Beyond “Awareness”: If We Teach Inclusive Design, Will Students Act On It?

Rosalinda Garcia
Oregon State University
garciros@oregonstate.edu

Patricia Morreale
Kean University
pmorreal@kean.edu

Pankati Patel
Kean University
patpanka@kean.edu

Jimena Noa Guevara
Oregon State University
noaguevg@oregonstate.edu

Dahana Moz Ruiz
Kean University
mozruizd@kean.edu

Sabyatha Satish Kumar
Oregon State University
satishs@oregonstate.edu

Prisha Velhal
Oregon State University
velhalp@oregonstate.edu

Alec Busteed
Oregon State University
busteeda@oregonstate.edu

Margaret Burnett
Oregon State University
burnett@eecs.oregonstate.edu

ABSTRACT

Motivation: Many university CS programs have begun teaching various types of CS-related societal issues using approaches such as ethics, Responsible CS, inclusive design, and more. However, some recent research suggests that, although these programs have been able to teach awareness, students often fail to act upon this awareness. To address this problem, University X’s CS program tried an unusual approach—integrating hands-on inclusive design skills in small ways across all four years of the CS major. But did it work? That is, did the students who experienced this change across the major actually build more inclusive technology than the students who did not experience it?

Objectives: This paper aims to answer this through addressing two research questions: (RQ1): Did students who learned inclusive design across the curriculum *act* to create more inclusive software? (RQ2): How did inclusivity (or lack thereof) manifest in students’ projects?

Method: To investigate these RQs, we conducted a case study of 22 term-long CS projects built by 22 teams consisting of a total of 92 3rd- and 4th-year CS students. Half of the student teams had experienced courses that had integrated inclusive design and the other half had not. The inclusive design elements University X taught were those of the GenderMag inclusive design method, so evaluating the students’ term-long projects was done by GenderMag experts—industry-experienced UX and Software professionals with real-world GenderMag experience.

Results: The inclusiveness of students’ projects was higher Post-GenderMag, with fewer reports of inclusivity bugs and higher inclusivity ratings. Experts’ evaluations also revealed the ways in which bias (e.g. bias against risk-averse users) and inclusion (e.g. inclusion of users with diverse information processing styles) appeared in students’ projects.

Implications: We believe this to be the first published evidence that compares student-built technology’s inclusiveness before vs. after they have been taught inclusive design. These positive results suggest that teaching inclusive design across the curriculum can impact students beyond simply heightening *awareness*—moving them to *act* upon this new understanding by building technology that more inclusively serves a wider spectrum of society.

CCS CONCEPTS

• **Human-centered computing**; • **Applied computingEducation**;

KEYWORDS

GenderMag, diversity, inclusion, broadening participation, CS education, Inclusive Design, HCI education

ACM Reference Format:
Rosalinda Garcia, Patricia Morreale, Pankati Patel, Jimena Noa Guevara, Dahana Moz Ruiz, Sabyatha Satish Kumar, Prisha Velhal, Alec Busteed, and Margaret Burnett. 2024. Beyond “Awareness”: If We Teach Inclusive Design, Will Students Act On It?. In *ACM Conference on International Computing Education Research V.1 (ICER ’24 Vol. 1), August 13–15, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3632620.3671101

1 INTRODUCTION

If you build it, [they] will come—Field of Dreams [38]

The above quote from the famous movie, “Field of Dreams,” tells a farmer that if he builds a baseball field, a player or players will come and play baseball in it. Not just *watch* baseball. Not just *“be aware”* of baseball—but *actively* play it.

This paper considers an analogous scenario for Computer Science (CS) education by investigating the following take on the Field of Dreams quote: if CS educators ”build it”—i.e., if they *teach* CS students concrete skills across the CS curriculum to address CS-related societal issues—the students will *act* upon them.

But will they? Already, groups of CS educators have begun to teach students principles of ethics, Responsible CS, inclusive design, and so on, with the hope that their students will act upon these new

ICER ’24 Vol. 1, August 13–15, 2024, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-0475-8/24/08
https://doi.org/10.1145/3632620.3671101

principles. However, recent research has indicated that students struggle to act on their awareness of social issues in computing. For example, Gray et al. found that students, including UX Design students, struggled to apply ethical design and decision-making even when they were aware of the ethical concerns [13]. Similarly, Oleson et al found that HCI faculty and students reported that HCI students experienced difficulties in preventing bias in their work [30].

To begin addressing this problem, we devised an across-the-curriculum inclusive design approach that aspired to produce students who would take action based on what they learned [10]. We tried it out with the Computer Science Department at University X [9], an urban Hispanic-Serving university in the Eastern U.S. Under this approach, 13 CS faculty at University X taught elements of an inclusive design method (GenderMag, described in Section 2). However, they did not teach them all at once—instead, the elements were sprinkled throughout 12 courses in the 4-year CS degree [9].

The results were promising. The field study reported significantly improved retention, grades, and education climate [9]. However, in that work, we did not analyze whether the students who experienced this approach acted upon their new knowledge of inclusive design by *building more inclusive software* than previous students have done.

That is the question that this paper aims to answer. To do so, we gathered student produced data and analyzed it using empirical case study methodology [46]. Specifically, we collected students' final deliverables for their term-long projects from offerings of 5 courses that had been offered both the year before and the year in which the across-curriculum inclusive design approach had begun. The projects we collected were the work of 92 students, which they did in 22 teams of about 4 students each. Outside experts then evaluated the inclusivity of students' projects, via a rubric we provided. We used their results to answer the following research questions:

- RQ1: Did students who learned inclusive design across the curriculum *act* to create more inclusive software?
- RQ2: How did inclusivity (or lack thereof) manifest in students' projects?

There were many reasons why the answers to the above research questions could turn out differently from the ideal. Ideally, the answers to RQ1 would be "yes, students acted" and to RQ2 would illustrate a variety of students' inclusivity successes in the context of the software they built. However, (Reason 1) because inclusive design was not the main point of any particular course, students could *easily ignore* (or memorize and forget) the inclusive design skills they had been taught. In fact, (Reason 2) the projects were term-long, so even if a course had a specific inclusive design assignment for the project earlier in the term (with obvious implications that the students' grade depended on the use of inclusive design), the projects were iterative, changing from one deliverable to the next with no final evaluation of the inclusiveness. Thus, there was *no final grade-connected "reward"* for designing inclusively. Further, (Reason 3) even if students were inspired enough to act (RQ1), they *might not have learned enough* inclusive design under this approach to actually *succeed* at building more inclusive software.

Thus, in answering the above research questions, our paper makes the following contributions:

- The first evaluation of whether an across-curriculum approach to any form of inclusive design can *move students to act* upon their newfound knowledge;
- The first evaluation of whether students experiencing an across-curriculum approach to any form of inclusive design *gain skills enabling them to succeed* at improving the inclusivity of the software they build; and
- The first comparison of the inclusivity of projects designed and built by students who have learned any form of inclusive design with projects designed and built by students who have not learned inclusive design.

Positionality Statement: The authors of this paper span various races (including Asian, Latinx, and White) and backgrounds, and experience levels. As some of us are women, including women of color, members of our team have experienced the lack of diversity and inclusion in CS education and in technology itself firsthand. Two commonalities we share are an interest in leveraging HCI to improve CS education for diverse students, and an interest in changing CS education to produce software practitioners whose CS skills arm them to *act* to produce more inclusive software than past practitioners have done. In pursuing these goals, we recognize that our access to higher education and academic research places us in a position of privilege. These privileges provided us with the access and influence necessary to complete this work and we remain committed to using these privileges to make experiences with CS education and technology better for everyone.

2 BACKGROUND AND RELATED WORK

2.1 GenderMag

The inclusive design elements that students learned in their CS course(s) came from the GenderMag method. GenderMag is an existing evidence-based method to identify and fix inclusivity issues within software [3]. GenderMag can be used on any type of technology in which the user needs to think and problem-solve [3]. For example, it has been used to find and/or fix inclusivity issues in open-source software project sites [5, 15, 33], learning management systems [16, 40], a search tool [44], a digital library [8], a code review tool [28], a tech support site [16], educational games and classroom materials [3, 6] and on multiple projects at a large software company [4].

GenderMag is an analytical method that combines personas with a variant of the cognitive walkthrough [26] so, like other analytical methods, it does not require bringing in users to test interfaces. It uses questions similar to those in Spencer's streamlined cognitive walkthrough [42], but specializes them by focusing evaluators' attention on *why* a *particular* user might have difficulty with a particular action, especially focusing on the five GenderMag "facets":

Before a subgoal/action: *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?*

Table 1: Each of the GenderMag persona's facet value definitions [3].

	Abi	Pat	Tim
Motivations	Uses technology to accomplish tasks	Uses technology to accomplish tasks	Enjoys learning all available features
Computer Self-efficacy	Lower relative to peers	Medium	Higher relative to peers
Attitude Toward Risk	Risk-averse	Risk-averse	Risk tolerant
Information Processing Style	Comprehensive	Comprehensive	Selective
Learning Style (Process vs Tinkering)	Process-oriented	Tinkering, but mindfully	Tinkering, sometimes to excess

The five GenderMag "facets" capture five ranges of problem-solving attributes users bring to their use of technology: (1) Motivations for technology usage, (2) Information Processing Styles, (3) Computer Self-efficacy, (4) Learning Style[1], and (5) Attitude Toward Risk. These facets have ranges of values, which are captured by a set of GenderMag personas named Abi, Pat, and Tim as described in Table 1. For example, Abi's motivations are to accomplish the task at hand, which is one endpoint in the Motivations facet; Tim's motivations are to explore new technology features, which is the other endpoint in the Motivations facet. Pat provides a third set of values that are a mix of Tim's and Abi's with some unique values as well. More details on each facet will be presented in the pertinent results section.

The central idea behind GenderMag is that if a technology feature does not simultaneously support *both* ends of a facet value range, it is not inclusive—and the feature is said to have an "inclusivity bug." The bugs are problem-solver inclusivity bugs as they disproportionately affect individuals with specific problem-solving styles, but they are also gender inclusivity bugs because the facet values statistically cluster around genders [3, 43, 44].

2.2 Related Work on Teaching Inclusive Design

Recent research has pointed to the need to teach inclusive design elements in CS education. For example, Oleson et al. found that HCI students (and HCI faculty) reported 18 difficulties in learning design, including difficulties applying design and preventing bias in their projects [30]. To head off difficulties like these, many educational institutions have begun to teach forms of inclusive design to technology students. Much of this research falls under Responsible CS, ethics, and accessibility, all of which can contribute to building students' ability to design systems for diverse populations with stakeholders in mind. For example, many Responsible CS approaches have the goal of teaching students' mindfulness of societal implications of their work, including developing for diverse populations [7, 27]. One example of analogous research in the ethics domain is that of Grosz et al., who presented a new approach to include ethics throughout the CS curriculum to help students learn about the social impacts of their work, including inclusive design and equal opportunity considerations [14]. Similarly, Horton et al. created ethics modules for undergraduate CS and subsequently taught them in Intro CS, Software Engineering, Machine Learning, and Algorithms courses [17, 18]. Examples abound in the accessibility domain, aiming to raise students' awareness of barriers faced by users with challenged motor, cognitive, hearing, etc., abilities, and/or to teach students' accessibility programming skills, (e.g., [32]).

Research has shown that beneficial outcomes to students often ensue from these efforts. Some of these positive outcomes have been in students' awareness of the need for inclusive design. For example, researchers have found that including accessibility in CS courses can help raise student awareness and help students value the need to design for users with disabilities in mainstream software [19, 25]. Similarly, Shapiro et al.'s investigation of role-playing inclusivity exercise in first-year, fourth-year, and online graduate level CS courses suggested that the approach could help students identify multiple perspectives [39]. Kuang et al. integrated accessibility assignments into core CS courses and found improvements in students' confidence in implementing accessibility [21]. Oleson et al. developed the new CIDER (Critique, Imagine, Design, Expand, Repeat) design evaluation method for education settings, and found that it helped students both respond to bias and value inclusivity in their design work [31].

Other beneficial outcomes have related to the inclusivity of the education climate. For example, Blaser et al. proposed that the inclusion of inclusive or universal design may help include women and students with disabilities [2], especially as women may be more drawn to inclusive design [12]. Letaw et al. did an action research study of teaching small pieces of inclusive design in online database and software engineering courses, and reported increases in climate, awareness, and respect for users' diversity [23]. Our prior action research study on teaching snippets of inclusive design across the CS curriculum reported benefits in retention, students' grades, and climate [9].

However, reports of skill gains by students have been mostly informal impressions from faculty interviews or self-reports by students (e.g. [9, 20, 22, 23, 36, 41, 45]). For example, in Letaw et al.'s study [23], students reported having gained new inclusive design abilities. Watchhorn et al. investigated outcomes from teaching inclusive/universal design to architecture and occupational therapy students, using both online and face-to-face teaching methods, including virtual and real-life simulation activities [45]. In their investigation, students self-reported improved new learning gains related to universal design. Kearney-Volpe et al. evaluated the use of accessibility modules in technology courses as part of the

[1]We use "learning style" to refer to the GenderMag facet about learning new technologies via process versus via tinkering as opposed to the education community's use of the term "learning styles" indicating learning through different formats (auditory, visuals, etc.).

TeachAccess program and found improvements in students' self-reports of confidence in applying accessibility principles [20]. Both the Letaw study and our prior work also produced self-reports and faculty impressions of newly gained design skills [9, 23].

Still, a few researchers have gone beyond student and faculty impressions in their reports of students' skill gains, primarily in the accessibility domain. Ludi reported on the incorporation of stakeholder interactions in accessibility related projects within a requirements and specifications course to improve students' accessibility skills [25]. To evaluate, they looked at students' mentions of accessibility in their final exam questions about example requirements and found that students who had stakeholder interactions mentioned accessibility more. Bar-El and Worsley used a similar approach for a course on Inclusive Making with qualitative analysis of students' design process with and without stakeholders revealing that the stakeholders helped students improve their process and projects [1]. In a different approach, Levy and Gandy evaluated the impact of a 60-minute accessibility lecture on students' audio cue design for low vision users [24]. They presented students with a short pre-recorded recording of a videogame, prompted students to add audio cues, evaluated the accessibility of students' additions, and found that students who experience the lecture added more accessible audio cues.

Closest to our own work in teaching GenderMag elements is that of researchers who have begun to investigate how to teach inclusive design through incorporating GenderMag into CS courses both online and in-person. The first investigation into teaching with GenderMag was Oleson et al.'s investigation into how to teach inclusive design [29]. Their action research investigation into how HCI- oriented faculty teach GenderMag in face-to-face university CS classes produced 11 elements of inclusive PCK (Pedagogic Content Knowledge) that can be leveraged to teach inclusive design in courses. Letaw et al. then leveraged some of these PCK elements to teach portions of GenderMag in two online CS courses. The results revealed that learning GenderMag-based material helped students feel included and improved their interest in finishing the CS major [23]. We also leveraged Oleson et al.'s PCK elements and created a faculty-oriented curriculum on how to teach GenderMag, termed the "Matchmaker Curriculum" [10]. The Matchmaker Curriculum's goal is to enable faculty to pair up with teaching resources they deem suitable for their particular courses, their comfort levels, and their teaching styles.

Our field study into the use of the Matchmaker Curriculum with CS faculty found that it achieved its Learning Outcome goals with 88% of the faculty [10]. In Patel et al.'s interview study with faculty who had experienced the curriculum, faculty generally reported it was feasible to apply and had positive impacts on students' work practices [34]. W investigated several student outcomes from the courses these faculty members taught, and reported higher grades, better retention, and a more inclusive climate [9].

However, none of these works have compared whether the technology products such students created were indeed more inclusive than products built by students not educated in inclusive design. That is the gap this paper aims to fill.

3 METHODOLOGY

To answer our research questions, we conducted an empirical case study [46], which is a form of field study. In our case study, we collected student projects that had been created over several terms in 2020-2022, and asked a group of outside experts to evaluate the student projects' inclusivity. The projects were created by undergraduate CS students between Spring 2020 and Fall 2022 at University X, a regional Hispanic-serving institution in the Eastern U.S.

University X was where we used the Matchmaker Curriculum, so Fall 2021 was when faculty first taught their courses with GenderMag concepts. Thus, the Pre-GenderMag students took these courses prior to Fall 2021 and had never learned GenderMag. The Post-GenderMag students were enrolled from Fall 2021 onward and had learned elements of GenderMag through one or more classes taught by Matchmaker "alums" (faculty), who had learned how to teach those elements via the Matchmaker Curriculum.

As is common in empirical case studies [46], we had no control over what the students learned or which faculty member they learned it from. In fact, prior to teaching the Post-GenderMag students, each faculty member had chosen for themselves the elements of GenderMag they wanted to integrate into their courses [9, 10]. In total, the Post-GenderMag students were taught by 13 University X faculty integrating GenderMag elements in whatever ways they chose, over multiple sections of 12 CS courses spread across all four years of the CS major.

Still, there was a pattern to these 12 courses, by level. In early courses, such as CS0/CS1, faculty would *introduce* one to two elements of inclusive design. For example, an early programming project may be updated to introduce the GenderMag facets and ask students to reflect on their own facet values. In mid-level courses, faculty gradually added more *breadth and depth* of inclusive design to have students *apply* what they were learning in new ways. For example, in the mid-level Web design course, faculty guided students through GenderMag walkthroughs with pre-set use-cases/scenarios. Finally, in upper-level courses, such as the Software Engineering and Capstone courses, the students used GenderMag hands-on to evaluate the software they were creating for their courses. The Post-GenderMag students' projects we collected were from the upper-level integrations of GenderMag, and the Pre-GenderMag students' projects came from the same courses but before the GenderMag elements had been added.

3.1 Participants

92 student participants created the projects in our study. They were undergraduate students at University X, majoring in either the Computer Science (CS) department's CS program or its Information Technology (IT) program. These 92 students were enrolled in one or more of the following junior/senior (3rd- and 4th-year) courses: Software Engineering, Human-Computer Interaction, Senior Project, CS Capstone, and IT Capstone courses. We collected their completed projects from faculty who taught these courses and obtained students' consent to use their projects with an IRB-approved consent form. Student participants did not have to do any additional work beyond the course projects they had already completed. The 92 students created their projects in groups (22

groups total) with an average size of 4 students. We describe these projects in Section 3.3.

The other type of participants was GenderMag experts, whose task was to evaluate the student projects' inclusivity. The GenderMag expert participants were UX and Software professionals who had between 1 and 5 years of experience with GenderMag. We recruited experts through social media postings and emails to known GenderMag experts in the field. 8 experts agreed to participate and completed an IRB-approved consent form. One expert was unable to continue, leaving 7 experts who completed the evaluations. We detail the rubric experts used for their evaluations in Section 3.2.

3.2 The Experts' Instructions and Rubric

To familiarize the experts with the projects we assigned them, we provided the experts with the student-created project names, student-written descriptions, student-written scenario (the persona's main task or goal), persona's "given" current subgoal (stepping-stone goal that will help accomplish their main goal), and the next hoped-for action the persona "should" take (specific action that is the developers' intended way to progress). We gathered all of this information from students' artifacts (e.g. from the problem statements in their documentation or presentations) and chose the subgoal and action as described in Section 3.3. The instructions had experts assume the persona knew the description of the application, the scenario, and the subgoal—but not had no prior knowledge of the action to fulfill that subgoal. For example, the information we gave the experts for Project 12Post (Project number 12 from Post-GenderMag students) was:

> "App Name: Recipes 4 You
>
> App Description: A website where users can search for recipes and submit recipe suggestions.
>
> Scenario (Overall Goal): Find a recipe to make for dinner
>
> Subgoal: View dinner recipes
>
> Action: Click 'Recipes"'

To maintain consistency between the evaluations completed by each expert, we created an evaluation rubric by drawing upon elements of GenderMag and taking inspiration from the Letaw et al.'s methodology [23]. However, Letaw et al.'s context was different from ours: they gathered expert evaluations on projects from a single Software Engineering course, within-student progress before vs. after they had learned GenderMag elements. In contrast, our experts evaluated between-student projects from multiple courses that occurred in different years, and our investigation had different research questions, so Letaw et al.'s rubric was not entirely suitable for our projects. We detail the rubric we created below and include the full version in the Supplemental Documents [11].

Our rubric (Figure 1) was set directly into the standard Inclusivity Bug Report form used in GenderMag walkthroughs. A GenderMag walkthrough walks through a scenario with a series of subgoals and actions. To minimize expert workload and to have uniform sequences across all student projects, we set up the process to include only 2 screens from each project, which required experts to evaluate the project both before (rubric Question A) and after (rubric Question B) one action. As mentioned previously in Section 2, these questions were: before a user could take the action (rubric Question A: "Will <Persona> know what to do at this step? Why/what facets?") and after taking the action (rubric Question B: "If <Persona> does the right thing, will s/he know that s/he did the right thing and is making progress toward their goal? Why/what facets?").

To capture inclusivity through the lens of all three personas, we included Inclusivity Bug Reports for each persona. Each one was labeled with the persona name and image. Thus, for each student project, experts answered 3 sets of A+B questions (6 questions per project), plus the 4 additional questions we describe next, for a total of 10 questions per project.

For each project, the experts also gave overall ratings of the inclusivity, via three Likert-scale rating questions and one ranking question to the end of the rubric. The three Likert-scale questions prompted experts to rate how well the project worked for each of the three personas on a 5-point scale from "terrible" to "wonderful". The ranking question then asked them to rank the project's suitability in terms of which persona the project supported best, from 1-3. Figure 2 shows these questions, and the Supplemental Documents provides the full rubric [11].

3.3 Collecting and preparing the projects

We gathered students' projects from upper-level (junior/senior) Software Engineering, HCI, Capstone, and Senior Project courses. We chose these courses because they all included term-long projects that student teams design and implement throughout the term. The use of term-long projects was to minimize effects of some particular assignment emphasizing inclusive design, which might have caused students to view inclusive design as the "main point" of the assignment; term-long projects instead tend to emphasize *everything* taught during the term (e.g., all of the software engineering practices taught, etc.). The CS Department Chair assessed these courses' term-long projects to be of similar complexity:

> CS Department Chair (interview): *"<The projects are> about the same <complexity>. HCI might be a little junior. Capstone might be a little senior. But we've got people taking those courses in both their senior and junior year."*

We collected students' finalized project submissions from faculty who taught these courses. The students' final project artifacts varied, including some or all of presentation slides, presentation video recordings, documentation, and websites. From all the projects we received, we reduced to the 22 student projects that were "evaluable"—11 projects from Post-GenderMag students and 11 projects from Pre-GenderMag students. We defined "evaluable" as having the following characteristics: available consent, having readable screenshots, having simple contexts, and having Question A/B screenshots in sequential order.

For each project, the Question A screen we chose was the homepage of the project. This allowed us to exclude sign-up or sign-in workflows as not every project had these and to avoid experts needing additional context. Then we chose the scenario and subgoal. When possible, we used scenarios and subgoals described by students in their presentations or documentation. When students did

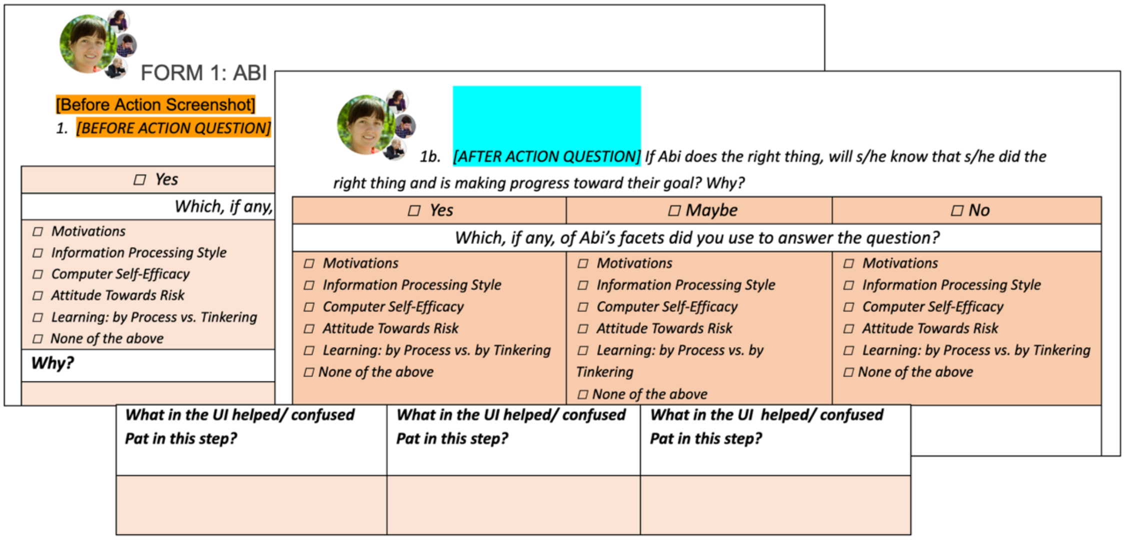

Figure 1: Rubric: The Inclusivity Bug Report the experts filled out. Snippets of (Left) Question A (labeled "1" in figure), and (Right) Question B (labeled "1b" in figure), both set up for the Abi persona. Experts could choose multiple facets as needed for their responses as is done in the full GenderMag method. (Bottom) This section appeared below both Question A and Question B, encouraging the experts to point to specific UI components involved in their judgments.

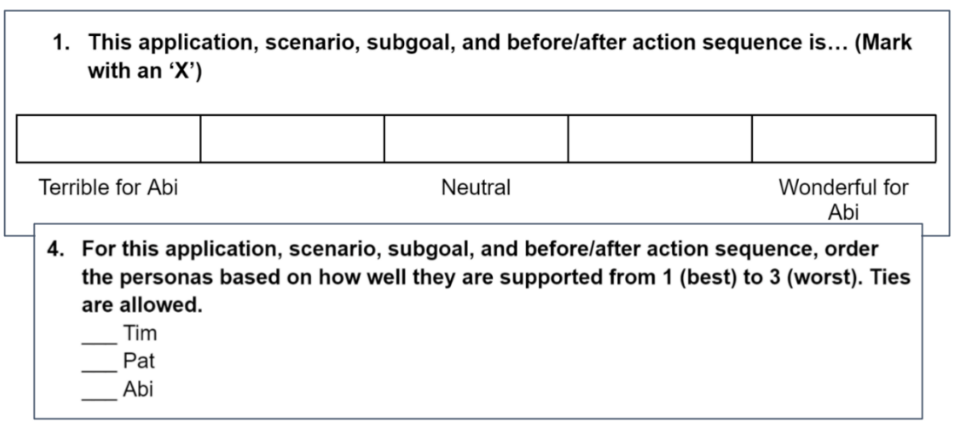

Figure 2: Rubric: (Top) The Likert scale rating question for the Abi persona. (Bottom) The persona ranking question.

not demonstrate specific scenarios/subgoals in their artifacts, we chose them based on students' descriptions of the primary purpose of the application. Based on the scenario, we then chose the action persona "should" take to make progress toward the subgoal. Finally, we chose the Question B screen to be the screen directly after the action takes place. For example, in Figure 3, the action is to click "Recipes" which navigates to another screen which we used as the Question B screen.

Finally, we redacted any information about the students, University X, or any real-life organization associated with the projects and filled out 22 rubrics with the project information and screenshots.

All screens, descriptions, subgoals, and actions we provided to experts can be found in each project rubric in the Supplemental Documents [11].

3.4 Project Assignment to Expert Evaluators, Data Collection, and Analysis

After asking the experts how many projects they were willing to evaluate, we randomly assigned each expert to their desired number of project rubrics. We provided experts with Google Drive folders containing the rubrics as Google Docs that they could complete in any order. We constrained the random assignment such that each project had two assigned experts, and each expert received

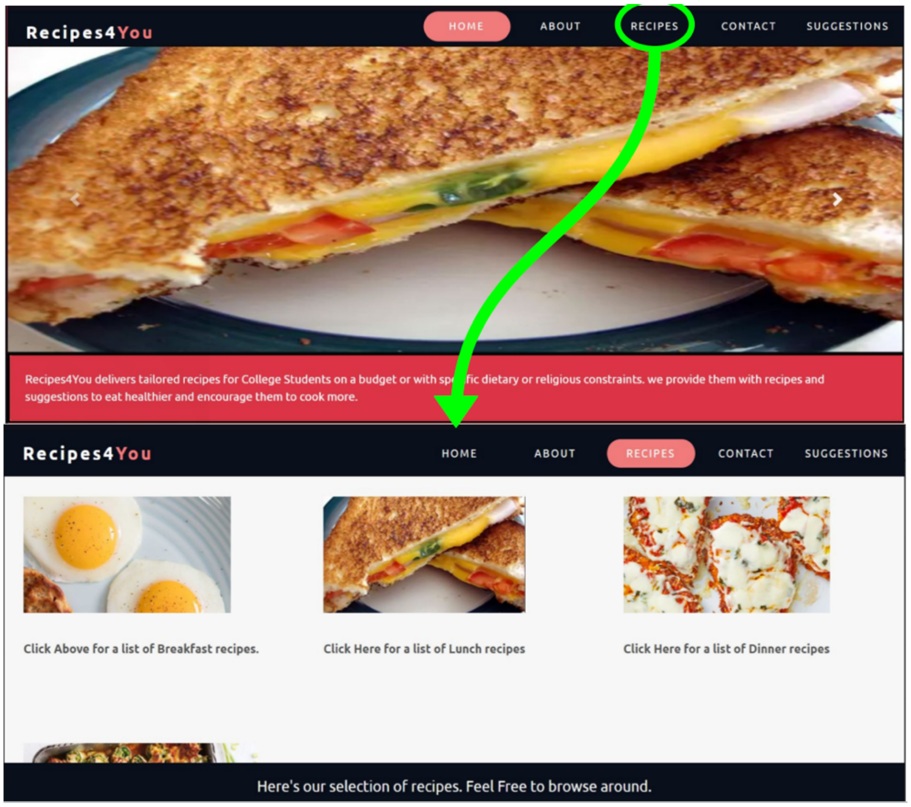

Figure 3: The Queston A (top) and Question B (bottom) screenshots from Project 12Post made by Senior Project students. They described their application as a website where students can search for recipes and submit recipe suggestions. We added a green circle and transition arrow to show that the action of clicking the recipe button navigates to the Question B screen.

approximately half of their assigned projects from Post-GenderMag students and half from Pre-GenderMag students. Experts did not know which projects were in which group or how many were from each group.

Experts returned their evaluations to us, but sometimes pairs of experts did not agree. We defined non-agreement as: the inclusivity ratings differing by more than 2 points for any of the three GenderMag personas. In these cases, we then assigned the project to a third expert from the original pool of 7 experts. By the end of all evaluations, the experts each evaluated 2-18 projects (mean: 8, median: 7), each expert evaluating approximately half from Post-GenderMag students and half from Pre-GenderMag students.

Our data analysis used a combination of descriptive statistics (counts, averages, etc.) and qualitative analysis. Qualitatively, we considered qualitative data in the form of experts' open answer responses to the “Why” and UI component sections of questions A and B. These data allowed us to understand the experts' evaluations and the inclusivity issues or lack thereof in the projects students created. We did not use inferential statistics because (1) as an empirical case study, our investigation did not use controls, so confounds abounded; and (2) even if we had wanted to do inferential statistics, none were suitable for our research questions because answering them would have mixed dependent data with independent data, which violates the assumption of inferential statistics. Instead, as with most case studies, we make extensive use of triangulation to safeguard the rigor of our results. (We return to these points in Section 5 Discussion.)

4 RESULTS

4.1 RQ1: Did learning GenderMag move students beyond “awareness” to action?

The experts' Likert-scale inclusivity ratings (recall Figure 2) directly answered RQ1—whether Post-GenderMag students acted upon their new inclusive design knowledge by building software that was more inclusive than that of the Pre-GenderMag students. As Figure 4 shows, the answer was yes: Post-GenderMag students created their projects more inclusively than the Pre-GenderMag students did for all three personas. Specifically, the Post-GenderMag projects were rated to be 10%, 6%, and 8% more inclusive than the Pre-GenderMag projects for Abi, Pat, and Tim, respectively.

The distribution of the projects' inclusivity rating scores[2], separated by persona, helped to reveal exactly where the Post-GenderMag students made progress and where they did not. For example, although it was already clear from Figure 4 that Abi-like

[2]As explained in the Methodology section, 2-3 experts evaluated each assignment, which totaled 30 inclusivity ratings of the Pre-GenderMag students' projects and 29 inclusivity ratings of the Post-GenderMag students' projects.

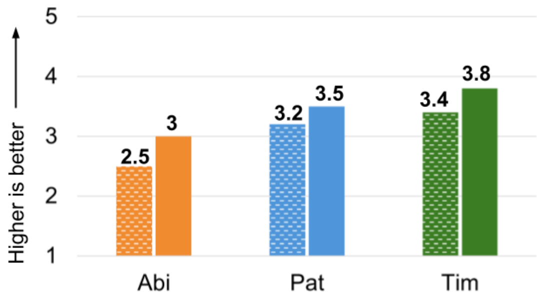

Figure 4: Average inclusivity rating per project, for the three personas, on a scale of 1 ("terrible") to 5 ("wonderful"). For all three personas, the Post-GenderMag students' projects had higher inclusivity ratings than the Pre-GenderMag students. Patterned bar=Pre-GenderMag students' projects (n=11), solid bar=Post-GenderMag (n=11); Orange=Abi, Blue=Pat Green=Tim.

users were the least included of the three personas, what the distributions of the experts' evaluations (Figure 5) add is whether students' projects changed by becoming more/less terrible (e.g., moving from terrible to neutral) or becoming more/less wonderful (e.g., moving from neutral to wonderful).

The Abi (left/orange) portion of the figure reveals that Post-GenderMag students were able to make progress on reducing the "terribleness" of the projects for Abi by 16%. Students also increased the "wonderfulness" of the projects for Abi, reflected by an 11% increase in "high" ratings. These terribleness and wonderfulness improvements for Abi of 16% and 11%, respectively, were larger than improvements for either of the other two personas (Pat: +1%(wrong direction) terribleness, -5%(wrong direction) wonderfulness; Tim -13% terribleness, +2% wonderfulness).

That the above inclusivity improvements were strongest for Abi-like users is important because, as all the figures in this section show, Abi-like users were the least supported in the Post-GenderMag and Pre-GenderMag students' projects alike. The fact that students struggled most to support the Abi persona was not surprising as prior GenderMag evaluations of professional software designs have also found lower support for Abi than for Tim [44]. Still, the Post-GenderMag Abi inclusivity rating average of 3 was lower than even the *Pre-GenderMag* ratings for Pat and Tim. Thus, although the Post-GenderMag inclusivity for Abi was higher than the Pre-GenderMag inclusivity, there is still much room for improvement.

For Abi, Post-GenderMag students' projects vs. Pre-GenderMag students' projects followed an "ideal" pattern: less terribleness and more wonderfulness. However, for Pat, this did not happen. As the middle (blue) section of Figure 5 shows, Pat's inclusivity "terribleness" was much lower than Abi's, but it was almost identical between students' Post-GenderMag and Pre-GenderMag projects—and the very small change (+1%) was in the wrong direction. The "wonderfulness" inclusivity for Pat changed by a larger amount (-5%)—again, in the wrong direction. Our data does not shed light on why this occurred, but note that the total change for Pat (Figure 4) was small, so the difference may come from combining the Low's into a single bucket for Figure 5. Another factor possibly explaining the smaller change in Pat than in the other two personas is that the "endpoints" of the facet ranges (described earlier in Section 2), which are represented by Abi and Tim, were more salient to the Post-GenderMag students, and thus easier to design for.

For Tim, the increase in wonderfulness was in the right direction (+2%) but was very small. One interpretation of the Tim results could be that the wonderfulness score was already fairly high (nearly 60% "wonderful), so bringing wonderfulness above this level might require more sophisticated inclusive design skills than students had yet gained. On the other hand, the Post-GenderMag students' projects still outshone the Pre-GenderMag students' projects on inclusivity for Tim, and the way they achieved this was by reducing terribleness, which dropped by a full 13%.

These results for all three personas were robust across different aggregation approaches and different measures. We have previously discussed the inclusivity ratings in Figure 4 and counts of

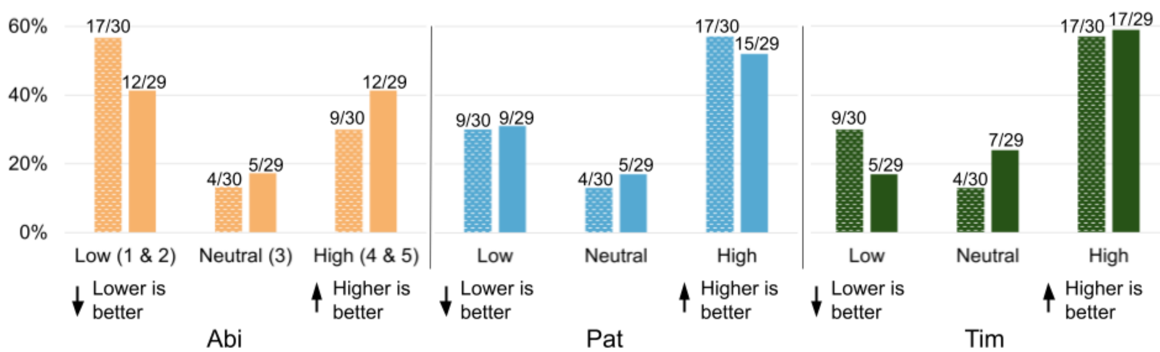

Figure 5: Counts of 59 Likert-scale inclusivity ratings of the student projects' "terribleness" to "wonderfulness," on a scale of 1 to 5 with the labels: 1=Terrible, 3=Neutral, 5=Wonderful. The distributions show that the Abi higher Post-GenderMag ratings shown earlier came from both less terribleness and more wonderfulness; the Tim higher Post-GenderMag ratings came from less terribleness (but not more wonderfulness); but where the Pat higher Post-GenderMag ratings came from is not revealed by these data. Patterned bar=Pre-GenderMag students' projects (100%=out of 30), solid bar=Post-GenderMag (100%=out of 29); Orange =Abi, Blue =Pat, Green =Tim.

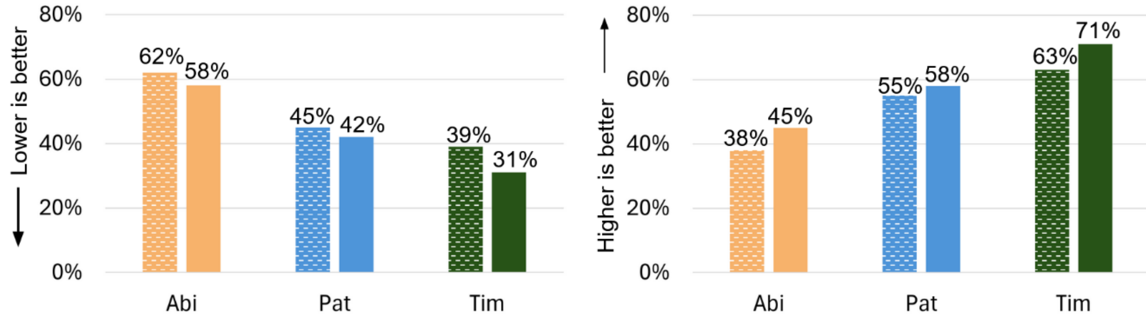

Figure 6: Inclusivity Bug Report responses. (Left): Average percent per project of expert's "no" and "maybe" responses (indicating questionable suitability to the persona) and (Right): "yes" responses (indicating suitability). For all three personas, Inclusivity Bug Reports showed fewer inclusivity bugs (left) and higher positive inclusivity outcomes (right) for the Post-GenderMag students' projects than for the Pre-GenderMag students' projects. Colors, patterns: Same as Figure 4.

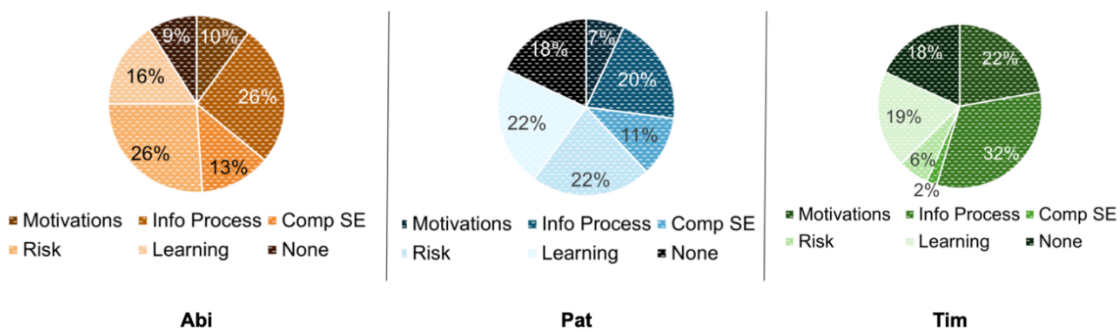

Figure 7: For the 11 Pre-GenderMag student projects, percentage each facet made up of the total facets experts used to explain inclusivity bugs reported on the Inclusivity Bug Reports for each project. 100%= average number of facets for "no"/"maybe" responses per project. In the Pre-GenderMag students' projects, the facets most troublesome for Abi (Info: 26%, Risk: 26%) were different from those for Pat (Risk: 22%, Learning: 22%) and different from those for Tim (Info: 32% and Motivations: 22%).

experts' high and low inclusivity ratings (Figure 5), the latter of which was consistent with Figure 4 for two of the three personas. In addition, averaging inclusivity ratings of the same Likert-scale questions by *expert* (not shown) gave nearly identical results to those in Figure 4's averages by *project*. Figure 4's results were also consistent with the experts' No/Maybe/Yes responses in the Inclusivity Bug Reports (recall Figure 1). Specifically, as the left bar chart in Figure 6 shows, the percentages of inclusivity bugs for all three personas were lower for Post-GenderMag students' projects than for Pre-GenderMag students' projects, and as the right bar chart shows, positive inclusivity outcomes were higher for Post-GenderMag students' projects than for Pre-GenderMag students' projects (4%, 3%, and 8% fewer inclusivity bugs for Abi, Pat, and Tim, respectively; and 7%, 3%, and 8% more positive outcomes).

4.2 RQ2: Inclusivity and Lack Thereof: The Facets Explain

4.2.1 The student projects' inclusivity bugs. In what ways were the inclusivity bugs problematic? Recall that Questions A and B on the Inclusivity Bug Report asked experts to report why they chose their answer and if there were associated facets (recall Table 1). Their responses included explicit uses of a facet 647 times. Their "maybe" and "no" responses identified the student projects' "inclusivity bugs"—aspects of the project with questionable suitability to one or more of the personas. These "maybe" and "no" responses usually pointed to the particular facet values that explained *why* the inclusivity bug was problematic. Figure 7 shows, by persona, which facet values students struggled to support in the Pre-GenderMag students' projects.

As discussed in Section 4.1, students' Post-GenderMag project were also not perfect—they still had a number of inclusivity bugs for all personas. For these students' projects, we show the facet values students struggled with in Figure 8.

Attitude Toward Risk and Information Processing Style were related to most of students' inclusivity bugs, as shown in Figures 7 and 8, for the Abi persona. The Attitude Toward Risk facet captures how people feel about using new technologies. For a variety of reasons, Abi is risk-averse about using new technologies. The Information Processing Style facet captures different ways people collect information. Abi's Information Processing Style is comprehensive, meaning they prefer to form an understanding of the problem by

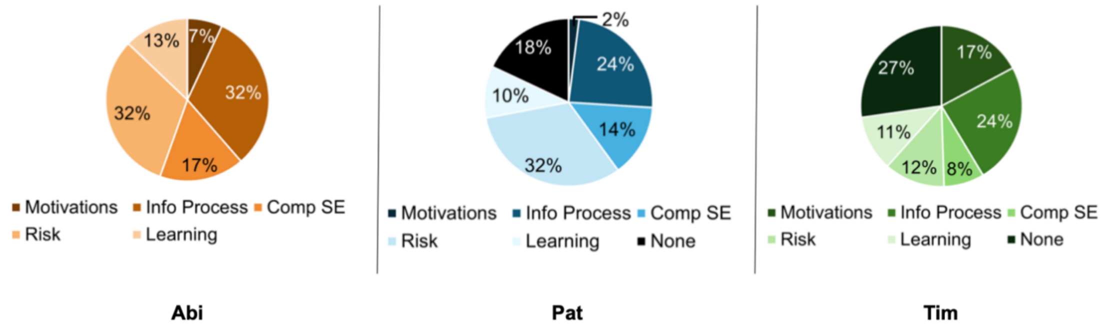

Figure 8: For the 11 Post-GenderMag student projects, average per assignment percentage of each facet experts used to explain their "no" and "maybe" responses. Risk and Information Processing together dominated the unsupported facets for Abi and Pat, with Motivations rarely mentioned. However, for Tim, Motivations were one of the top two unsupported facets.

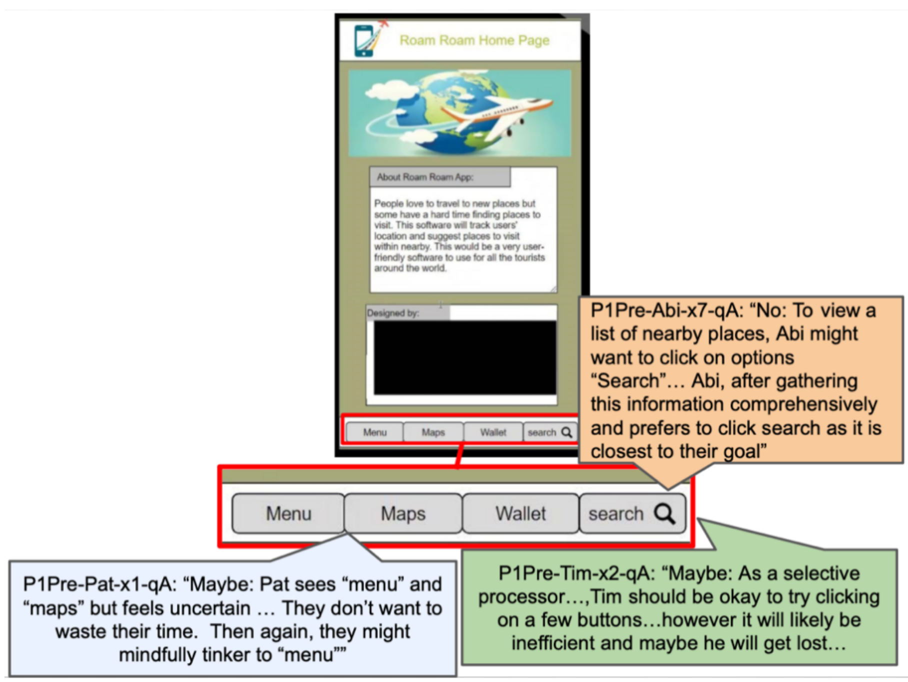

Figure 9: The Question A screenshot of Project 1Pre (HCI Students' travel planning application) with quotes from the Inclusivity Bug Reports[3]. To answer Question A, the scenario was that the persona was trying to find a nearby point of interest and the subgoal was that the persona had already decided to find a list of nearby places. Experts reported several inclusivity bugs relating to the labeling of buttons at the bottom of the screen.

collecting a large batch of the information available *before* taking action to solve the problem.

For example, Project 1Pre (Figure 9) had an inclusivity bug for comprehensive information processors like Abi. In this project, students implemented an app to help people find places to visit while travelling. Students' button labels on this page were problematic for Abi's Information Processing Style because they would read all the options, but the label on the button Abi "should" click on did not appear to be the closest option to their goal.

Further evidence of students struggling to support risk-averse and comprehensive processing users is given by the high representation of the Attitude Toward Risk and Information Processing Style in inclusivity bugs for Pat. Recall that the five GenderMag facets have ranges of values where Abi and Tim capture the end points and Pat provides a third set of values. Pat's Information Processing Style and Attitude Toward Risk are similar to Abi's: they both process comprehensively, they read everything on the screen, and

[3]Quote IDs are as follows: P1Pre-Abi-x7-qA = Project 1 by Pre-GenderMag students, response about the Abi persona, by Expert 7 on question A

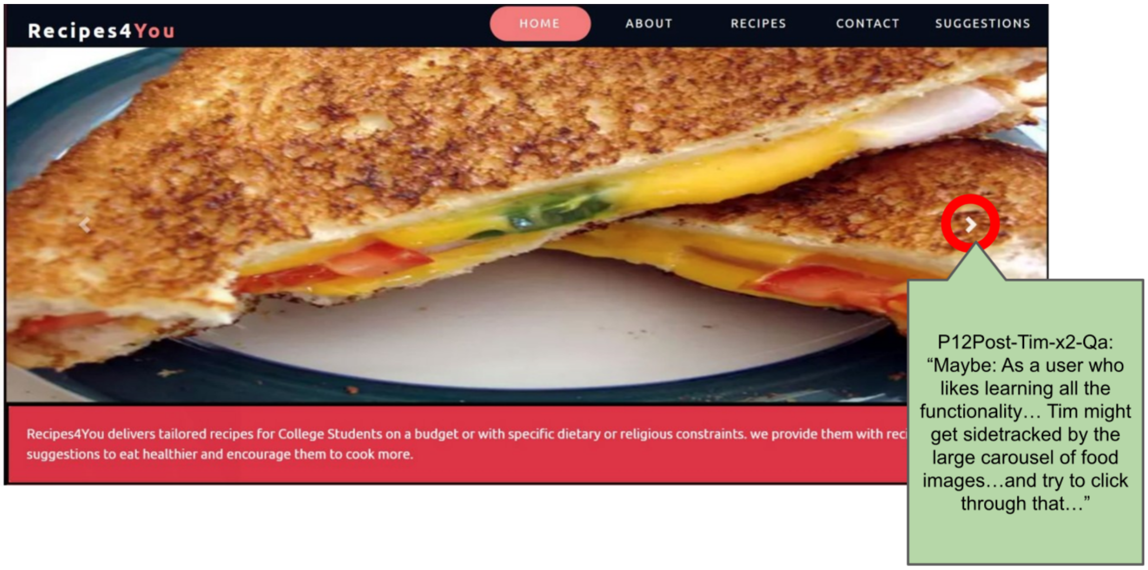

Figure 10: The Question A screen from Project 12Post made by Senior Project students. The scenario was that the persona was trying to find a recipe to make for dinner and the subgoal was that the persona had decided they needed to view a list of dinner recipes. While the desired action was at the top (click the "Recipes" button), Inclusivity Bug Reports showed Tim's attention may be drawn to the large image carousel in the middle. (We added a red circle around the carousel button)

are risk-averse in part due to a lack of spare time. This similarity is shown in the way that students' button labels in Project 1Pre were also a problem for Pat's Attitude Toward Risk and Information Processing Style.

This same inclusivity bug also connected to Pat's Learning Style, which was Pat's other facet value that students especially struggled with in Pre-GenderMag projects. The Learning Style facet captures the ways in which people tend to learn new technology. Pat's Learning Style is mindful tinkering with new technology (which differs from Abi's by-process Learning Style that prefers tutorials and step-by-step directions and also differs from Tim's tinkering a bit excessively).

For Tim, students struggled most in supporting Information Processing Style and Motivations. Compared to Abi and Pat's comprehensive Information Processing Style, Tim is a selective processor. This means they tend to follow a depth-first pattern, pursuing the first promising piece of information then backing out if need be. For example, the same buttons in Project 1Pre that caused bugs for Abi's and Pat's Information Processing were also a source of bugs for Tim's Information Processing Style. This is interesting as it showed students' struggling with this facet across the entire spectrum of the Information Processing Style facet values.

An example of a bug associated with Tim's Motivations arose in Project 12Post (Figure 10; previously shown in Section 3 as Figure 3). Recall that for this project, students built a website for users to search for recipes and submit recipe suggestions. The Motivations facet is about the reasons why people use technology. Tim likes to use technology to learn all available functionality (whereas Abi uses technology to achieve a task and Pat learns new technology when needed).

Finally, Project 2Pre provided an example of an inclusivity bug for Abi's Computer Self-efficacy. The Computer Self-efficacy facet captures people's self-confidence relative to their peers when using technology and where they place blame when problems happen. Abi's Computer Self-efficacy is lower relative to their peers and they tend to blame themselves when something goes wrong or if they do not understand something. As shown in Figure 11, students added options without indicating their importance or when they should be used, which may cause Abi-like users to be unsure.

4.2.2 The student projects' inclusive design successes. In contrast to the "maybe's" and "no's" indicating inclusivity bugs in students' projects, experts' "yes" responses to Questions A and B indicated students' support for the personas. Figure 12 shows the most successful facets per persona for students' Pre-GenderMag projects. However, note that we are not comparing the "yes" responses to the "maybe" and "no" responses as we have determined that it is not a fair comparison. We will discuss this further in Section 5 (Discussion).

As the Pre-GenderMag averages in Figure 12 show, students were able to support Information Processing Style more than any other facet for all three personas. Recall that Abi and Pat's Information Processing Style is comprehensive while Tim's is selective. As the figure shows, despite students creating many bugs associated with Information Processing Style, they also frequently were able to support the whole range of facet values.

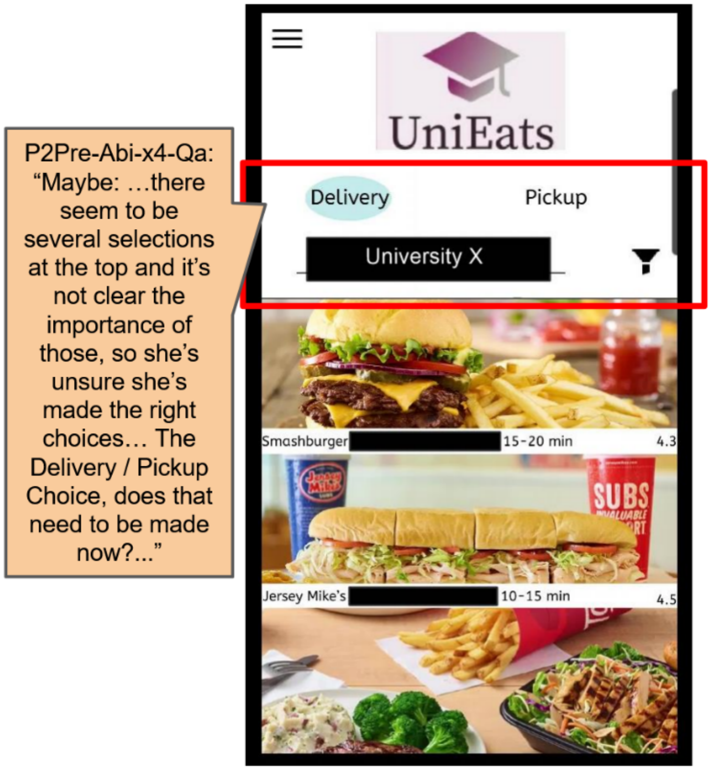

Figure 11: The Question A screenshot of Project 2Pre (HCI students' food ordering application). The scenario was that the persona was trying place a food order and the subgoal was to do so by finding the Smashburger menu. Expert 4 indicated that Abi may be concerned about the options at the top of the screen and if they have made the right choices to find exactly what they are looking for.

Figure 12: Average per project distribution of facets from experts' "yes" responses for each persona for the 11 Pre-GenderMag project evaluations. Each distribution shows the average percentage each facet represents of the total facets used for each project. The Information Processing Style facet was most well supported for all personas (Abi: 41%, Pat 38%, Tim 45%).

For students' Post-GenderMag projects', there was also support of diverse Information Processing Styles in the Post-GenderMag projects as shown by the distributions in Figure 13.

Project 18Post provides an example of students' support of Information Processing Styles for all three personas. Students created this project, shown in Figure 14, as an inventory management website. On the Question B screen, students' clear labeling and large headings supported the Information Processing Style of all three personas. This success provides interesting contrast with the Information Processing inclusivity bugs caused by the unclear labels in Project 1Pre (Figure 16).

Beyond Information Processing, students in both the Pre- and Post-GenderMag courses were able to support a variety of facet values. For example, the students who implemented Project 22Post

Figure 13: Average distribution of facets from experts' "yes" responses for each persona from the 11 Post-GenderMag design evaluations. Information Processing was the most supported facet, making up more than half the successes for all personas (Abi: 62%, Pat 53%, Tim 54%).

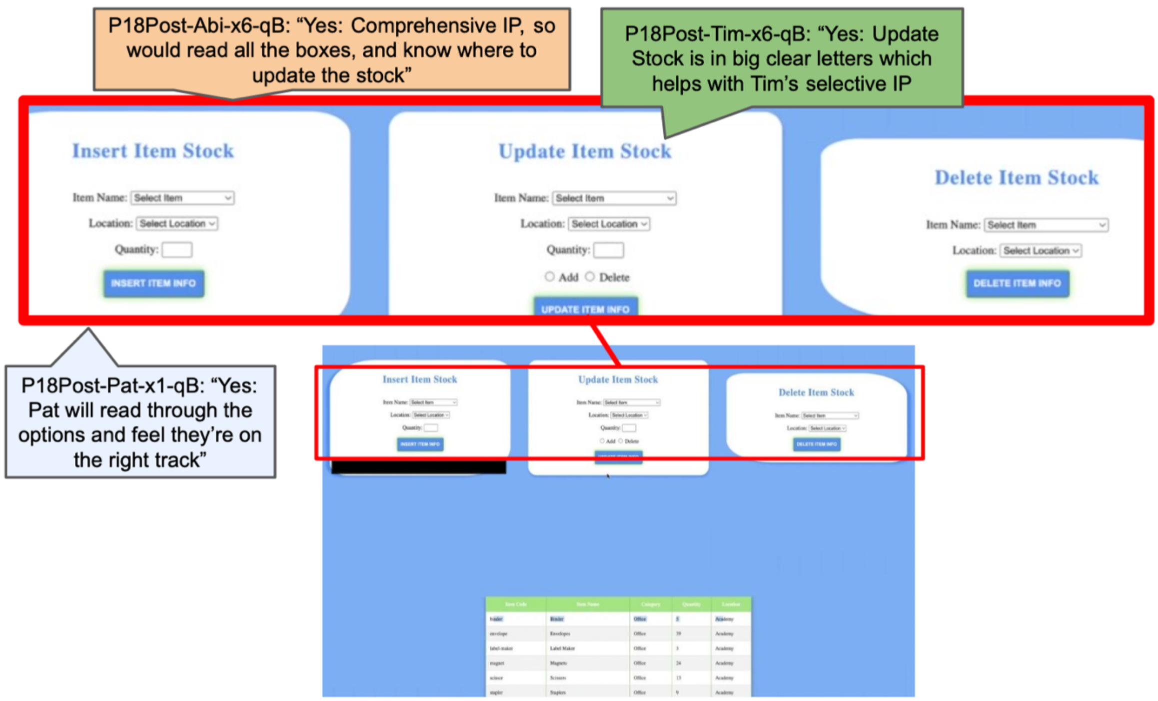

Figure 14: The Question B screenshot of Project 18Post, an inventory management website built by Software Engineering students. For the evaluation, the scenario was that the persona wanted to log that 2 binders in inventory were no longer available and the subgoal was to update the stock of binders. At this point, the personas had taken the action to click the "Manage Stock" which navigated to this page. Inclusivity Bug Reports showed that the labeling supported selective information processing while the options on the page helped comprehensive information processors.

supported Abi's Attitude Toward Risk as shown in Figure 15. Abi is risk-averse about using new features so providing a call to action may help Abi move forward.

Another example of other successes came from the Question B screen in the Project 1Pre, shown in Figure 16. Previously we explained that in the Question A screen, the students' project lacked support for all personas (Figure 9). In contrast, students supported all personas on the Question B screen. Students' clear label and simple interface helped Abi, Pat, and Tim. For Abi, the design supported their Motivations of using technology to complete their tasks. For Pat, it supported their Learning Style of mindful tinkering. Finally for Tim, it supported their higher self-confidence when using technology due to their higher Computer Self-efficacy.

This final example also helps to illustrate how for some students' support for facet values or lack thereof varied even within a single project throughout the project. Though students' Question A

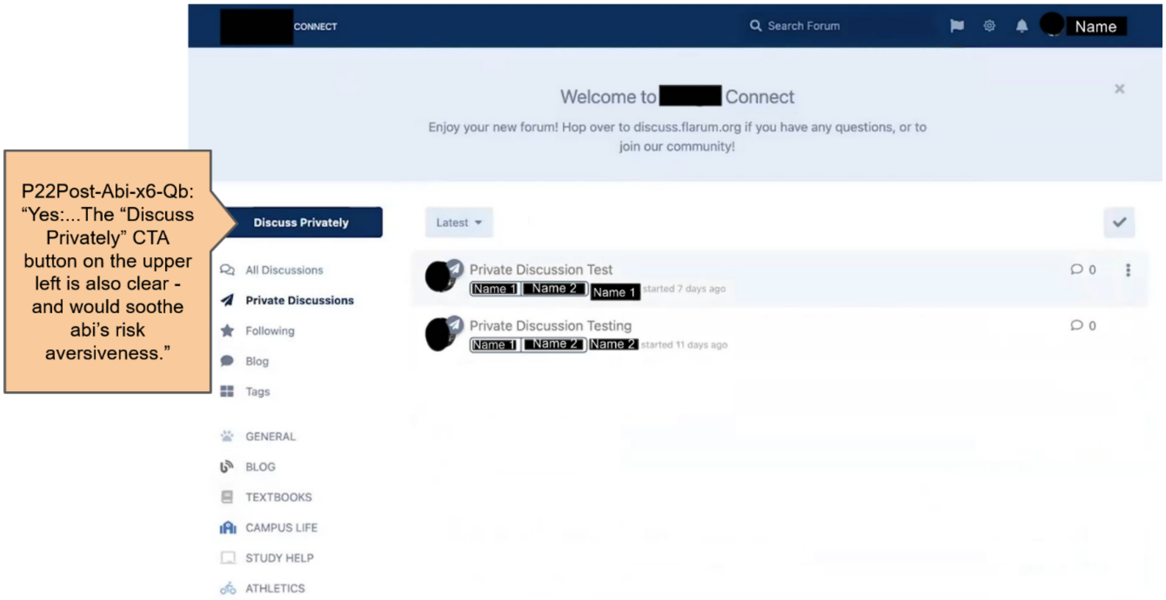

Figure 15: The Question B screenshot of Project 21Post, Software Engineering students' online university discussion forum. For the evaluation, scenario was that the persona wanted to send a private message to a friend and the subgoal was to switch to private discussions. Expert 6 commented that students' use of a call to action (CTA) button helped Abi know they made progress toward their goal.

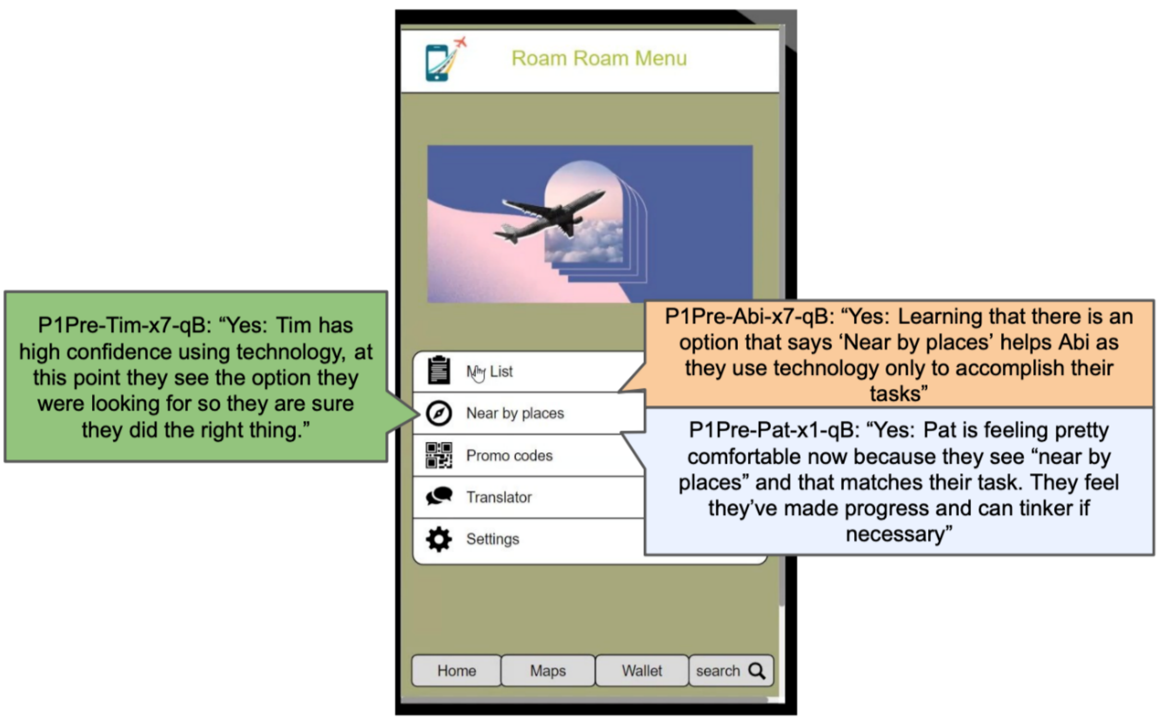

Figure 16: The Question B screenshot of HCI students' Project 1Pre (previously shown in Figure 9) with experts' comments. Recall the subgoal was to find a list of nearby places. In contrast to the Question A screen (Figure 9), Inclusivity Bug Reports showed that the clear labelling of the "near by places" button on the Question B screen worked well for all personas.

screen from this project demonstrated a variety of inclusivity bugs (described in Section 4.2.1), students were able to support all three personas in the Question B screen.

5 DISCUSSION

5.1 Why and why not?

Our results suggest the approach University X followed of teaching elements of GenderMag across-the-curriculum did succeed at moving students to act on their new knowledge. Post-GenderMag students demonstrated improved inclusive design skills with better support for all three personas and support for the full spectrum of facet values. The faculty had hoped that the approach would move students from just "awareness" to action for several reasons, and we believe the following two may have been especially influential: (1) the actionable nature of the GenderMag elements students experienced, and (2) the integration of inclusive design into many of the CS courses, with the goal that inclusive design was "simply" part of creating high-quality software.

Aspect (1) was a result of leveraging a concretely actionable inclusive design method, which was GenderMag in this case study. We believe that teaching an actionable approach helped students learn actionable skills through the elements of GenderMag embedded in their degree program. We also believe that any inclusive design method that includes concretely actionable elements could be substituted for GenderMag, such as concrete inclusive design guidelines/heuristics.

Aspect (2) was to move inclusive design into the courses and projects that students are already completing. By doing this, using inclusive design became part of students' "day jobs" (the CS work that they do for school every day). This contrasts with approaches that add inclusive design to the curriculum through different, isolated courses or modules. When inclusive design is sidelined into specialized courses or modules, we believe the message this sends to students is that inclusive design is extra, not a necessary daily consideration.

That said, there is still room for improvement in Post-GenderMag students' inclusive design skills. As Figures 4 and 6 indicated, the students' inclusion of Abi-like users was still below their inclusion of Pat- and Tim-like users in Post-GenderMag projects.

However, there is still hope for CS students to create even more inclusive project than they did in this case study. Our results captured only the first academic year in which these across-the-curriculum changes were implemented. We hope that as time passes and more students experience the full four-year curriculum with embedded GenderMag elements they will gain the skills necessary to close the gap between Abi and Pat/Tim.

5.2 Threats and Limitations

5.2.1 Data Limitations. The first limitation to our work was that we were not able to use inferential statistics on our data, due to a combination of the lack of controls, relatively small sample sizes, and the dependencies within our data. The environment for our study removed our ability to impose any controls on the projects, so the many varying external factors alone made statistical analysis inappropriate for our case study. Moreover, one statistical assumption behind many statistical tests is that the data points are independent of one another [37], and our data violated this assumption. For example, Project 1Pre might be evaluated by expert A, B, and C, and Project 2Pre might be evaluated by expert B and D. We cannot combine these data because both projects were evaluated by expert B, whose evaluations are not independent from each other. However, the other experts' evaluations of the projects would be independent so we could not use a paired test with an assumption of all points being dependent. Additionally, we do not know if any students were involved in more than one of the project groups. If our project pool had been very large or our expert pool had been very large, we might have been able to divide up the mapping of experts to projects to avoid the dependence between evaluations, but as is, there were no suitable ways to use statistical inference to compare them. Instead, we used extensive triangulation to safeguard our results as described in Section 5.3.

Another limitation relating to our data was in the experts' rankings responses. Recall that the evaluation rubric asked experts to rank the personas in order of how well they were supported from 1 (high) to 3 (low) (Figure 2). However, the experts appeared to interpret the way to report edge cases differently. For example, some reported ties for Abi/Pat/Tim as 3/3/2, as 3/3/1, or as 2/2/1. Since we had was no way to know if these rankings meant the same thing, we dropped the rankings. That said, we do not view losing these data as a serious limitation because all the other measures were so consistent with one another.

A final data limitation was the difference between experts' "yes" responses and their "maybe"/"no" responses. After we compiled the data, we discovered that experts reported fewer facets associated with their "yes" answers compared to their "maybe"/"no" answers. For example, "maybe" and "no" responses for the Abi persona were on average associated with 2.5 facets per response while "yes" answers were only associated with on average 1.5 facets per response. While we do not know why this occurred, we concluded this difference meant that comparing the positive and negative answers in terms of facet usage was not a fair comparison. Thus, further investigation is needed to compare the difference in students' support of the facets Pre- and Post-GenderMag.

5.2.2 Context Threats. In our case study, we investigated the inclusivity of students' work products in a specific context over a specific period of time. Thus, our results are particular to this context of students at University X in these particular courses. Our results are also specific to the particular projects and experts that were included in the study. Had we chosen different projects or different GenderMag experts, we may have had different results. Because of this, it is not possible for us to generalize beyond this context without further investigation.

Because of the timing of our study, it is also likely that we captured only part of the impact on students' use of inclusive design. The University X faculty began teaching with GenderMag in Fall of 2021 and our data was collected from the first three terms that students experienced the updated classes. As a result, the students in our study had not experienced the full scope of the curriculum changes across all four years of their degree program. Thus, we cannot conclude that this is the final measurement of students' inclusivity at University X. Still, our results provide encouraging evidence that courses with small elements of GenderMag can help

	Evidence Sources: Inclusivity Ratings			Evidence Sources: Inclusivity Bug Reports (Q. A & B)		Totals
	Avg. Per Project	Avg Per Eval.	Distribution by Eval.	% Maybe/No	% Yes	
Abi	✓	✓	✓	✓	✓	5
Pat	✓	✓		✓	✓	4
Tim	✓	✓	✓	✓	✓	5
Total	3	3	2	3	3	14

Figure 17: Evidence used to triangulate the results of RQ1. For each source of evidence, 14/15 sources of evidence produced the same conclusions.

students design more inclusively even if they have not been exposed to it for the full four years.

5.3 Triangulation

As with other case studies, our methodology included extensive triangulation to identify results in multiple ways through multiple pieces of evidence [46]. This was done both to safeguard against threats to construct validity and to add confidence to our results.

As Figure 17 shows, RQ1 was cross-confirmed with 2 sources of evidence. Within each source of evidence, we also considered multiple perspectives as applicable. For each perspective we considered, we triangulated further by considering it from the viewpoint of each of the three GenderMag personas to cover the full spectrum of facet values. In total, the result that students created more inclusive projects in Post-GenderMag courses was triangulated by 14 of the 15 total viewpoints.

For RQ2, we did not triangulate a specific outcome or finding but instead considered the details of students' support of each facet through the quantitative perspective of the facets' distribution and the qualitative perspective of experts' responses to Questions A and B. Similar to our triangulation within each perspective for RQ1, we also considered students' support for all three personas through these two lenses.

CONCLUDING REMARKS

In this paper, we have presented a case study to investigate whether an approach University X has been trying would move students to take societal action through the technology they build. Specifically, University X CS educators engaged in *teaching* elements of GenderMag in ways tightly integrated through most courses in the 4-year CS curriculum—and the Post-GenderMag students who experienced this curriculum moved *to take action,* building more inclusive software than their predecessors did.

These results are very encouraging, because the particular student projects in this case study were not specifically inclusive design projects: they were term-long projects in more general courses, namely Software Engineering, HCI, Senior Project, and two Capstone projects. Thus, these projects were not tied to a *"reward"* for designing inclusively. Further, the projects we collected were after only one year, so it was far from certain that students would have learned enough inclusive design skills to succeed at building more inclusive software than their predecessors.

- Specifically (RQ1:) students who learned inclusive design across the curriculum did *act* to create more inclusive software. Post-GenderMag student projects received higher ratings for all personas, and the Inclusivity Bug Reports for their projects had fewer inclusivity bugs that for Pre-GenderMag student projects. These results were very robust, triangulating to the same conclusion in 14/15 measures (Section 5.3).
- RQ2 revealed the students' biggest challenges. Even the Post-GenderMag students struggled particularly often to support all three of Abi's, Pat's, and Tim's Information Processing facet values—which is interesting in that Tim's information processing value is the opposite of Abi's and Pat's. Other particularly challenging facets were Abi's and Pat's Risk values, and Tim's Motivations (Section 4.2.1).
- That said, RQ2 also revealed that the troublesome Information Processing facet above was also a strong source of the students' inclusivity successes. For all three personas, the majority of Post-GenderMag students' successes were tied with the Information Processing facet. The three personas' Learning Styles and Motivations were also reasonably strong sources of inclusivity successes (Section 4.2.2).

These results not only provide encouraging evidence for the across-the-curriculum inclusive design approach, but also show it is possible to lead students beyond "awareness" to action. Perhaps this is the most important result: that it is possible to effectively educate tomorrow's CS professionals to build more inclusive technology than their predecessors have done.

ACKNOWLEDGMENTS

We thank the seven experts for volunteering their time and effort. We also thank the students and faculty at University X for their contributions and participation. This work was supported in part by USDA-NIFA/NSF grant number 2021-67021-35344, and by NSF grant numbers 1901031, 2042324, 2042341, 2345334, and 1928452. Any opinions expressed are those of the authors alone, and do not necessarily reflect the views of the sponsors.

REFERENCES

[1] Bar-El, David, and Marcelo Worsley. "Making the maker movement more inclusive: Lessons learned from a course on accessibility in making." International Journal of Child-Computer Interaction 29 (2021): 100285.
[2] 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.

[3] 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

[4] 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

[5] Amreeta Chatterjee, Mariam Guizani, Catherine Stevens, Jillian Emard, Mary Evelyn May, Margaret Burnett, and Iftekhar Ahmed. 2021. AID: An automated detector for gender-inclusivity bugs in OSS project pages. In 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE), pp. 1423-1435.

[6] Amreeta Chatterjee, Lara Letaw, Rosalinda Garcia, Doshna Umma Reddy, Rudrajit Choudhuri, Sabyatha Sathish 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 - Volume 1 (ICER '22), Vol. 1. Association for Computing Machinery, New York, NY, USA, 356–372. https://doi.org/10.1145/3501385.3543973

[7] 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.

[8] 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

[9] Rosalinda Garcia, Patricia Morreale, Lara Letaw, Amreeta Chatterjee, Pankati Patel, Sarah Yang, Isaac Tijerina Escobar, Geraldine Jimena Noa, and Margaret Burnett. 2023. "Regular" CS × inclusive design = smarter students and greater diversity. ACM Trans. Comput. Educ. 23, 3, Article 34 (July 2023), 34 pages. https://doi.org/10.1145/3603535

[10] Rosalinda Garcia, Patricia Morreale, Gail Verdi, Heather Garcia, Jimena Noa Guevara, Spencer Madsen, Maria Jesus Alzugaray-Orellana, Elizabeth Li, and Margaret Burnett. 2024. The Matchmaker Inclusive Design Curriculum: A Faculty-Enabling Curriculum to Teach Inclusive Design Throughout Undergraduate CS. In Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/3613904.3642475

[11] Rosalinda Garcia, Patricia Morreale, Pankati Patel, Jimena Noa Guevara, Dahana Moz Ruiz, Sabyatha Satish Kumar, Prisha Velhal, Alec Busteed, and Margaret Burnett. 2024. Supplemental Documents. https://figshare.com/articles/online_resource/ICER_2024_Supplemental_Documents/25506220

[12] Jerilee Grandy 1994. Gender and ethnic differences among science and engineering majors: Experiences, achievements, and expectations. ETS Research Report Series 1994, 1 (1994) ,i–63.

[13] 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). Association for Computing Machinery, New York, NY, USA, 225–238. https://doi.org/10.1145/3446871.3469754

[14] B.J. Grosz, D.G. Grant, K. Vredenburgh, J. Behrends, L. Hu, A. Simmons, and J. 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

[15] Mariam Guizani, Igor Steinmacher, Jillian Emard, Abrar Fallatah, Margaret Burnett, and Anita Sarma. 2022. How to debug inclusivity bugs? a debugging process with information architecture. In Proceedings of the 2022 ACM/IEEE International Conference on Software Engineering: Software Engineering in Society, pp. 90-101.

[16] 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.

[17] Diane Horton, Sheila A. McIlraith, Nina Wang, Maryam Majedi, Emma McClure, and Benjamin Wald. 2022. Embedding Ethics in Computer Science Courses: Does it Work? In Proceedings of the 53rd ACM Technical Symposium on Computer Science Education - Volume 1 (SIGCSE 2022), Vol. 1. Association for Computing Machinery, New York, NY, USA, 481–487. https://doi.org/10.1145/3478431.3499407

[18] Diane Horton, David Liu, Sheila A. McIlraith, Steven Coyne, and Nina Wang. 2024. Do Embedded Ethics Modules Have Impact Beyond the Classroom? In Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2024). Association for Computing Machinery, New York, NY, USA, 533–539. https://doi.org/10.1145/3626252.3630834

[19] 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

[20] Kearney-Volpe, Claire, Devorah Kletenik, Kate Sonka, Deborah Sturm, and Amy Hurst. "Evaluating instructor strategy and student learning through digital accessibility course enhancements." In Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility, pp. 377-388. 2019.

[21] Emily Kuang, Selah Bellscheidt, Di Pham, Kristen Shinohara, Catherine M Baker, and Yasmine N. Elglaly. 2024. Mapping Accessibility Assignments into Core Computer Science Topics: An Empirical Study with Interviews and Surveys of Instructors and Students. In Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 656, 1–16. https://doi.org/10.1145/3613904.3642097

[22] Landa-Avila, Irma C., and Carlos Aceves-Gonzalez. Inclusive human-centered design: Experiences and challenges to teaching design engineering students. In Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018) Volume VII: Ergonomics in Design, Design for All, Activity Theories for Work Analysis and Design, Affective Design 20, pp. 1558-1570. Springer International Publishing, 2019.

[23] 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 17th ACM Conference on International Computing Edu- cation Research (ICER 2021), August 16–19, 2021, 18 pages. https://doi.org/10.1145/3446871.3469742

[24] Levy, Laura, and Maribeth Gandy. "Teaching video game design accessibility: toward effective pedagogic interventions in accessible design." Universal Access in Human-Computer Interaction. Theory, Methods and Tools: 13th International Conference, UAHCI 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part I 21. Springer International Publishing, 2019.

[25] 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.

[26] 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.

[27] Mozilla Foundation. Responsible Computer Science Challenge. Retrieved September 5, 2019 from http://responsiblecs.org .

[28] Emerson Murphy-Hill, Alberto Elizondo, Ambar Murillo, Marian Harbach, Bogdan Vasilescu, Delphine Carlson, and Florian Dessloch. 2024. GenderMag improves discoverability in the field, especially for women. In IEEE/ACM International Conference on Software Engineering (ICSE), pp. 973-973.

[29] 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

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

[31] Alannah Oleson, Meron Solomon, Christopher Perdriau, and Amy Ko. 2023. Teaching inclusive design skills with the CIDER assumption elicitation technique. ACM Trans. Comput.-Hum. Interact. 30, 1, Article 6 (March 2023), 49 pages. https://doi.org/10.1145/3549074

[32] Alannah Oleson, Amy J. Ko, Richard Ladner (2024). Teaching Accessible Computing. https://bookish.press/tac, retrieved 3/12/2024.

[33] 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).

[34] Pankati Patel, Dahana Moz-Ruiz, Rosalinda Garcia, Amreeta Chatterjee, Patricia Morreale, and Margaret Burnett. 2024. From workshops to classrooms: Faculty experiences with implementing inclusive design principles. In Proceedings of the 55th ACM Technical Symposium on Computing Science Education (SIGCSE 2024), March 20-23,2024, Portland, OR, USA. 7 pages. https://doi.org/10.1145/3626252.3630861

[35] Pew Research Center. 2020. On the Cusp of Adulthood and Facing an Uncertain Future: What We Know About Gen Z So Far. Retrieved April 12, 2023 from 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/.

[36] Poor, G. Michael, Laura M. Leventhal, Julie Barnes, Duke R. Hutchings, Paul Albee, and Laura Campbell. "No user left behind: Including accessibility in student projects and the impact on CS students' attitudes." ACM Transactions on Computing Education (TOCE) 12, no. 2 (2012): 1-22.

[37] Ramsey, F., Schafer, D. 2002. The Statistical Sleuth: A Course in Methods of Data Analysis. Duxbury Press. ISBN: 0534386709.

[38] Phil Alden Robinson (director),1989. Field of Dreams, Lawrence Gordon and Charles Gordon (producers).

[39] Ben R. Shapiro, Emma Lovegall,Amanda Meng, Jason Borenstein, and 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. 1034–1040.

[40] 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

[41] Shi, Weishi, Heather Moses, Qi Yu, Samuel Malachowsky, and Daniel E. Krutz. "ALL: Supporting Experiential Accessibility Education and Inclusive Software Development." ACM Transactions on Software Engineering and Methodology 33, no. 2 (2023): 1-30.

[42] Rick Spencer. 2000. The streamlined cognitive walkthrough method, working around social constraints encountered in a software development company. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 353-359.

[43] 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.

[44] 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

[45] Watchorn, Valerie, Helen Larkin, Susan Ang, and Danielle Hitch. Strategies and effectiveness of teaching universal design in a cross-faculty setting. Teaching in Higher Education 18, no. 5 (2013): 477-490.

[46] Yin, Robert K. 2009. Case study research: design and methods. India, SAGE Publications.