Motivating Literature and Evaluation of the Teaching Practices Game: Preparing Teaching Assistants to Promote Inclusivity

Audra Lane, Ruth Mekonnen, Catherine Jang, Phoebe Chen
Harvey Mudd College
Claremont, CA, USA
{alane,rmekonnen,cjang,phchen}@hmc.edu

Colleen M. Lewis
University of Illinois at Urbana-Champaign
Urbana, IL, USA
ColleenL@illinois.edu

ABSTRACT

In the US, there are longstanding patterns of underrepresentation in computing (see Table 1). To make CS more inclusive, we can train computer science (CS) teaching assistants (TAs) to create inclusive classrooms. The Teaching Practices Game is a scenario-based card game meant to prepare CS TAs for difficult situations they may encounter and help them promote diversity and inclusion. Game participants (N=86) were surveyed from multiple institutions. The majority of survey respondents (N=86) agreed or strongly agreed that the game taught them new strategies for responding to difficult teaching situations (83%) and for discussing diversity and inclusion (69%). Additionally, respondents reported that the game made them more confident (74%) and more likely (66%) to respond to biased statements. Contrary to our hypotheses, we found that there were small and statistically insignificant differences in enjoyment, learning, or likelihood of response to biased statements between participants who do and do not identify as underrepresented in CS. A primary contribution of the research is a review of professional development practices and diversity training strategies that can inform the training of TAs, and we discuss the extent to which best practices from previous research were incorporated in the game.

CCS CONCEPTS

• **Social and professional topics → Computing education**; **Computer science education**.

KEYWORDS

diversity; stereotypes; microaggressions; professional development

1 INTRODUCTION

As shown in Table 1, numerous demographic groups are underrepresented in CS education and the CS labor force in the US[1]. There are many advantages to Broadening Participation in Computing (BPC), or increasing the number of CS students and CS professionals from groups currently underrepresented in computing. In colleges and universities, efforts to increase diversity benefit students both socially and academically [35]. While not all students have the same experience [24], students who report encountering a diversity of ideas and identities in college report higher levels of satisfaction and sense of community, and they show evidence of increased critical thinking and motivation [35]. In the CS industry. Diversity offers a variety of perspectives, so it increases creativity and innovation [46]. A more diverse set of employees can also help a company reach a more diverse customer base [3]. In short, BPC can benefit both individuals from groups underrepresented in computing and the companies that hire them.

Implicit bias, structural inequality and microaggressions are among the many causes of longstanding patterns of underrepresentation [5, 24, 36, 44, 45, 56]. If we can reduce the bias that CS students have to face in undergraduate CS classrooms [15, 50, 56, 58, 62], we can likely better encourage retention and recruitment.

The Teaching Practices Game [28, 29] is designed for CS teaching assistants (TAs). TAs likely play an important role in creating an inclusive environment that reduces the impact of racism, sexism, and other discrimination. The game constitutes working through a deck of cards in two rounds. Each card contains a potential situation and a sample response for the situation. In Round 1, the cards address teaching situations. In Round 2, the situations focus on diversity-related discussions.

The Teaching Practices Game is designed to be played over the course of 90 minutes; the game is explained (10 minutes), small groups of three to six play Round 1 (20 minutes), the whole group discusses Round 1 (20 minutes), small groups play Round 2 (20 minutes), and the whole group discusses Round 2 (20 minutes). Round 1 and Round 2 consist of 29 and 23 cards, respectively. It is expected that groups will only discuss 5-10 cards in each round.

For each card, one player in each small group starts as the judge. The judge reads the situation described on the card and listens to each player's description of how they would handle the situation. Then the judge selects the "winner" of the round, which is the player that the judge believes has best responded to the situation on the card. Finally the judge reads the card's sample response to the other players and the role of the judge rotates.

Faculty and staff at academic institutions in the US received copies of the game at an in-person workshop or by mail. The intention was that these faculty and staff would then lead TAs in playing the game at a TA training. At the research team's request, after the game, players were invited to complete a survey about the game.

[1] While Asian men are not classified as underpresented in computing by the National Science Foundation [39], Asian people in the US face discrimination [75]. Additionally, the racial category of Asian is not homogeneous. Subgroups of the Asian-American population (Chinese-American, Korean-American, etc.) differ in terms of historical ease of immigration and access to resources, wealth, and education [75].

SIGCSE '21, March 17-20, 2021, Toronto, Canada

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

Table 1: Representation in CS by Race and Gender in the US

Demographic Group	US Pop.[1]	All BA/BS[2]	CS BA/BS[3]	All Labor Force[4]	CS Labor Force[5]
White Male	29.6%	24.7%	45.0%	42.0%	47.0%
White Female	30.5%	32.3%	8.3%	35.6%	12.6%
Black Male	6.2%	3.4%	5.4%	6.0%	3.8%
Black Female	7.0%	6.0%	1.9%	6.6%	3.4%
Hispanic Male	9.3%	5.3%	8.1%	9.9%	4.5%
Hispanic Female	9.2%	8.3%	1.8%	7.5%	1.4%
Asian Male	2.9%	3.4%	10.7%	3.3%	17.7%
Asian Female	3.2%	4.0%	3.5%	3.0%	7.5%
Pacific Islander or Native Hawaiian Male	0.09%	0.09%	0.2%	0.2%	0.3%
Pacific Islander or Native Hawaiian Female	0.09%	0.1%	0.04%	0.2%	0.07%[6]
American Indian or Alaska Native Male	0.4%	0.2%	0.2%	0.6%	0.1%[6]
American Indian or Alaska Native Female	0.4%	0.3%	0.1%	0.5%	0.07%[6]
Two or More Races Male	1.7%	1.4%	2.5%	1.1%	1.2%
Two or More Races Female	1.7%	2.1%	0.7%	1.0%	0.4%

[1] Population statistics are from the U.S. Census Bureau [65, 68–70]. Sources vary by demographic group: White [69], Black [69], Hispanic [68], Asian [70], Pacific Islander or Native Hawaiian [65], American Indian or Alaska Native [66], Two or More Races [67]. In population and labor force statistics, people who identify their ethnicity as Hispanic are also counted within their selected race.

[2] All BA/BS statistics are from the 2018-2019 academic year and were calculated from data gathered by The Integrated Postsecondary Education Data System (IPEDS) at the National Center for Education Statistics (NCES) [1].

[3] CS BA/BS statistics are from the 2015-2016 school year; they were gathered by IPEDS and downloaded from the BPCnet Resource Portal [72].

[4] All Labor Force statistics are from 2018 and were calculated from data gathered by the U.S. Bureau of Labor Statistics [71].

[5] The CS labor force statistics are from 2017 and were calculated from data gathered by the National Center for Science and Engineering Statistics (NCES) [40].

[6] For Pacific Islander or Native Hawaiian Female, and American Indian or Alaska Native Male and Female, the percentages given are the percent of the total science and engineering labor force because CS-specific data was not available.

Our analysis of the survey data was guided by four research questions: **RQ1:** Overall, how do participants rate the game with respect to its goals? The goals of the game are that players will learn, have fun while discussing difficult topics, and become more likely to intervene to create an inclusive climate. Our remaining research questions investigated differences in students' experiences based upon their identities: Compared to players who identify as underrepresented in computing, do players of the game who do not identify as underrepresented in computing rate the game: **RQ2:** lower in terms of their enjoyment? **RQ3:** higher in terms of their learning? **RQ4:** higher in terms of their likelihood of response to biased statements?

We preregistered our research questions, hypotheses, and data analysis methods on the Open Science Framework (osf.io). Contrary to our hypotheses, there were small and statistically insignificant differences between the survey responses from students who do and do not identify as underrepresented in computing.

2 PREVIOUS RESEARCH AND APPLICATION OF BEST PRACTICES IN THE GAME

Here we review two bodies of literature that motivate our design and analyses. First, professional development practices that support teachers in improving their practice. Second, diversity training strategies. For each, we discuss the extent to which best practices from previous research were incorporated into the game.

2.1 Professional Development Practices

In a meta-analysis of 35 studies of professional development (PD) programs for teachers, Darling-Hammond et al. identified seven "design elements of effective professional development." The Teaching Practices Game attempts to incorporate five out of the seven.

"Models and modeling" (Included) refers to providing participants with examples of best practices [12]. The sample answers provide potential best practices for the situation.

"Peer-to-peer collaboration" (Included) refers to group work and participants relying on each other for help [12]. The game is played among peers, and each player shares what they would do in the situation. Players can learn from their peers' responses.

"Active learning" (Included) refers to teaching through hands-on activities where teachers can "analyze, try out, and reflect on new strategies" [12, p. 7]. With each new card, players have the opportunity to practice the strategies they have learned from the sample answers and their peers' past answers.

"Feedback and reflection" (Included) refers to participants receiving feedback and reflecting on the PD program or their performance during said PD program [12]. For each card, one player serves as the judge, who, when deciding the winner of the round, provides feedback to said winner. As for reflection, it is built into the whole-group discussions. Also, when a player's response differs from the sample answer or the winner's response, that player can reflect on how their response could be improved.

"Content-focused" (Included) refers to when PD programs address the participants' discipline [12]. A content-focused PD program for CS would teach CS content and CS pedagogical strategies [12]. Some of the cards present pedagogical and diversity-related issues as they would arise in the CS classroom specifically.

"Coaching and expert support" (Not included) refers to participants receiving individual instruction or feedback from the PD instructor or some other expert [12]. While the workshop host could hypothetically provide coaching and expert support, the game typically only involves peer-to-peer collaboration.

"Long in duration" (Not included) refers to when programs occur in multiple sessions over a long period of time (e.g., a year or a semester) [12]. Some CS faculty report using a few cards at each meeting with their TAs. In the future, we will encourage this because when PD participants repeatedly practice new strategies, they are more likely to apply these strategies in the real-life classroom [12]. However, in most cases, the game is not long in duration.

2.2 Diversity Training Practices

2.2.1 Background. Here we introduce previous research about stereotypes, implicit bias, structural inequality, and microaggressions and the harm caused by each within an educational context. Section 2.2.2 addresses methods of reducing or mitigating this harm.

Stereotypes - First, let us consider how stereotypes inhibit inclusivity in computing. Stereotypes are "a socially shared set of beliefs about traits that are characteristic of members of a social category" [23, p. 14]. Stereotypes can be negative (e.g., that women are bad at math) or positive (e.g., that Asian people are good at math). More specifically, negative stereotypes can have harmful effects due to stereotype threat, which is defined as "being at risk of confirming, as self-characteristic, a negative stereotype about one's group" [57]. When people are reminded of a potentially-relevant, negative stereotype about a group to which they belong, their performance tends to conform to that stereotype [57, 63].

Positive stereotypes might seem benign, but have harmful effects as well [26, 53, 55]. Positive stereotypes can invalidate the agency and individuality of group members [55] and create a hierarchy by implying negative characteristics about other groups [53]. Kay et. al. [26] argue that positive stereotypes are more likely to go unnoticed and unquestioned and that positive stereotypes can indirectly serve as evidence of a biological or innate difference between racial groups, which indirectly serves to support negative stereotypes.

Implicit Bias - Discrimination may contribute to patterns of underrepresentation in the US within the labor force [5, 38, 44] and colleges [24, 56]. As such, discrimination may be one explanation for underrepresentation in computing [1, 40, 65, 68–72]. Discrimination can be caused by explicit prejudice or implicit bias.

Payne and Vuletich define explicit prejudice as "*consciously intended* [emphasis added] and overtly expressed beliefs that favor some groups over others" [44, p. 50][2]. Whereas, implicit bias, a subtler form of bias, is a "mental association based on race, gender, and other social categories that may lead to discrimination *without conscious intent* [emphasis added]" [44, p. 49]. A study of 34 countries demonstrated that when a country has a greater gender gap in science and math scores that favors males, citizens of that country are more likely to implicitly associate males and science [42]. Payne, Vuletich and Lundberg use this correlation to argue that implicit biases are more directly correlated to the cultural trends or "social context" [44] of an environment[3] rather than the belief or values of an individual [44]. Although implicit bias is unintentional, it can still lead to harmful, discriminatory results [5, 36, 38, 44, 45]. For instance, studies have shown that implicit bias can negatively impact the fairness of research opportunities [36], hiring decisions [5, 38], and evaluation of writing skills [45].

Structural Inequality - Discrimination can also be embedded in policies and practices that perpetuate structural inequality such as structural racism. "Structural racism refers to the totality of ways in which societies foster racial discrimination through mutually reinforcing systems of housing, education, employment, earnings, benefits, credit, media, health care, and criminal justice" [4, p. 1453].

If colleges and universities have policies (e.g., competitive enrollment policies [41]) that make it harder or impossible to become a CS major without high school (HS) CS experience, those policies can be seen as reinforcing structural inequality. Low-income and Black students have the least access to CS classes in HS [20]. Hispanic HS students are less likely to use computers at home and at school [22]. Girls are "less likely to be encouraged by teachers and parents that they would be good at CS" [21] and fewer girls than boys say they have learnt CS in HS [21]. Competitive enrollment policies can enforce structural racism and sexism because of their disproportionate negative impact on the admission to CS degree programs of students who are from groups underrepresented in CS.

Microaggressions - Implicit bias affects the classroom through microaggressions, which are intentional or unintentional [61] "brief, everyday exchanges that send denigrating messages to certain individuals because of their group membership" [59, p. xvi]. One microaggresion especially relevant to the classroom is "Ascription of Intelligence," [61, p. 276] based upon a person's race. Examples include "asking an Asian person to help with a math or science problem" [61, p. 276] or saying "you are a credit to your race" [61, p. 276] to a Black or Latinx student. Both deploy the stereotype that Asian people are more intelligent than other people of color [61].

From our perspective, despite being "brief, everyday exchanges," [59, p. xvi] the denigrating effects of microaggressions seem inconsistent with the prefix "micro." Microaggressions have notable negative outcomes in educational contexts [50, 56, 58, 62]. Racial microaggressions can negatively impact the mental health of students who experience them [60], activate stereotype threat [58], lead to less productivity and problem-solving efficiency [15, 50], and create unwelcoming and less inclusive campus environments [24, 56].

2.2.2 Strategies to Mitigate the Impact of Implicit Bias. Previous research has demonstrated that diversity training can only temporarily reduce implicit biases [6, 17], and diversity training should aim to reduce the *impact* of implicit bias on decisions rather than reduce implicit bias *itself* [44]. We identified three strategies for effective diversity and inclusion workshops, namely challenging stereotypes, encouraging cognitive dissonance, and expecting resistance. Unlike the recommendations for effective professional development for teachers, which came from a meta-analysis [12], recommendations for effective diversity and inclusion workshops came from a range of resources [9, 11, 16, 19, 25, 33, 34, 37, 64, 73].

Challenge stereotypes - In order to decrease the harmful effects of stereotypes, we need to acknowledge and challenge them. One useful way, according to Cohoon and Tychonievich, is to react to stereotypes with surprise [64], which "can be an honest response" that "help[s] combat both the students' belief in the stereotype and their belief that the stereotype is widely believed" [64].

This is exemplified by the Teaching Practices Game, which deals with situations in which students bring up stereotypes in the classroom[4]. For instance, card 37 has the prompt: "A TA says 'It is

[2]Some scholars argue that explicit prejudice has decreased in the US [7, 51], but the recent increase of hate crimes in the US [43] suggests otherwise.

[3]including geographical regions, institutions and organizations

[4]Cards involving stereotypes include 30, 31, 33, 37, 46, 47, 48, 50, 51, and 53 [28].

tough TA-ing for a woman professor because women are so emotional'" [28]. The underlying stereotype here is that "women are emotional, whereas men are rational" [16, p. 303]. The sample answer suggests responding by asking what they meant and noting that all genders may have or express certain emotions, and a leaders' humanity may help them connect with and motivate employees.

Encourage cognitive dissonance - Training can help participants respond to biased statements. However, resulting confrontations may elicit "guilt, self-criticism and dissatisfaction" [9, p. 178] known as "negative self-directed affect" [9, p. 178] (See also [10, 48, 49]). Although these reactions are negative, they can motivate a change in behavior "if the confrontation points out a discrepancy between one's behavior and one's egalitarian self concept" [9, p. 175] (See also [11, 14, 49]). Discrepancy strategies "make students aware of how their potential biases are divergent from important values and standards they hold" [9, p. 175]. Hence, "discrepancy strategies" are crucial for effective confrontations [9, 37].

The sample answers in the Teaching Practices Game promote confrontation of implicit bias using discrepancy strategies. Some sample answers[5] point out how a student's egalitarian intentions or self-concept diverges from the negative impact of their biased statements. More specifically, card 42 is a clear example of a discrepancy strategy. The prompt on card 42 is "Someone says 'I can't be racist! I'm a good person'" [28]. The sample answer suggests explaining that despite good intentions someone's behavior can still have a racist impact and implicit bias manifests in ways that are difficult to detect [28]. This is a discrepancy strategy because it points towards the discrepancy between one's "egalitarian self-concept" [9, p. 175] of being a "good person" and one's biased actions. However, reassurance of the goodness of the student contradicts the need for "negative self-directed affect" to bring about a change in behavior. But we must also consider that an excessive threat to one's self-image could lead to backlash [9, 11].

Expect resistance - Diversity-related conversations can lead to backlash [73]. Backlash could arise when students find it difficult to recognize the negative impact of their biased statement or acknowledge their privilege [19, 73]. Peggy McIntosh first introduced white privilege as "a corollary aspect" of racism that grants White people an unearned advantage [34, p. 10]. But privilege can also be extended to anyone whose identities are "historically linked to social or political advantages" [73, p. 118] including, but not limited to, men [33, 73], cisgender, able-bodied, and/or heterosexual people [73]. When trying to make classrooms more inclusive, dialogue that addresses privilege is critical, but it can lead to defensiveness [73]. For instance, when a student is told that their privilege played a role in their achievements, they could feel that their efforts are being discounted and get defensive [64]. Defensiveness hinders the engagement of students because it shifts the conversation away from deeper explorations of injustice and of how social justice relates to oneself [19, 73]. So diversity training should prepare educators for future encounters with defensiveness.

With this in mind, Sherry K. Watt developed the Privileged Identity Exploration (PIE) Model. Using the PIE model, Watt clarifies that defensiveness is primal and normal and exists in eight forms called defense modes. Watt suggests that responses to defense modes must exhibit "unconditional positive regard and non-judgmental understanding for students" [73]. Furthermore, Johnson and Longerbeam recommend "empathic listening skills to better manage defensive reactions" [25] and keep students engaged in diversity-dialogues.

Some cards address defense modes[6] and their sample answers model strategies to minimize backlash. Here we focus on one defense mode, False Envy. Watt illustrates False Envy as follows: a White student, in a conversion about racial injustice states, "people from other races are cool. They have an identity to claim as unique, that bends the social rules of normal, yet they are still normal and very strong. And what I wouldn't give to have a tan all of the time" [73, p. 122]. Here, the student's "surface-level admirations" shift the focus away from "deeper explorations of the complexities of race in society" [73, p. 122], so it hinders diversity-dialogues.

Out of the cards that address False Envy[7]; card 36's sample answer applies the best practices of positive, empathic and non-judgemental feedback [25, 73]. The prompt on card 36 is "A student says that it's unfair that some students get extra time on exams just because they have a 'so-called disability'" [28]. Here, the student exhibits False Envy because they do not recognize the privilege of the able-bodied and/or neurotypical student. Moreover, the student focuses on the surface-level admiration of receiving accommodations without explorations of *why* they are provided these. The sample answer acknowledges the benefit of extra time and then attempts to shift the conversation back to a deeper exploration of why people with disabilities are provided with accommodations.

3 METHODS

3.1 Data Collection

People received copies of the game to lead TAs in playing the game at a TA training. At the research team's request, after the game, players were invited to complete a survey about the game without compensation. The survey data was collected between October 7, 2019 and February 3, 2020 from a total of 86 survey respondents[8]. After data collection, but before data observation and analysis, we preregistered our study (osf.io).

3.2 Survey Contents and Indices

Questions 1 through 7 (provided in Table 2) were five-point Likert items. We decided[9] to convert the Likert responses (see Table 2) to numbers 1 through 5. Question 8 collected minimal demographic data (Section 3.2.1). Question 9 asked, "How many minutes did you spend playing the Teaching Practices Game?" (min=10, max=60, median=25, mean=26.59, std=11.82, N=81). Question 10 asked, "About how many cards did your small group discuss?"

[5]Cards encouraging cognitive dissonance include 31, 34, 42, 47, 50 [28].

[6]The game addresses defense modes of Denial (Cards 34, 38, and 52), Rationalization (Card 44), Principium (Cards 44), and False Envy (Cards 36, 43, 48, and 50) [28].

[7]Cards involving False Envy include 36, 43, 48, and 50 [28]

[8]There appear to be 5 clusters of survey responses (5 or more responses submitted within two hours of each other). We interpret these clusters as 5 distinct workshops. 13 responses do not seem to belong to any of the clusters, which may be late submissions.

[9]The options are not linearly spaced in the same way that their numerical counterparts are. For example, an "agree" is not necessarily halfway between "neither agree nor disagree" and "strongly agree" [30]. However, it is a common practice to convert Likert responses to numbers [13], and doing so may make our analysis and results more accessible to readers. As such, we decided to make this conversion.

Table 2: Summary Statistics for Survey Questions using a Likert Scale

Survey Question	Mean	Std Dev	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	N
1. I would recommend the Teaching Practices Game to other teaching assistants.	4.21	0.81	2% (N=2)	1% (N=1)	7% (N=6)	52% (N=45)	37% (N=32)	86
2. I learned new strategies for responding to difficult teaching situations.	4.08	0.83	1% (N=1)	3% (N=3)	13% (N=11)	51% (N=44)	31% (N=27)	86
3. I learned new strategies for discussing diversity and inclusion.	3.83	0.91	1% (N=1)	7% (N=6)	23% (N=20)	45% (N=39)	23% (N=20)	86
4. The [Game] helped me feel more confident responding to biased statements.	3.83	0.75	1% (N=1)	3% (N=3)	21% (N=18)	60% (N=52)	14% (N=12)	86
5. The Teaching Practices Game made me more likely to respond to biased statements.	3.81	0.79	1% (N=1)	1% (N=1)	31% (N=27)	48% (N=41)	19% (N=16)	86
6. Playing the Teaching Practices Game was fun.	3.98	0.84	1% (N=1)	3% (N=3)	19% (N=16)	50% (N=43)	27% (N=23)	86
7. I would like the opportunity to read through the remaining cards from the [Game].	3.92	0.92	1% (N=1)	6% (N=5)	21% (N=18)	44% (N=37)	28% (N=24)	85

(min=3, max=53, median=10, mean=12.94, std=8.64, N=82). Question 11 asked, "How many people (including yourself) were in your small group?" (min=2, max=25, median=5, mean=5, std=2.57, N=82).

Table 3: Summary of Indices and Research Results

	Group	N	Mean	SD
RQ2 - Enjoyment (Survey question 6)	Underrep	39	4.05	0.89
	Well Rep	42	3.90 (p = 0.5)	0.79
	All	86	3.98	0.84
RQ3 - Learning (Survey questions 2, 3)	Underrep	39	3.92	0.88
	Well Rep	42	3.93 (p = 1)	0.67
	All	86	3.93	0.77
RQ4 - Likelihood of response to bias... (Survey questions 4, 5)	Underrep	39	3.69	0.82
	Well Rep	42	3.89 (p = 0.2)	0.51
	All	86	3.82	0.69

3.2.1 Underrepresented in Computing - RQ2-4. We separated the survey respondents into two groups: survey respondents who *do* and *do not* identify as underrepresented in computing. Survey respondents were assigned to one of the groups based on their response to question 8: "I identify as underrepresented in computing because of my gender, race, ethnicity, and/or a disability"[10] with options "yes" (N=39), "no" (N=42), or "decline to state" (N=5)[11].

3.2.2 Enjoyment - RQ2. To answer RQ2, we assessed a survey respondent's enjoyment of the game by their response to survey question 6: "Playing the Teaching Practices Game was fun."

3.2.3 Learning - RQ3. To answer RQ3, we assessed a survey respondent's learning from the game by their responses to question 2 and 3 shown in Table 2. Responses to these questions were averaged to create an index (Cronbach's alpha = 0.74; see Table 3)[12].

3.2.4 Likelihood of Response to Biased Statements - RQ4. To answer RQ4, we assessed a respondent's likelihood of response to biased statements with questions 4 and 5 (see Table 2). These were combined into an index as was done for learning (Cronbach's alpha = 0.74).

3.3 Hypotheses

We hypothesized that survey respondents who *do not* identify as underrepresented in computing would report less enjoyment if they were uncomfortable with diversity-related topics in the game (RQ2), report more learning if they were previously unfamiliar with diversity-related topics (RQ3), and report being more likely to respond to biased statements if the game helped them build this confidence (RQ4). In contrast, we hypothesized that survey respondents who *do* identify as underrepresented in computing would report more enjoyment if they were comfortable with diversity-related topics in the game or found discussing their experiences cathartic (RQ2), report less learning if they were previously familiar with diversity-related topics (RQ3), and report being no more likely to respond to biased statements if they already felt comfortable or face more social repercussions for responding [2] (RQ4). In our study registration (osf.io), we selected a significance threshold of 0.05.

3.4 Limitations

Completing the survey was not required, and our methods do not allow us to calculate response rate. Our research only measures how a participant *reports feeling* about their learning and their likelihood of response to biased statements. Participants may have

[10]It is best practice to disaggregate demographic data by race, gender, and disability status because these multiple dimensions of identity can shape a person's experiences [8, 52]. However, given our expectation of small sample sizes, a quantitative analysis of disaggregated survey responses would be underpowered. With small sample sizes, interviews would be more appropriate to understand how participants' identities shaped their experiences. Unfortunately, this was not feasible. We acknowledge that this creates two extremely heterogeneous groups.

[11]The data of all respondents who selected "decline to state" was not included in any analyses, but are included in descriptive statistics shown in Table 3.

[12]If a survey respondent only answered one of the survey questions in an index, that response is used as their index.

positive biases in their self-evaluation. We did not account for group composition, which can impact participants' experience [47].

3.5 Data Analysis

For each hypothesis (RQ2-4), a permutation test[13] was performed on the two groups' scores or indices for that hypothesis. A permutation test is similar to a two-tailed t-test; both tests compare the means and distributions of two groups, and both produce a p-value. However, unlike t-tests, permutation tests do not assume a normal distribution. Our permutation tests compared the scores or indices of survey respondents who *do not* and *do* identify as underrepresented.

4 RESULTS

The majority of survey respondents (N=86) agreed or strongly agreed that the game taught them new strategies for responding to difficult teaching situations (83%) and for discussing diversity and inclusion (69%). Additionally, participants responded that the game made them more confident (74%) and more likely (66%) to respond to biased statements. Contrary to our hypotheses, there are small and statistically insignificant differences in enjoyment, learning, or likelihood of response to biased statements between participants who *do* identify and *do not* identify as underrepresented in CS.

5 CONCLUSION

There continue to be patterns of underrepresentation in computing. Table 1 shows CS Bachelor's degrees awarded and CS labor-force participation when disaggregated by gender and race with the US [1, 40, 65, 68–72]. It is beneficial to address these patterns of underpresresentation [3, 35, 46]. Even more, they are *necessary* to address because they are created and reinforced by stereotypes, implicit biases, structural inequity, and microaggressions. From our perspective, CS departments have a responsibility to address patterns of unwelcoming campus environments [18, 31] even if they are not entirely unique to CS [24, 56]. The development of the Teaching Practices Game was motivated by the belief that TAs play an important role in making CS courses inclusive and/or toxic and may be best positioned to effect that culture change. Survey responses to the game were generally positive (RQ1). We researched whether those who identify as underrepresented and those who do not perceive the game differently, in terms of enjoyment (RQ2), learning (RQ3), and likelihood of response to biased statements in the future (RQ4). We found that there was no statistically significant difference between the two populations. A primary contribution of the research is a review of professional development practices and diversity training strategies that can inform the training of TAs.

[13] In general, a permutation test compares two groups of values: typically the treatment group and the control group. A test statistic is generated by taking the absolute value of the difference between the means of the two groups. We then perform the permutations that give the test its name; the two groups are randomly shuffled, meaning all the numbers from the original treatment and control groups are combined into one group then redistributed into two new "control" and "treatment" groups. This shuffling is repeated many times, and a test statistic is calculated for each new set of "control" and "treatment" groups. Eventually we have enough test statistics to generate an approximate test statistic distribution; we find our p-value based on where the original test statistic falls on that distribution. The original test statistic is the one calculated from the real, not yet permuted groups. All other test statistics were generated from the thousands of permuted groups. To calculate a p-value, we find the number of test statistics greater than or equal to the original test statistic, then we divide that number by the number of permutations we performed. Leeper provides a more in-depth explanation of the permutation test [27], and we also highly recommend Wilber's visual explanation [74]. To perform our permutation tests, we used an R script provided by Marin Stats Lectures [32]. We also used an excel spreadsheet from the University of Connecticut to calculate Cronbach's alpha for our two indices [54].

ACKNOWLEDGMENTS

The Teaching Practices Game was developed in collaboration with Phillip T. Conrad. The Teaching Practices Game expanded upon Microaggressions: The Game! that was developed in collaboration with Kyla McMullen, Catherine Ashcraft, Wendy DuBow, and Helen Hu. The production of the cards was funded by a SIGCSE Special Project grant. This material is based upon work supported by the National Science Foundation (IUSE 1821136).

REFERENCES

[1] U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System. 2019. Number of awards conferred by Title IV institutions, by race/ethnicity, level of award, and gender: United States, 2018–19. https://nces.ed.gov/ipeds/search?query= Accessed: 2020-07-22.

[2] Sara Ahmed. 2009. Embodying diversity: Problems and paradoxes for Black feminists. *Race Ethnicity and Education* 12, 1 (2009), 41–52.

[3] Jalal Armache. 2012. Diversity in the Workplace: Benefits and Challenges. *Journal of International Diversity* 2012, 1 (2012), 59–75.

[4] Zinzi D Bailey, Nancy Krieger, Madina Agénor, Jasmine Graves, Natalia Linos, and Mary T Bassett. 2017. Structural racism and health inequities in the USA: evidence and interventions. *The Lancet* 389, 10077 (2017), 1453–1463.

[5] Marianne Bertrand and Sendhil Mullainathan. 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review* 94, 4 (2004), 991–1013.

[6] Katerina Bezrukova, Chester S Spell, Jamie L Perry, and Karen A Jehn. 2016. A meta-analytical integration of over 40 years of research on diversity training evaluation. *Psychological Bulletin* 142, 11 (2016), 1227.

[7] Tessa E.S. Charlesworth and Mahzarin R. Banaji. 2019. Research: How Americans' Biases Are Changing (or Not) Over Time. https://hbr.org/2019/08/research-on-many-issues-americans-biases-are-decreasing

[8] Kimberle Crenshaw. 1990. Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stan. L. Rev.* 43 (1990), 1241.

[9] Chelsea Crittle and Keith B Maddox. 2017. Confronting bias through teaching: Insights from social psychology. *Teaching of Psychology* 44, 2 (2017), 174–180.

[10] Alexander M Czopp and Margo J Monteith. 2003. Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality and Social Psychology Bulletin* 29, 4 (2003), 532–544.

[11] Alexander M Czopp, Margo J Monteith, and Aimee Y Mark. 2006. Standing up for a change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology* 90, 5 (2006), 784.

[12] Linda Darling-Hammond, Maria E Hyler, and Madelyn Gardner. 2017. Effective teacher professional development. , 4–16 pages.

[13] JFC De Winter and Dimitra Dodou. 2010. Five-Point Likert Items: t test versus Mann-Whitney-Wilcoxon (Addendum added October 2012). *Practical Assessment, Research, and Evaluation* 15, 1 (2010), 11.

[14] Patricia G Devine, Margo J Monteith, Julia R Zuwerink, and Andrew J Elliot. 1991. Prejudice with and without compunction. *Journal of Personality and Social psychology* 60, 6 (1991), 817.

[15] John F Dovidio. 2001. On the nature of contemporary prejudice: The third wave. *Journal of Social Issues* 57, 4 (2001), 829–849.

[16] Agneta H Fischer. 1993. Sex differences in emotionality: Fact or stereotype? *Feminism & Psychology* 3, 3 (1993), 303–318.

[17] Chloë FitzGerald, Angela Martin, Delphine Berner, and Samia Hurst. 2019. Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: a systematic review. *BMC psychology* 7, 1 (2019), 29.

[18] Kathy Garvin-Doxas and Lecia J Barker. 2004. Communication in computer science classrooms: understanding defensive climates as a means of creating supportive behaviors. *Journal on Educational Resources in Computing (JERIC)* 4, 1 (2004), 2–es.

[19] Geneva Gay and Kipchoge Kirkland. 2003. Developing cultural critical consciousness and self-reflection in preservice teacher education. *Theory Into Practice* 42, 3 (2003), 181–187.

[20] Google Inc. and Gallup Inc. 2015. Searching for Computer Science: Access and Barriers in U.S. K-12 Education. https://services.google.com/fh/files/misc/searching-for-computer-science_report.pdf

[21] Google Inc. and Gallup Inc. 2016. Computer Science Learning: Closing the Gap: Girls. https://goo.gl/dF7shN

[22] Google Inc. and Gallup Inc. 2016. Computer Science Learning: Closing the Gap: Hispanic Students. https://services.google.com/fh/files/misc/computer-science-learning-closing-the-gap-hispanic-brief.pdf.

[23] Anthony G Greenwald and Mahzarin R Banaji. 1995. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review* 102, 1 (1995), 4.

[24] Shaun R Harper and Sylvia Hurtado. 2007. Nine themes in campus racial climates and implications for institutional transformation. *New Directions for Student Services* 2007, 120 (2007), 7–24.

[25] Dawn R Johnson and Susan D Longerbeam. 2007. Implications for the Privileged Identity Exploration Model in Student Affairs Theory and Practice. *College Student Affairs Journal* 26, 2 (2007), 216–221.

[26] Aaron C Kay, Martin V Day, Mark P Zanna, and A David Nussbaum. 2013. The insidious (and ironic) effects of positive stereotypes. *Journal of Experimental Social Psychology* 49, 2 (2013), 287–291.

[27] Thomas J Leeper. 2013. Permutation Tests. https://thomasleeper.com/Rcourse/Tutorials/permutationtests.html

[28] Colleen M. Lewis and Phillip Conrad. 2020. Teaching Practices Game. https://www.csteachingtips.org/cards

[29] Colleen M. Lewis and Phillip Conrad. 2020. Teaching Practices Game: Interactive Resources for Training Teaching Assistants. In *SIGCSE Proceedings*. 1110–1111.

[30] Salvatore S. Mangiafico. 2016. Introduction to Likert Data. https://rcompanion.org/handbook/E_01.html

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

[32] Mike Marin and Ladan Hamadani. 2019. R Datasets. https://www.statslectures.com/r-scripts-datasets

[33] Peggy McIntosh. 1988. *White privilege and male privilege: A personal account of coming to see correspondences through work in women's studies*. Vol. 189. Wellesley College, Center for Research on Women, Wellesley, MA.

[34] Peggy McIntosh. 1989. White Privilege: Unpacking the Invisible Knapsack. *Peace and Freedom* (July/August 1989), 10–12.

[35] Jeffrey F Milem. 2003. The Educational Benefits of Diversity: Evidence from Multiple Sectors. In *Compelling Interest: Examining the Evidence on Racial Dynamics in Higher Education*. Stanford University Press, Stanford, CA, 126–169.

[36] Katherine L Milkman, Modupe Akinola, and Dolly Chugh. 2015. What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology* 100, 6 (2015), 1678.

[37] Margo J Monteith. 1993. Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology* 65, 3 (1993), 469.

[38] Corinne A Moss-Racusin, John F Dovidio, Victoria L Brescoll, Mark J Graham, and Jo Handelsman. 2012. Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences* 109, 41 (2012), 16474–16479.

[39] National Science Foundation. 2019. Broadening Participation in Computing (BPC). https://www.nsf.gov/cise/bpc/. Accessed: 2020-07-25.

[40] National Science Foundation, National Center for Science and Engineering Statistics. 2017. Employed scientists and engineers, by ethnicity, race, occupation, highest degree level, and sex: 2017. https://ncses.nsf.gov/pubs/nsf19304/data.

[41] An Nguyen and Colleen M Lewis. 2020. Competitive Enrollment Policies in Computing Departments Negatively Predict First-Year Students' Sense of Belonging, Self-Efficacy, and Perception of Department. In *SIGSCE Proceedings*. 685–691.

[42] Brian A Nosek, Frederick L Smyth, Natarajan Sriram, Nicole M Lindner, Thierry Devos, Alfonso Ayala, Yoav Bar-Anan, Robin Bergh, Huajian Cai, Karen Gonsalkorale, et al. 2009. National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences* 106, 26 (2009), 10593–10597.

[43] Federal Bureau of Investigation. 2018. 2017 Hate Crime Statistics Released. https://www.fbi.gov/news/stories/2017-hate-crime-statistics-released-111318

[44] B Keith Payne and Heidi A Vuletich. 2018. Policy insights from advances in implicit bias research. *Policy Insights from the Behavioral and Brain Sciences* 5, 1 (2018), 49–56.

[45] Arin N Reeves. 2014. *Written in black & white: Exploring confirmation bias in racialized perceptions of writing skills*. Technical Report. Nextions LLC., Chicago, IL. Yellow Paper Series.

[46] Marie-Élène Roberge and Rolf Van Dick. 2010. Recognizing the benefits of diversity: When and how does diversity increase group performance? *Human Resource Management Review* 20, 4 (2010), 295–308.

[47] Loriann Roberson, Carol T Kulik, and Molly B Pepper. 2001. Designing effective diversity training: Influence of group composition and trainee experience. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 22, 8 (2001), 871–885.

[48] Milton Rokeach and Ray Cochkane. 1972. Self-Confrontation and Confrontation With Another as Determinants of Long-Term Value Change. *Journal of Applied Social Psychology* 2, 4 (1972), 283–292.

[49] Milton Rokeach and D Daniel McLellan. 1972. Feedback of Information About the Values and Attitudes of Self and Others as Determinants of Long-Term Cognitive and Behavioral Change. *Journal of Applied Social Psychology* 2, 3 (1972), 236–251.

[50] Jessica Salvatore and J Nicole Shelton. 2007. Cognitive costs of exposure to racial prejudice. *Psychological Science* 18, 9 (2007), 810–815.

[51] Howard Schuman, Charlotte Steeh, Lawrence Bobo, and Maria Krysan. 1997. *Racial attitudes in America: Trends and interpretations*. Harvard University Press.

[52] Kimberly A Scott, Kimberly M Sheridan, and Kevin Clark. 2015. Culturally responsive computing: A theory revisited. *Learning, Media and Technology* 40, 4 (2015), 412–436.

[53] Niral Shah. 2017. Race, ideology, and academic ability: A relational analysis of racial narratives in mathematics. *Teachers College Record* 119, 7 (2017), 1–42.

[54] Del Siegle. 2015. Excel Spreadsheet to Calculate Instrument Reliability Estimates. https://researchbasics.education.uconn.edu/excel-spreadsheet-to-calculate-instrument-reliability-estimates/

[55] John Oliver Siy and Sapna Cheryan. 2013. When compliments fail to flatter: American individualism and responses to positive stereotypes. *Journal of Personality and Social Psychology* 104, 1 (2013), 87.

[56] Daniel Solorzano, Miguel Ceja, and Tara Yosso. 2000. Critical race theory, racial microaggressions, and campus racial climate: The experiences of African American college students. *Journal of Negro Education* 69, 1/2 (2000), 60–73.

[57] Claude M Steele and Joshua Aronson. 1995. Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology* 69, 5 (1995), 797.

[58] Claude M Steele, Steven J Spencer, and Joshua Aronson. 2002. Contending with group image: The psychology of stereotype and social identity threat. In *Advances in experimental social psychology*. Vol. 34. Elsevier, 379–440.

[59] Derald Wing Sue. 2010. *Microaggressions in everyday life: Race, gender, and sexual orientation*. John Wiley & Sons, Hoboken, NJ.

[60] Derald Wing Sue, Christina M Capodilupo, and Aisha Holder. 2008. Racial microaggressions in the life experience of Black Americans. *Professional Psychology: Research and Practice* 39, 3 (2008), 329.

[61] Derald Wing Sue, Christina M Capodilupo, Gina C Torino, Jennifer M Bucceri, Aisha Holder, Kevin L Nadal, and Marta Esquilin. 2007. Racial microaggressions in everyday life: implications for clinical practice. *American Psychologist* 62, 4 (2007), 271.

[62] Derald Wing Sue, Annie I Lin, Gina C Torino, Christina M Capodilupo, and David P Rivera. 2009. Racial microaggressions and difficult dialogues on race in the classroom. *Cultural Diversity and Ethnic Minority Psychology* 15, 2 (2009).

[63] Valerie Jones Taylor and Gregory M Walton. 2011. Stereotype threat undermines academic learning. *Personality and Social Psychology Bulletin* 37, 8 (2011), 1055–1067.

[64] Luther Tychonievich and James P. Cohoon. 2020. Lessons Learned from Providing Hundreds of Hours of Diversity Training. In *SIGSCE Proceedings*. 206–212.

[65] United States Census Bureau. 2017. Selected Population Profile in the United States: Native Hawaiian and Other Pacific Islander Alone (500-599. https://data.census.gov/cedsci/table?t=050-NativeHawaiianandOtherPacificIslanderalone(500-599)&tid=ACSSPP1Y2017.S0201&q=S0201. Accessed: 2020-07-21.

[66] United States Census Bureau. 2017. Sex by Age (American Indian and Alaska Native Alone). https://data.census.gov/cedsci/table?q=B01001C-SEXBYAGEAMERICANINDIANANDALASKANATIVEALONE&tid=ACSDT1Y2018.B01001C. Accessed: 2020-07-22.

[67] United States Census Bureau. 2018. Selected Population Profile in the United States: Two or More Races. https://data.census.gov/cedsci/table?t=100-Twoormoreraces&tid=ACSSPP1Y2018.S0201&hidePreview=false&vintage=2018#. Accessed: 2020-08-02.

[68] United States Census Bureau. 2018. Table 1. Population by Sex, Age, Hispanic Origin, and Race: 2018. https://www.census.gov/data/tables/2018/demo/hispanic-origin/2018-cps.html. Accessed: 2020-07-21.

[69] United States Census Bureau. 2019. Table 1. Population by Sex and Age, for Black Alone and White Alone, Not Hispanic: 2019. https://www.census.gov/content/census/en/data/tables/2019/demo/race/ppl-ba19.html. Accessed: 2020-07-21.

[70] United States Census Bureau. 2019. Table 15. Population by Sex and Age, for Asian Alone or in Combination and White Alone, Not Hispanic: 2019. https://www.census.gov/content/census/en/data/tables/2019/demo/race/ppl-ac19.html.

[71] U.S. Bureau of Labor Statistics. 2018. Table 1. Employment status of the civilian noninstitutional population 16 years and older by gender and race, 2018 annual averages. https://www.bls.gov/opub/reports/race-and-ethnicity/2018/home.htm.

[72] U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System. 2016. Statistics for Completed Computing Degrees. BPCnet Resource Portal. https://bpcnet.org/statistics/

[73] Sherry K Watt. 2007. Difficult Dialogues, Privilege and Social Justice: Uses of the Privileged Identity Exploration (PIE) Model in Student Affairs Practice. *College Student Affairs Journal* 26, 2 (2007), 114–126.

[74] Jared Wilber. 2019. The Permutation Test. https://www.jwilber.me/permutationtest/

[75] Frieda Wong and Richard Halgin. 2006. The "model minority": Bane or blessing for Asian Americans? *Journal of Multicultural Counseling and Development* 34, 1 (2006), 38–49.