facing an Uncertain Future: What We Know About Gen Z So Far. https://www.pewresearch.org/social-trends/2020/05/14/on-the-cusp-of-adulthood-and-facing-an-uncertain-future-what-we-know-about-gen-z-so-far-2/ (retrieved April 12, 2023).

[81] Vahab Pournaghshband and Paola Medel. 2020. Promoting diversity-inclusive computer science pedagogies: A multidimensional perspective. In Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education. 219–224.

[82] Cynthia Putnam, Maria Dahman, Emma Rose, Jinghui Cheng, and Glenn Bradford. 2016. Best practices for teaching accessibility in university classrooms: cultivating awareness, understanding, and appreciation for diverse users. ACM Transactions on Accessible Computing (TACCESS) 8, 4 (2016), 1–26.

[83] Yim Register and Amy J. Ko. 2020. Learning machine learning with personal data helps stakeholders ground advocacy arguments in model mechanics. In Proceedings of the 2020 ACM Conference on International Computing Education Research (ICER '20). Association for Computing Machinery, New York, NY, USA, 67-78. DOI: https://doi.org/10.1145/3372782.3406252

[84] Shamima Nasrin Runa, Brett A. Becker, and Catherine Mooney. Variations in sense of belonging in undergraduate computing students through the COVID-19 pandemic. 2022. Proceedings of the 2022 Conference on United Kingdom & Ireland Computing Education Research.

[85] John P. Salerno, Jackson Devadas, M. Pease, Bryanna Nketia, & Jessica N. Fish. 2020. Sexual and gender minority stress amid the COVID-19 pandemic: Implications for LGBTQ young persons' mental health and well-being. Public Health Reports, 135(6), 721-727.

[86] Ben R. Shapiro, Emma Lovegall, Amanda Meng, Jason Borenstein & Ellen Zegura. 2021. Using role-play to scale the integration of ethics across the Computer Science curriculum. In Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (pp. 1034-1040).

[87] Arun Shekhar and Nicola Marsden. 2018. Cognitive walkthrough of a learning management system with gendered personas. In Proceedings of the 4th Conference on Gender & IT (Heilbronn, Germany) (GenderIT '18). Association for Computing Machinery, New York, NY, USA, 191–198. DOI: https://doi.org/10.1145/3196839.3196869

[88] Lee Shulman. 1987. Knowledge and teaching: Foundations of the new reform. Harvard educational review 57, 1 (1987), 1–23.

[89] Angela A. Siegel, Mark Zarb, Emma Anderson, Brent Crane, Alice Gao, Celine Latulipe, Ellie Lovellette, Fiona McNeill, and Debbie Meharg. 2022. "The impact of COVID-19 on the CS student learning experience: How the pandemic has shaped the educational landscape." In Proceedings of the 2022 Working Group Reports on Innovation and Technology in Computer Science Education, pp. 165-190.

[90] Steven E Stemler. 2004. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. Practical Assessment, Research, and Evaluation 9, 1 (2004), 4.

[91] Chris Stephenson, Alison Derbenwick Miller, Christine Alvarado, Lecia Barker, Valerie Barr, Tracy Camp, Carol Frieze, Colleen Lewis, Erin Cannon Mindell, Lee Limbird, Debra Richardson, Mehran Sahami, Elsa Villa, Henry Walker, and Stuart Zweben. 2018. Retention in Computer Science undergraduate programs in the U.S.: Data challenges and promising interventions. New York, NY. ACM.

[92] Jane G. Stout and Heather M. Wright. 2016. Lesbian, gay, bisexual, transgender, and queer students' sense of belonging in computing: An intersectional approach. Computing in Science & Engineering 18, no. 3, 24-30.

[93] Ernest T. Stringer. 2007. Action Research. Sage.

[94] Simone Stumpf, Anicia Peters, Shaowen Bardzell, Margaret Burnett, Daniela Busse, Jessica Cauchard, and Elizabeth Churchill. 2020. Gender-inclusive HCI research and design: A conceptual review. Foundations and Trends in Human–Computer Interaction 13, 1 (2020), 1–69.

[95] Charlotte Chengyun Teng, Sophia Hon, Angela Wang, and William Tsai. 2022. "Impact of COVID-19 discrimination fear on psychological distress among East Asian college students: The moderating role of emotion regulation." American journal of orthopsychiatry.

[96] Jannet P. Trammell, Nataria T. Joseph, & Jennifer A. Harriger. 2021. Racial and ethnic minority disparities in COVID-19 related health, health beliefs and behaviors, and well-being among students. Journal of American College Health, 1-7.

[97] U.S. Government Accountability Office. 2022. Women in the Workforce: The Gender Pay Gap Is Greater for Certain Racial and Ethnic Groups and Varies by Education Level, GAO-23-106041. https://www.gao.gov/products/gao-23-106041 Date of last access: April 16, 2023.

[98] Mihaela Vorvoreanu, Lingyi Zhang, Yun-Han Huang, Claudia Hilderbrand, Zoe Steine-Hanson, and Margaret Burnett. 2019. From gender biases to gender-inclusive design: an empirical investigation. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, 14 pages. DOI: https://doi.org/10.1145/3290605.3300283

[99] Annalu Waller, Vicki L Hanson, and David Sloan. 2009. Including accessibility within and beyond undergraduate computing courses. In Proceedings of the 11th international ACM SIGACCESS Conference on Computers and Accessibility. 155–162.

[100] Markeya S. Peteranetz, Shiyuan Wang, Duane F. Shell, Abraham A. Flanigan, & Leen-Kiat Soh. 2018. Examining the impact of computational creativity exercises on college computer science students' learning, achievement, self-efficacy, and creativity. In Proceedings of the 49th ACM Technical Symposium on Computer Science Education, 155-160.

[101] Etienne Wenger-Trayner and Beverly Wenger-Trayner. 2015. Communities of practice. A brief introduction. Retrieved from: https://wenger-trayner.com/introduction-to-communities-of-practice/. Date of last retrieval: January 30, 2023.

[102] Linda L. Werner, Brian Hanks, and Charlie McDowell. 2004. Pair-programming helps female computer science students. Journal on Educational Resources in Computing (JERIC) 4, 1, 4 pages.

[103] Richmond Y. Wong and Tonya Nguyen. 2021. Timelines: A world-building activity for values advocacy. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Number 616. Association for Computing Machinery, New York, NY, USA, 1ś15. DOI: https://doi.org/10.1145/3411764.3445447

[104] C. Wu, Qian, Y., & Wilkes, R. 2021. Anti-Asian discrimination and the Asian-white mental health gap during COVID-19. Ethnic and Racial Studies, 44(5), 819-835.

[105] Julia Yates, & Anke C. Plagnol. 2022. Female computer science students: A qualitative exploration of women's experiences studying computer science at university in the UK. Education and Information Technologies, 27(3), 3079-3105.

[106] 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. 1053–1059.

[107] Nick Young and Shriram Krishnamurthi. 2021. Early post-secondary student performance of adversarial thinking. In Proceedings of the 14th ACM Conference on International Computing Education Research (ICER 2021), Virtual Event, 12 pages. DOI: https://doi.org/10.1145/3446871.3469743

[108] Zippia. 2023. 25+ Telling Diversity In High Tech Statistics [2023]: Tech Demographics + Trends. https://www.zippia.com/advice/diversity-in-high-tech-statistics/ Date of last access: April 16, 2023.