Rethinking the Design of Online Professor Reputation Systems

Haley Tatum

Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_theses

Part of the Graphics and Human Computer Interfaces Commons, Other Computer Sciences Commons, Software Engineering Commons, and the Systems Architecture Commons

Recommended Citation

https://digitalcommons.lsu.edu/gradschool_theses/5515

This Thesis is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Master's Theses by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
RETHINKING THE DESIGN OF ONLINE PROFESSOR REPUTATION SYSTEMS

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science

in

The Division of Computer Science and Engineering

by
Haley Elizabeth Tatum
B.S. Computer Science, Louisiana State University, 2020
May 2022
Dedication

To my family.
Acknowledgments

I would like to express my deepest gratitude and appreciation for my advisor, Dr. Anas Mahmoud, for his ongoing support during my M.S. study. His patience, motivation, dedication, and expertise have been invaluable to both my research and my career. I could not have imagined having a better advisor and mentor for my time at Louisiana State University. I also would like to thank my thesis committee, Dr. Bijaya Karki and Dr. Andrew Webb for their encouragement and guidance throughout this process.

Additionally, I would like to thank and acknowledge Adrienne Steele. Your guidance, advice, and motivational leadership have molded me into the person I am today. I cannot imagine my experience at LSU without the encouragement I received from you or from my peers in the Society of Peer Mentors. Thank you for believing in me. To all my dear friends, thank you for your support. I am lucky to have intelligent and ambitious peers. I will always cherish the laughter, accomplishments, and celebrations I had with each of you.

Finally, I would like to thank my family. Without your unconditional love and support, my journey at LSU would never have been possible. I want to thank my wonderful parents, David and Debby, and my grandmother, Clara for believing in me and supporting me every step of my career. This accomplishment is as much yours as it is mine or even more so. You are my inspiration.
# Table of Contents

ACKNOWLEDGEMENTS ........................................ iv

LIST OF TABLES ............................................. v

LIST OF FIGURES ........................................... vi

ABSTRACT .................................................... vii

CHAPTER

1. INTRODUCTION ........................................... 1

2. RESEARCH METHOD ...................................... 3
   2.1. Research Questions ................................ 3
   2.2. Identifying Related Evidence ...................... 4
   2.3. Including and Excluding Evidence ................. 5
   2.4. Quantitative Analysis .............................. 7
   2.5. Evidence Categorization ............................ 7

3. BIAS IN OPR SYSTEMS .................................. 10
   3.1. Sexism and Physical Appearance .................. 11
   3.2. Racism ............................................. 15
   3.3. Ageism ............................................. 17
   3.4. Classism .......................................... 18
   3.5. Discipline and Difficulty .......................... 20
   3.6. Other .............................................. 22

4. DESIGN PROPOSALS FOR BIAS MITIGATION .......... 24
   4.1. Preventive Design Strategies ...................... 25
   4.2. Corrective Design Strategies ..................... 30

5. DISCUSSION ............................................. 32

6. CONCLUSIONS AND FUTURE WORK .................... 35

REFERENCES ............................................... 36

VITA ......................................................... 43
List of Tables

2.1. The Decision Making Protocol. ............................................. 6

2.2. The Research Venues of our Included Primary Studies and their Scope. ...... 9

3.1. Different Types of Bias in the Literature, their Definition, and Papers Classified Under Each Type ............................. 10
List of Figures

2.1. The Number of Publications Included per Year from 2010 to 2020. ........... 7
2.2. A Schematic Diagram of our Primary Study Search Process ..................... 8
3.1. RateMyProfessors Profile with Chili Pepper ........................................... 11
4.1. A Feature Diagram of the Suggested Bias-Mitigation Design Strategies....... 25
4.2. A Yelp! Reviewer with Elite Badge......................................................... 27
4.3. RateMyProfessors Rating Tags ................................................................. 28
4.4. Amazon’s Rate Features of a Vacuum Cleaner ........................................... 29
Abstract

Online Professor Reputation (OPR) systems, such as RateMyProfessors.com (RMP), are frequently used by college students to post and access peer evaluations of their professors. However, recent evidence has shown that these platforms suffer from major bias problems. Failing to address bias in online professor ratings not only leads to negative expectations and experiences in class, but also poor performance on exams. To address these concerns, in this thesis, we study bias in OPR systems from a software design point of view. At the first phase of our analysis, we conduct a systematic literature review of 23 interdisciplinary studies on bias problems affecting OPR systems. Our objective is to systematically categorize and synthesize existing evidence and identify features of OPR systems which enable offline patterns of bias to flourish online. In the second phase, we propose several preventive and corrective software design strategies to mitigate bias in OPR systems. Our objective is to highlight evidence-based design tactics that software engineers can use to develop OPR systems that are immune to bias by design.
Chapter 1. Introduction

Online Professor Rating (OPR) systems, such as RateMyProfessors.com (RMP), offer anonymous student-generated evaluations of professors in the form of text reviews, numerical ratings, and descriptive tags [33]. Over the past few years, OPR systems have received a growth in popularity among college students [25]. Recent research has shown that students find OPR systems to be helpful, credible, and a valuable source of information when deciding which classes to enroll in [30, 55, 7]. As of 2021, the largest OPR system, RMP, has indexed more than 19 million reviews of around 1.7 million professors across almost 7,500 universities and colleges from all over the world.

This rapid growth of trust in OPR systems by students and the mainstream media has granted such platforms considerable attention in educational research [22, 73, 24, 31]. In general, this line of research aims to uncover students’ motivations for writing reviews, expose bias in these reviews, examine the pairwise correlations among the different rating criteria, and study the overall quality of reviews and their impact on students academic decisions in different disciplines [36, 55, 26, 52].

Despite their popularity, recent research has exposed several limitations impacting current OPR systems. For instance, as with most anonymous systems, there is no verification process for users, thus the legitimacy of information on OPR systems is often questionable [32]. Any person, regardless of their academic status, has the ability to rate a professor even if they have never enrolled in their courses. Reviews can be written by students, colleagues, family members, and even professors themselves [52]. Furthermore, numerous studies have exposed major bias problems affecting OPR systems. Bias is the tendency to feel or show favoritism towards someone or something, usually in a way considered to be unfair. Patterns of systematic bias have been detected in reviews, ratings, and descriptive
tags (e.g., helpful, caring, easy, etc). These patterns include problems of racism, sexism, classicism, and ageism [15, 60, 53, 58, 21, 68, 31, 59, 18, 46]. Bias in OPR reviews has been linked to poor attitudes of students towards professors, lower expectations about the educational outcomes of courses, and even sub-optimal performance on exams [14, 13].

Existing research often investigates bias in OPR systems from a socio-academic point of view. However, evidence on the specific features of OPR systems which enable a complex social problem such as bias to transfer online is still underdeveloped. In general, existing evidence often appear across a broad range of interdisciplinary venues, including education, psychology, and social work. Identifying, synthesizing, and interpreting such evidence requires considerable manual effort. This can be particularly problematic in agile environments, where the majority of effort is geared towards solution development rather than problem research. To bridge this knowledge gap, in this thesis, we conduct a first-of-its-kind effort to systematically consolidate interdisciplinary research on bias in OPR systems. Our objectives are to a) systematically categorize and synthesize this body of research and b) propose several evidence-based design strategies to help developers deliver OPR systems that are immune to bias by design.

The remainder of this thesis is organized as follows. Chapter 2 describes our research method, which includes our research questions and the process for identifying primary studies. Chapter 3 categorizes and summarizes existing evidence on bias in OPR systems. Chapter 4 proposes several evidence-based design strategies for countering issues of bias in OPR systems. Chapter 5 discusses the implications of our findings and the limitations of our study. Finally, Chapter 6 concludes the paper.
Chapter 2. Research Method

In this thesis, we adapt an evidence-based approach for mitigating issues of bias in OPR systems. Evidence-Based Software Engineering (EBSE) refers to a research methodology which aims to integrate high quality research evidence into the decision-making process during the different phases of the software engineering process, from idea conception, to formal specifications, to software implementation, testing, delivery, and maintenance [34, 48]. Evidence-based Software Engineering studies are often described as a multi-step process [34]. In the first step, the research goals are articulated as research questions. In the second step, evidence related to the formulated questions is located. In the third step, the quality of extracted evidence is systematically evaluated using a pre-defined systematic process of quality assessment. In the fourth step, high quality evidence is interpreted in the context of software engineering and applicable evidence is then integrated into software engineering practices. Finally, in the fifth step, the process is reviewed for future improvements.

In this chapter, we conduct a Systematic Literature Review (SLR) of research on bias in OPR systems. SLR is a research methodology that consists of three main steps: planning the review, conducting the review, and reporting the results [35]. Next, we describe each of these steps in detail.

2.1. Research Questions

The first step of evidence-based research is to formulate the research questions. Such questions establish and justify the need for information as well as identify the scope of studies (papers) to be included in the evidence search process. In this thesis, our research questions are:
• **RQ1:** *What type of evidence is available on bias in OPR systems?* Under this research question, we locate and quantitatively analyze existing evidence on the different forms of bias affecting popular OPR systems. Our objective is to generate basic statistics (e.g., year, venue, and discipline) of this body of research. This information is intended to create a frame of reference for researchers and practitioners interested in the topic [50].

• **RQ2:** *What are the main types of bias affecting OPR systems?* Under this question, we identify, through a systematic coding process, the specific forms of bias affecting OPR systems and the main features of OPR systems which enable these forms of bias to transfer online.

• **RQ3:** *Can issues of bias affecting OPR systems be mitigated?* Under this question, we identify multiple software design strategies that can be used to mitigate issues of bias affecting OPR systems. We also discuss the potential impact of these suggested design tactics on the overall reliability and authenticity of these systems.

2.2. Identifying Related Evidence

Our scope includes primary studies tackling bias issues in current Online Professor Reputation Systems. To identify these papers, we followed a three-step process. At the first step, we searched Google Scholar, ACM Digital Library, IEEE Xplore, and arXiv for evidence. To conduct the search process, we started by formulating our search query. The synonyms *discrimination, prejudice,* and *bigotry* were included in our search query to account for language variations. In addition to these terms, we considered the specific types of bias, such as *racism, sexism, ageism,* and *classism.* Our search query can be described as follows:
We enforced the time scope of 2010 - 2020 on selected papers to make sure that we only tackle recent evidence. In total, 133 papers were found using our iterative search process.

2.3. Including and Excluding Evidence

Inclusion and exclusion criteria are commonly used in review-based studies to identify high quality studies that are related to the research goals of the study. Such criteria are determined beforehand during the planning phase. Our inclusion criteria in this thesis are:

- Studies (Books, papers, and technical reports) which explicitly address bias in public online professor reputation systems.
- Studies that are published in English.

The following exclusion criteria are used to eliminate any papers that were irrelevant to our research questions:

- Papers less than 4 pages and grey literature.
- Existing SLRs on the topic.
- Duplicate papers. In case of duplication, the most recent version is selected. In case of a paper extension both papers are considered.
To include and exclude papers, each paper was examined by three experts individually, including both authors and an external expert. All three experts hold professional degrees in software engineering and human computer interaction. Each expert examined through the title, abstract, and the body of each of the 133 papers to determine their degree of relevance to our research questions. Each expert marked each paper as either Include, Neutral, or Exclude. The paper was then included or excluded based on the majority vote as shown in Table 2.1.

<table>
<thead>
<tr>
<th>Expert1</th>
<th>Expert2</th>
<th>Expert3</th>
<th>Final Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>INCLUDE</td>
<td>INCLUDE</td>
<td>INCLUDE/EXCLUDE/NEUTRAL</td>
<td>Include</td>
</tr>
<tr>
<td>EXCLUDE</td>
<td>EXCLUDE</td>
<td>INCLUDE/EXCLUDE/NEUTRAL</td>
<td>Exclude</td>
</tr>
<tr>
<td>NEUTRAL</td>
<td>NEUTRAL</td>
<td>INCLUDE/EXCLUDE/NEUTRAL</td>
<td>Consensus Meeting</td>
</tr>
<tr>
<td>INCLUDE</td>
<td>EXCLUDE</td>
<td>NEUTRAL</td>
<td>Consensus Meeting</td>
</tr>
</tbody>
</table>

Conflicts were resolved using majority voting. After applying the inclusion and exclusion criteria to the initial set of retrieved papers, 18 papers were included in our analysis. We observed during this process was that a large number of papers were specific to Student Evaluations of Teachers (SET) with no discussion of aspects related to OPR systems. These papers were excluded. Furthermore, papers which study the motivations of students using OPR systems without explicitly discussing bias were excluded.

To reduce the risk of false negatives (paper omission), a backward-snowballing was applied on the set of included primary studies [72]. The process involved going through the list of citations for each of the already included studies to identify any potentially-relevant missed studies. In total, 25 more papers were identified and reviewed. Of these papers, 4 were included, raising the number of primary studies to be included in our study to 23
papers. The number of included primary studies each year from 2010 to 2020 are shown in Fig. 2.1. A summary of the entire evidence search process is depicted in the diagram in Fig. 2.2.

![Bar chart showing the number of publications included per year from 2010 to 2020.](chart.png)

*Figure 2.1. The Number of Publications Included per Year from 2010 to 2020.*

### 2.4. Quantitative Analysis

To answer RQ$_1$, we examined each paper to determine which OPR systems are discussed or studied. Our results show that the most commonly studied OPR system is RateMyProfessors (RMP), with 21 of the 23 studies investigating this platform. Other studied platforms include Koofers.com, which was examined in one of the papers [25], and Misprofesores.com, a version of RMP used in Mexico, examined by one paper [49]. The primary studies included in our review mainly appeared in educational venues. Some of these venues are domain specific (e.g. Economics [49] and Business [5]). These venues and their specific scopes are shown in Table 2.2.

### 2.5. Evidence Categorization

A grounded-theory approach is adopted to categorize evidence in the set of included studies [10]. The coding process involved three software experts individually examining the
At the end of our classification process, five different categories of bias have emerged in our set of primary studies: sexism, racism, ageism, classism, and discipline and difficulty. In the following chapter, we summarize the evidence classified under each category.
Table 2.2. The Research Venues of our Included Primary Studies and their Scope.

<table>
<thead>
<tr>
<th>Venue</th>
<th>Scope</th>
<th>Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment and Evaluation in Higher Education</td>
<td>Higher Education</td>
<td>[55, 8]</td>
</tr>
<tr>
<td>Practical Assessment, Research and Evaluation</td>
<td>Higher Education</td>
<td>[4, 3, 64, 54]</td>
</tr>
<tr>
<td>Research in Higher Education Journal</td>
<td>Higher Education</td>
<td>[29, 59]</td>
</tr>
<tr>
<td>Diversity in Higher Education</td>
<td>Higher Education</td>
<td>[53]</td>
</tr>
<tr>
<td>Economics of Education Review</td>
<td>Economics</td>
<td>[49]</td>
</tr>
<tr>
<td>ACM SIGCSE Technical Sym. on Comp. Sci. Education</td>
<td>Computer Science</td>
<td>[21]</td>
</tr>
<tr>
<td>Journal of Marketing Education</td>
<td>Marketing</td>
<td>[1]</td>
</tr>
<tr>
<td>The International Journal of Management Education</td>
<td>Management</td>
<td>[9]</td>
</tr>
<tr>
<td>Machine Learning and Applications</td>
<td>Data Science</td>
<td>[2]</td>
</tr>
<tr>
<td>Data Science</td>
<td>Data Science</td>
<td>[43]</td>
</tr>
<tr>
<td>Advances in Social Network Analysis and Mining</td>
<td>Web Science</td>
<td>[25]</td>
</tr>
<tr>
<td>Information Systems Education Journal</td>
<td>Information Systems</td>
<td>[57]</td>
</tr>
<tr>
<td>PLoS One</td>
<td>Cross Disciplinary</td>
<td>[46]</td>
</tr>
<tr>
<td>American Journal of Criminal Justice</td>
<td>Criminal Justice</td>
<td>[31]</td>
</tr>
<tr>
<td>Workshop on Noisy User-generated Text</td>
<td>Linguistics</td>
<td>[68]</td>
</tr>
<tr>
<td>Language in Society</td>
<td>Linguistics</td>
<td>[60]</td>
</tr>
</tbody>
</table>
Chapter 3. Bias in OPR Systems

Our review shows that different types of bias commonly manifest within the textual comments [56], numerical ratings [47], and descriptive tags [43] of OPR systems. In general, bias can be either implicit or explicit. Implicit bias is a type of discriminatory behavior or tendency that the user is not actively aware of, including any unconscious association or belief (mainly stereotypical) toward any particular group of people [47]. Such bias is often detected through correlation analysis between the numerical ratings and personal attributes of professors, such as their age [59], ethnicity [53, 68, 21], or even social class [8, 31, 15]. Explicit (conscious) bias refers to more obvious types of bias that are intentional and controllable, such as a student leaving a racially charged review for a professor. Such bias is often detected by identifying discriminatory language or prejudice in the textual comments professors receive [59, 68].

Different types of bias are identified in the literature. Table 3.1 shows these types, their description, and the papers investigating each type. In what follows, we discuss each of these types.

Table 3.1. Different Types of Bias in the Literature, their Definition, and Papers Classified Under Each Type

<table>
<thead>
<tr>
<th>Type</th>
<th>Bias against:</th>
<th>Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sexism &amp; Attractiveness</td>
<td>gender, sexual orientation or physical appearance</td>
<td>[31, 2, 55, 43, 5, 49, 68, 15, 46, 64, 21,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>57, 29, 53]</td>
</tr>
<tr>
<td>Racism</td>
<td>ethnicity, color, nationality, or accent</td>
<td>[31, 46, 53, 21, 60]</td>
</tr>
<tr>
<td>Ageism</td>
<td>older or younger aged individuals, years of experience</td>
<td>[15, 46, 59]</td>
</tr>
<tr>
<td>Classism</td>
<td>tenure, social, or professional status</td>
<td>[15, 31, 46, 9, 8, 29]</td>
</tr>
<tr>
<td>Discipline</td>
<td>course topic, department, field, or difficulty level of content</td>
<td>[2, 55, 57, 15, 5, 59, 31, 29, 46, 9, 3, 4, 54, 25, 8]</td>
</tr>
<tr>
<td>Other</td>
<td>Other types of discrimination</td>
<td>[1]</td>
</tr>
</tbody>
</table>
3.1. Sexism and Physical Appearance

Under this category, we include papers which discuss sexism, or bias based on gender or sexual orientation in OPR systems. Gender information is not readily available on OPR systems. To generate such information, researchers often resort to indirect methods of gender identification, such as gender classification based on pronouns used in the comments [55, 68, 5, 53, 15, 43, 21], matching names with U.S. Social Security names [2, 55, 15], third-party gender identification services [2], or manually analyzing professors’ pictures from their university profiles [53].

Up until 2018, students were able to rate professors on RMP based on their physical appearance or “hotness” [58]. Specifically, each professor’s profile displayed a chili pepper icon (Fig. 3.1) which correlated to how physically attractive they were perceived to be by students [58, 5, 42]. While this explicit form of sexism was (for obvious reasons) removed, sexism based on appearance can still be detected implicitly through analysis of numerical ratings.

Figure 3.1. RateMyProfessors Profile with Chili Pepper

Several studies have shown a significant correlation between the physical appearance and overall ratings of professors in OPR systems [55, 52, 58]. In general, professors who
are perceived as more attractive are often granted more positive ratings [68, 31, 59]. For instance, Johnson and Crews [31] examined the relation between 407 professors’ personal characteristics (collected from university profiles) and their RMP ratings of easiness, helpfulness, clarity, overall quality, and hotness, using multiple regression analyses. The results showed that professors were more likely to be rated significantly higher in clarity, helpfulness, and overall quality if they were male, white, and perceived as attractive.

Antoine et al. [2] used Latent Dirichlet Allocation (LDA) to model gender bias in a corpus of professor reviews retrieved from RMP (29,952 professors and 437,521 reviews). A series of regression models were then used to analyze the relations between professors’ ratings and the topics discovered. The results revealed that the frequency of words associated with negative teaching was higher for female professors. The results further showed that physical attractiveness and difficulty remained important determinants of the overall quality rating of professors. Specifically, the authors found that being perceived as “hot” led to an increase in the overall quality score by 0.74 points (on a 5-point scale).

Rosen [55] analyzed a large data set of RMP reviews (190,006 professors and 7,882,980 reviews) for correlations between instruction quality, easiness, physical attractiveness, discipline, and gender of professors. Significant positive correlations were detected between ratings of instruction quality and easiness and attractiveness. On average, female professors scored 0.4-0.5 points lower on metrics of instruction quality compared to male professors.

Liu et al. [43] collected and analyzed close to 10,000 RMP reviews, overall quality scores, and tag distributions of male and female professors from different academic disciplines. The results showed that in liberal arts disciplines, female professors were rated much lower in comparison to male professors. A distribution analysis of descriptive tags further showed that students preferred male professors more than female professors.
Boehmer and Wood [5] examined the hypothesis that RMP ratings showed a gender bias or preference in favor of male faculty. The data included 92 professors’ academic records as well as their RMP data. After controlling for reported grades, years of service, hotness, easiness, and academic department, male professors had on average 0.275 points more on a scale of 5 than female professors.

Arceo-Gomez and Campos-Vazquez [49] analyzed gender stereotypes in student evaluations of college professors on MisProfesores.com. The analysis was conducted using 600,000 evaluations of 64,577 professors and teachers in Mexico. The results showed that female professors received, on average, lower ratings than male professors. In particular, the analysis of students’ reviews showed that they referred more to the physical appearance and personality traits of female professors than male professors. Female professors were also described with more negative words such as “bad” or “strict” and often referred to as “teachers” while calling their male counterparts “professors”.

Waller and Gorman [68] used an ensemble of classifiers to detect objectifying language in RMP student reviews. Objectifying reviews were defined as reviews that described a professor’s physical appearance, demeanor, clothing style, or resemblance. The results showed that female professors with high quality ratings were significantly more likely to receive attractiveness commentary than male professors with the same ratings. The results also showed that the more difficult the professor is, the less likely their reviews would contain attractiveness commentary.

Fisher et al. [15] analyzed a sample of 6,094 professors ratings (3,134 female and 2,960 male professors) from RMP. A univariate analysis of variance showed that female professors in high-status departments (Law, Engineering, Business, and Computer Science) were evaluated more negatively than female professors in low-status departments (English, Hist-
tory, Philosophy, and Art History). Interestingly, this effect was reduced when the female professor was “hot” as indicated by their red chili pepper. Specifically, a “hot female” received a RMP quality score of 4.26 in comparison to 2.57 for a not “hot” female.

Murray et al. [46] analyzed public RMP student evaluations of 18,946 tenure and tenure-track faculty from 399 US universities. These evaluations were matched with career and research performance indicators (publications, citations, and grants) of professors obtained from the company Academic Analytics. A linear regression model revealed that the attractiveness of faculty (the presence of the chili pepper) was one of the factors most associated with higher student ratings. Furthermore, male faculty were associated with 0.11 points greater overall teaching quality compared to female faculty.

Theyson [64] analyzed a sample of 476 instructors (306 male and 170 female) profiles on RMP along with objective ratings of their attractiveness generated based on pictures from the websites of their universities. The results confirmed that instructors who were perceived to be hot and easy by their students received higher overall quality evaluations. The results also showed that both male and female instructors equally benefited from “hotness”.

Gordon and Alam [21] analyzed the RMP ratings of 39,000 Computer Science (CS) professors to examine the role of race and gender in their teaching evaluations. Using frequency distribution of RMP feedback tags, the analysis revealed that female professors were rated lower than their male counterparts in overall teaching quality. Female professors were also perceived to be less humorous, and less inspirational.

In a study of 820 RMP evaluations from information systems, marketing, and management courses from 34 universities, Sena and Crable [57] examined the relation between professors’ gender, courses’ level (introductory course vs. senior-level), and courses’ subject (management, marketing, information systems). The results showed that male professors
of information systems received higher overall average ratings than females in information systems, although the results were statistically-insignificant (p=0.18). In marketing, however, females had a higher (but still insignificant) average overall score than males.

Jalbert [29] analyzed RMP teaching evaluations of 300 business professors. The results showed that female professors received higher teaching ratings than male professors. Furthermore, a significant relationship was observed between professors “hotness” and students’ evaluation of teaching quality.

3.2. Racism

Under this category, we include papers which discuss racism, or bias based on race, ethnicity, or accent. Racist behavior on OPR systems can be detected explicitly through analyzing the textual reviews of professors, such as denoting the accent of a non-native professor in a review [61]. OPR systems do not list the race of professors. Therefore, researchers use other methods for determining or predicting race. Common methods of race (ethnicity) detection include using US-census data to predict the race based on last name [21, 46, 60], manually examining publicly accessible photos of professors to predict the race [31], or a combination of the two methods [53].

Bias based on race has been extensively studied in related literature. For instance, in their study of 407 criminal justice and criminology faculty members, Johnson and Crews [31] found that white instructors received significantly higher ratings than non-white instructors in the overall quality, helpfulness, and clarity ratings. Similarly, Murray et al. [46] found that faculty who were mentioned as having an accent on their RMP comments were associated with 0.17 points lower on overall quality. Furthermore, those with a commonly white surname were associated with 0.118 points higher scores for overall teaching quality.
than those with an unknown (or not commonly white) surname. Gordon and Alam [21] found that professors from racial minority backgrounds, especially Asian, tend to receive a higher frequency of negative tags such as “tough grader” than other races.

In a study by Reid [53], the RMP ratings of 3,717 faculty members (3,079 White, 142 Black, 238 Asian, 130 Latino, and 128 other) from 25 different liberal arts colleges were extracted to examine the role of perceived race and gender on RMP ratings. The authors found that racial minority faculty were rated significantly less favorable than their white counterparts in overall teaching quality, helpfulness, and clarity. The author additionally examined the relationship between instructor race and gender using analysis of variance (ANOVA). The addition of gender revealed that black male faculty were rated more negatively than any other combination of race and gender.

Subtirelu [60] compiled a list of reviews from 1,096 instructors with US last names and a list of reviews from 1,096 instructors with Chinese or Korean last names. Non-parametric Wilcoxon rank sum test was then used to determine the difference in clarity and helpfulness ratings between the different professor groups. The results showed that professors with common US last names scored higher on clarity and helpfulness. The authors also found that comments which mentioned an instructor’s accent were frequently followed by the word “but”, suggesting an initial problem with having an accent. Furthermore, students tend to comment on the language of Chinese or Korean instructors more frequently compared to instructors with common US last names [60]. For example, “Don’t take him unless you know Chinese. Because he obviously can’t speak English.”
3.3. Ageism

Under this category, we include papers which discuss ageism, or bias based on a professor’s age or their years of experience. OPR systems do not provide information on a professor’s age. Therefore, researchers often use alternative ways to predict age, such as calculating the number of years since a professor received their bachelor’s degree [59] or estimating age based on the year a professor was hired [15]. In some cases, academic age, or the number of years within a particular academic field, is used [46].

The impact of age, or tenure, of professors on their OPR system reviews has also been investigated in related literature. The goal of this type of analysis is to expose how a professor’s age would impact the students’ views of them, whether in comments or ratings. For instance, Fisher et al. [15]’s study suggested that less experienced (younger) female faculty were more vulnerable to backlash on RMP than more experienced (older) female faculty. Newly hired female professors were found to be the most susceptible to this backlash, potentially due to age, lack of experience, rank, or any combination of the three.

Murray et al. [46] measured the impact of scientific age, or the number of years since the instructor obtained their terminal degree (highest degree awarded for that field), on their RMP ratings. The authors discovered a negative relationship between scientific age and overall teaching quality, where each additional decade was associated with 0.13 points lower for the overall rating. Generally, the findings suggested that older faculty received lower evaluations.

Stonebraker and Stone [59] used a large data set from RMP to investigate the impact of a professors’ age on their perceived effectiveness. The authors found that age has a negative impact on student ratings of faculty. This effect was robust across genders, academic disciplines, and even type of institutions (universities and colleges) and seemed to
manifest when professors hit their mid-forties. This effect did not necessarily increase when a professor reached the retirement ages of 65 or 70. When the data was reduced to only professors who were perceived as “hot”, the effect of age entirely disappeared.

3.4. Classism

Classism, in general terms, refers to discrimination or prejudice against people from a different social class. In academia, classism refers to bias based on a professor’s academic title, department status, or educational background (university prestige). On RMP, the term “professor” is used to represent all instructors who have a profile on the platform regardless of their tenure-status or adjunct status. Therefore, in order to determine a professor’s professional rank, researchers must distinguish between full professors, assistant professors, associate professors, full-time instructors, or graduate students. Common methodologies to predict professional rank include using third party services such as Academic Analytics [46], university salary databases [9], and manually examining professors’ university profiles [31].

The impact of classism on professors ratings was clear in Fisher et al. [15]’s study, who found that female professors in high-status departments (Law, Engineering, Business, and Computer Science) were rated more negatively than female professors in low-status departments (English, History, Philosophy, and Art History).

In their study of criminal justice and criminology faculty, Johnson and Crews [31] reported that professional characteristics of instructors, such as the possession of a doctorate, tenure status, years of teaching experience, and scholarly publication rate produced a weaker association to RMP scores than physical characteristics.

Murray et al. [46] found that associate professors’ overall teaching quality ratings on RMP were associated with 0.05 points lower than assistant professors, and 0.14 points
lower than full professors. These findings suggest that the relationship between seniority and overall ratings may not be linear; those with experience, but not too much seniority, tend to do best. The authors further found an effect of research activity on overall quality ratings of faculty. For instance, professors with no publications received 0.034 points less for overall quality while those with a high level of publications were associated with about 0.022 points higher on the overall quality. Additionally, the authors found that faculty affiliated with public universities were associated with about 0.08 points lower in overall quality ratings than faculty affiliated with private universities.

In an analysis of 3,570 RMP ratings of 264 professors at 9 Florida universities, Constand et al. [9] examined how ratings can differ across different types of management courses. A series of t-tests revealed that an associate professor status tends to be negatively correlated with quality, clarity, helpfulness, and easiness scores. On average, instructors with the titles of associate professor and full-time instructor earned lower overall quality ratings than those with a full professor title.

Chiu et al. [8] performed extensive data exploration of 16,802 professors from 1,592 schools which received at least 20 ratings on RMP. The study examined the impact of being at a top university (according to Forbes’ America’s Top Colleges 2016 report) on professors ratings. The authors observed that the perceived quality of professors working for top colleges was 4.11 in comparison to 3.69 overall quality rating for professors working for other colleges. In general, the decline of perceived quality when class difficulty increases is not as steep for professors who come from top colleges.

Jalbert et al. [29] examined matched data between teaching and research performance for 300 business professors across 104 universities. Teaching evaluations were collected from RMP and research metrics were collected from the Social Science Research Network.
The results showed no statistically significant evidence of a relationship between teaching ratings and research quality or quantity. However, private university professors received significantly higher teaching ratings than public university professors.

3.5. Discipline and Difficulty

Under this subcategory, we classify papers which discuss bias against the academic discipline of courses; stemming from classroom characteristics such as a course’s topic or its perceived difficulty. In general, converging evidence indicates that professors teaching a course that is perceived to be difficult receive lower overall quality scores. For instance, Antoine et.al [2], found that an increase in perceived difficulty by a single unit led to a reduction of 0.54 points on the quality score. Similarly, Rosen [55] found a statistically significant dependency between the overall quality of a professor and their easiness score. In their study of information systems professors’ ratings on RMP, Sena and Crable [57] also found a significant negative correlation between the overall rating of professors and the perceived difficulty of the courses they teach. Fisher et al. [15] reported that female professors who taught courses perceived to be difficult, or were tough graders, had a higher chance of receiving worse evaluations, or backlash (low scores or negative comments) on RMP than those who were perceived as easy graders. Boehmer and Wood [5] reported that students assigned statistically significant higher ratings to professors who taught easy classes. In their study of age impact on ratings, Stonebraker and Stone [59] found that an extra point on the easiness scale can account for over 30 additional years of age. Johnson and Crews [31] reported that the perception that the instructor was an easy grader was the most significant predictor of helpfulness, clarity, and overall quality in the multiple regression models. Jalbert et al. [29] found that professors perceived as more difficult re-
ceived lower teaching quality ratings on RMP. Overall, students tend to attribute positive attributes to professors who provide them with higher grades and lower perceived stress.

In Chiu et al. [8]'s large-scale study of RMP ratings, the regression results indicated that a course with higher difficulty was often associated with a lower perceived quality of the professor. This effect was lowered when a professor comes from a top school. The results also seemed to differ by discipline, where the impact of perceived difficulty on perceived quality for STEM professors was in general significantly higher than for humanities professors.

In general, professors in STEM tend to receive worse instruction quality ratings compared to disciplines in the humanities and arts. Murray et al. [46] found course difficulty to be the most associated factor with lower RMP student ratings of professors. Professors in Engineering were rated the lowest in overall teaching quality, while faculty in Humanities were rated 0.18 points higher in overall quality ratings, followed by Medical Science (0.11+) and Natural Sciences (.07+).

Constand et al. [9]'s study of management professors revealed that those teaching core classes received significantly lower ratings on quality, helpfulness, clarity, and easiness compared to capstone class professors. Overall, students perceived core classes to be more difficult and less interesting.

Bleske-Rechek and Fritsch [3] examined the reliability of students’ ratings on RMP. The data included 366 instructors with 10 or more student ratings. The results showed that professors who were rated as easy consistently received significantly higher quality ratings. Furthermore, Math & Natural Sciences instructors were rated as less easy than instructors in Arts & Humanities and Social Sciences.
It is also important to point out that several researchers found that the correlation between easiness ratings and quality ratings was not always necessarily a product of student bias, thus disputing the argument that instructors instructors perceived as easy graders receive high evaluations. For example, Bleske-Rechek and Michels [4] analyzed anonymous self-report data on use of RMP from 208 students and RMP ratings of 322 instructors at a regional public university. The results uncovered a positive link between instructor easiness and instructor quality. However, these findings were not conclusive across departments. In another study, Rizvi [54] investigated the relationship between student evaluations of the instructors on RMP and the average grades awarded by those instructors. The data included the evaluations of the 419 instructors who taught for at least two semesters and had at least 10 RMP evaluations. The results uncovered a strong correlation between instructors’ overall evaluations and their easiness scores. However, the perceived easiness of an instructor did not always result in higher grades for students. The results also revealed that STEM instructors received lower overall evaluations than instructors in other fields where courses were perceived easier by the students. Hassan [25] analyzed the correlation between aggregate course outcomes and student perception of the course instructor on Koofers for 478 courses taught at Virginia Tech. The results showed that student ratings of course instructors were more correlated with course outcomes rather than student perception of the simplicity of course materials.

3.6. Other

In a field experiment of Marketing students, Ackerman and Chung [1] examined how student online course ratings can be influenced by existing ratings. Students viewed a video of a professor lecturing and then were presented with prior ratings from students who had
recently completed a questionnaire. Students were then asked to give their own rating of the class. The results suggested that prior visible positive and negative online ratings biased subsequent ratings, indicating that a mob mentality might affect OPR systems. The results also showed that online ratings influenced student course selection decisions.
Chapter 4. Design Proposals for Bias Mitigation

Reputation systems are commonly used as a *de facto* mechanism for establishing trust in online marketplaces [62]. Such systems can be either two-sided or one-sided. For example, in Airbnb, hosts can review their guests and guests can review their hosts. A main problem that is commonly associated with this approach is retaliation. For example, a subpar host might retaliate at a guest for leaving an honest review about the service they received. To mitigate this type of online retaliatory behavior, several online marketplaces, such as Amazon and EBay, entirely switched to single-sided reviews [27, 62]; only buyers can leave reviews for sellers. While this designed decision has eliminated retaliation, it limited the ability of the party at the receiving end of ratings and reviews to directly respond to malicious feedback [70].

OPR systems are an example of anonymous one-sided reputation systems, therefore, retaliation is highly unlikely unless the professor can identify students from their reviews. However, as our review revealed, OPR systems commonly suffer from malicious and biased feedback. To mitigate this problem, in this chapter, we discuss multiple software design strategies that have been proposed in the literature or deployed in practice to mitigate issues of bias (unfairly positive or unfairly negative ratings) in reputation systems. Our objective is to propose a set of evidence-based design solutions that software engineers of OPR systems can use to mitigate bias in their systems. Addressing issues of bias in OPR systems can not only help students make more informed course selection decisions, but also improve their performance in class. In fact, Edwards et al. reported that biased reviews can lead to negative expectations and experiences and poor performance on exams [14, 13]. Addressing bias in OPR systems can also reflect positively on professors, given that multiple studies have shown that professors might alter their teaching behavior (such as not covering...
In general, our suggested bias-mitigation strategies (shown in Fig. 4.1) can be categorized into two main categories: preventive and corrective.

<table>
<thead>
<tr>
<th>Bias Mitigation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preventive</strong></td>
</tr>
<tr>
<td>Controlled anonymity</td>
</tr>
<tr>
<td>Adaptive review system</td>
</tr>
<tr>
<td>Incentives</td>
</tr>
<tr>
<td><strong>Corrective</strong></td>
</tr>
<tr>
<td>Bias correction</td>
</tr>
<tr>
<td>Adjusted ratings</td>
</tr>
<tr>
<td>Biased Review filtering</td>
</tr>
<tr>
<td>Selective reveal of information</td>
</tr>
<tr>
<td>Controlled anonymity</td>
</tr>
<tr>
<td>Ratings Aggregation</td>
</tr>
<tr>
<td>Rater Credibility</td>
</tr>
</tbody>
</table>

Figure 4.1. A Feature Diagram of the Suggested Bias-Mitigation Design Strategies.

### 4.1. Preventive Design Strategies

Preventive design strategies are intended to stop bias \textit{a priori}, before a student leaves a malicious review or a biased rating for a professor. These strategies can be described as follows:

- **Controlled anonymity**: Public OPR systems are anonymous platforms. It is not possible to tie a comment or rating to a specific identifiable user. Previous evidence from other related fields has shown that complete online anonymity often enables a culture of toxicity that can diminish the perceptions of believability and accuracy of online reputation systems [36]. This effect was clear on a platform such as Yik Yak, an
anonymous social network that was forced to shutdown in 2017 due to uncontrollable levels of cyber-bullying [71]. To work around such a problem, a more controlled form of anonymity can be used [12]. For example, to prevent uncontrolled ratings in the health sector, Jabeen, et al. [28] proposed an anonymous reputation management technique. The goal of this design strategy is to keep the identity of service consumers concealed from service providers, yet visible to the system if required.

Controlled anonymity can be achieved by assigning pseudonyms to users. However, there is always a risk of misconduct if users are allowed to generate pseudonyms independently [17]. For example, students can create multiple pseudonyms using multiple email accounts. To work around this problem, OPR systems can ask students to sign up using a valid “.edu” email before they can rate a professor. Typically, students have only one “.edu” email at their institution. This will enable the platform to keep track of each student’s review behavior while keeping their identity completely concealed from professors. Furthermore, it prevents people who affiliated with a University from rating professors.

- **Incentives:** Converging evidence from the reputation systems literature suggests that some sort of incentives, such as virtual points, can encourage reviewers to be more balanced and provide less biased ratings. For example, Google awards their most active reviewers with a “Local Guide” badge. Local guides are also encouraged with other forms of incentives such as early access to new Google products. Yelp also grant an “Elite” status to recognize their active reviewers (e.g., Fig. 4.2). Elite reviewers are also rewarded with various types of perks. Zhang et al. [74] investigated the influence of incentives on Yelp’s reviewers’ behavior. The results showed that incentives can have immediate short term and long term impacts on review quality. Elite reviewers
significantly increased their contribution levels as well as the quality and length of their reviews while maintaining a lower rating variance. The authors also highlighted the positive impact of reevaluating Elite reviewers annually to maintain the quality of their incentive systems.

![Maria L. Elite 2021](image)

Baton Rouge, LA

8/21/2021

10 photos

Figure 4.2. A Yelp! Reviewer with Elite Badge.

Overall, the converging evidence suggests that a game-theoretical incentive mechanism may prompt students to provide more frequent and more objective reviews on OPR systems [20, 74, 63]. For example, a virtual point system can be designed to reward students who leave informative reviews, or if their reviews achieve a certain ratio of “likes” vs. “dislikes”. Rewards can be varied to keep students motivated to attain the next badge, thus, contribute more to the platform. In their study of bias in reputation systems, Ma et al. [44] reported that the effect of bias can diminish with more frequent, longer, and more informative reviews, given that bias in OPRs often comes from few early ratings [65]. Furthermore, badges can be reevaluated annually to maintain the quality of the incentive system.
• **Adaptive evaluations:** OPR systems offer a *one-size-fits-all* design for rating professors. However, as our review showed, professors from different disciplines are rated differently. These observations suggest that instead of rating professors based on vaguely defined constructs (e.g., Fig 4.3), professors should be rated based on dimensions that are related to their disciplines. Amazon, for instance, uses specific “Rate features” to rate different products. For example, a vacuum cleaner can be rated based on “Suction power”, “For cleaning up hair”, and “Maneuverability” (Fig. 4.4), while an office chair is rated based on “Easy to assemble”, “Sturdiness”, and “Comfort”. Similarly, Airbnb creates groups of listings (rental places) and asks users to rate specific features of these listings, such as the quality of WIFI in a house that is designated as a work space [16].

![RateMyProfessors Rating Tags](image)

Figure 4.3. RateMyProfessors Rating Tags

A similar design could be adapted to OPR systems, rather than using generic tags, such as “Caring” or “Hilarious”, adaptive tags can be automatically generated based on existing reviews for each department or college. Such tags should reflect the teaching qualities that students often consider when rating professors from different departments. This can be particularly useful for professors who teach classes perceived to be harder by rating them on aspects of the class other than the difficulty of exams or the frequency of homework.
Selective reveal of information: This design strategy is intended to control for mob mentality behavior over OPR systems. Our review revealed that new students can be significantly influenced by previous ratings [1, 38], therefore, it is important to minimize this effect in order to avoid biased-influenced reviews. This can be achieved by showing only the most recent reviews of professors, while masking the older reviews as well as the total number of reviews [51]. According to Hannák et al. [23], this design strategy can *level the playing field* for minority workers in freelancing platforms, while still providing prompt and testimonial feedback.
4.2. Corrective Design Strategies

Corrective design strategies can be used to eliminate bias after the fact, assuming the preventive strategies failed to stop bias. These tactics can be described as follows:

- **Bias correction:** This design change suggests that since bias can be objectively quantified, professors ratings can be adjusted to recompense for measurable sources of bias. In their work on reputation systems in Sharing Economy platforms (e.g., TaskRabbit and UpWork), Goel et al. [19] used the covariance between ethnicity and the aggregated reputation scores as a proxy for measuring bias. Implementing the suggested changes on a dataset of Airbnb reviews showed that adjusting for sensitive attributes such as ethnicity removed their affect, while the effect of other pertinent attributes remained significant. Similar approaches can be used in OPR systems where professors’ ratings are constantly monitored for bias and corrected accordingly.

- **Filtering biased reviews:** This design strategy is intended to systematically detect and filter out potentially biased reviews and ratings. The underlying assumption is that ratings typically follow a particular distribution, thus, extreme biased ratings should have a different statistical distribution than average fair ratings. For instance, Whitby et al. [70] proposed a statistical filtering algorithm to remove recommendations that do not lie between lower and upper quantiles. The algorithm assumes that recommendations follow a beta distribution, there is a large number of raters, and less than 30% of them behave consistently in a biased manner. Weng et al. [69] introduced a scalable entropy-based method for filtering out unfair feedback. The proposed method does not make any assumptions about the rating distribution. However, a rating is considered malicious if it appears to deviate from the majority...
opinion. In general, removing statistical outliers can help to eliminate biased reviews. However, large enough samples of reviews and ratings have to be present in order for outlier removal methods to achieve reasonable results [66].

- **Adjusted ratings**: Popular OPR platforms such as RMP display the arithmetic mean of each professor’s ratings. Existing research has revealed that this strategy can be severely misleading. For instance, Dai et al. [11] explored rating aggregation mechanisms in Yelp.com’s restaurant reviews. The authors showed that as the number of ratings grows over time, using a simple arithmetic average to aggregate ratings can produce misleading results that do not account for reviewer biases or reflect changes in the quality of service. To control for such effect, the authors suggested that an adjusted average rating should be used. For instance, more recent ratings should be assigned more weights to adapt to changes in quality of teaching (declining or improving quality). Furthermore, reviewers’ credibility should factor into their ratings. Credibility can be quantified by comparing the student’s rating of specific professors to the average rating and reviews of those professors [67] or by simply analyzing their rating history and attributes (number of ratings, general sentiment of their ratings, and history of bias). However, for this feature to be implemented, controlled anonymity should be first implemented so that the rating profiles of students can be tracked over time.
Chapter 5. Discussion

Our review shows that bias in OPR systems is a multi-dimensional socio-technical problem. The problem is further exacerbated by the rapidly accelerating pace of technological advances (access to smartphones) and the increased reliance of students on computer-mediated Word-Of-Mouth (WOM) systems for class enrollment decisions. Therefore, it is safe to say that there is no silver bullet for solving bias in OPR systems. However, satisficing solutions could be developed to alleviate the problem [45]. These solutions can be deduced from existing interdisciplinary evidence which detects and documents patterns of bias in reputation systems across a broad range of application fields through engineering studies and controlled experiments.

In this thesis, we outlined several concrete design strategies that could be used to mitigate issues of bias in OPR systems. These strategies have been validated in both research and practice and have been shown to prevent, mitigate, and correct biased content in online reputation systems. For instance, an OPR platform can start by asking students to sign up for the service using their “.edu” emails. Students should be assured that their identity will remain anonymous on the platform. This form of controlled anonymity enables the system to keep track of individual student rating history and behavior, thus, enhancing the validity and quality of reviews. The platform should also implement some sort of a reward mechanism to encourage students to leave more reviews. Evidence has shown that the difference in ratings between in-class and online evaluation can diminish when a more representative sample of students provide online evaluations [37]. The platform can also use more sophisticated mechanisms for aggregating ratings, such as showing the pattern of rating (declining or increasing rating), rather than just showing the arithmetic mean. Furthermore, ratings can be adjusted based on the credibility of individual raters which
can be estimated by tracking their rating behavior over the platform. The platform can also implement statistical-based filtering mechanisms to filter out statistical outliers.

It is important to point out that some of the proposed strategies are easier to implement than others. For instance, integrating a controlled anonymity feature for new users can be a simple fix. However, features such as bias correction, adaptive review systems, and filtering biased reviews require complex statistical models and data science components to discover patterns in reviews and correct bias accordingly. Therefore, before taking on any of the suggested design changes, developers should factor in the effort expected to develop, test, and maintain such features in their Return-ON-Investment (ROI) analysis.

In terms of threats to validity, we notice that some of our identified categories of bias were more well-defined than others. For example, several papers examined multiple types of bias in their analysis, such as sexism, racism, and ageism (e.g., [46, 31]). These papers appeared at multiple subsections. A descriptive summary of the paper, including the data and the research method, are mentioned at the first appearance of the paper in the text. Other limitations might stem from our inclusion/exclusion protocol. Specifically, we relied on our own assessment of the literature to include or exclude papers. To mitigate this threat, we applied a systematic coding protocol, which included individual coding of primary studies and majority voting. We further acknowledge the fact that the quality of our paper summaries might vary across different categories and sometimes within the same category. We tried to mitigate this threat by imposing a standard structure on our summaries.

Another threat to validity stems from the fact that none of the proposed design strategies have been independently evaluated in the context of OPR systems. This can be mainly attributed to the fact that long term reviewing behavior can be very challenging to replicate.
in a laboratory setting [16]. Ideally, suggested design tactics should be evaluated through engineering studies which are designed to evaluate any scientifically-developed ideas prior to implementation and to reduce the risk of unforeseen or underestimated circumstances[6]. For instance, OPR platforms can integrate our proposed changes into their platforms, release these changes for a selected group of users, and conduct market research on the outcomes. Such evaluation strategies are commonly used in platforms such as Airbnb, Yelp, and eBay which constantly roll out and evaluate bias-mitigating changes in their reputation systems [16, 40].
Chapter 6. Conclusions and Future work

In this paper, we explored the problem of bias in Online Professor Reputation (OPR) systems. In the first phase of our analysis, we conducted a systematic literature review of 23 papers on bias issues affecting OPR systems, published in the period between 2010 and 2020. Our results show that common forms of offline bias, such as sexism, racism, ageism, and classism are widely spread in OPR systems. The results show that other forms of academic bias, such as discrimination based on academic department or class difficulty are also common.

In the second phase of our analysis, we proposed several design strategies to mitigate issues of bias in OPR systems. These strategies can be divided into corrective and preventive strategies. Preventive strategies include design tactics that can be used to prevent biased behavior online, such as controlled anonymity, the use of incentives and reward mechanisms, integrating adaptive evaluation systems, and the selective reveal of rating information. Corrective design strategies include design tactics that can be applied after-the-fact to remove biased content that has made its way to the system. These tactics include bias correction in ratings, filtering review anomalies, and using adjusted rating systems that take into account the timeline of ratings as well as the reputation of the rater.

The work presented in this paper will be extended through engineering studies which will be conducted to examine the effect of our proposed solutions on the rating behavior of students and the reliability and authenticity of OPR systems.
References


Vita

Haley Tatum, born and raised in Louisiana, received her Bachelors of Science in Computer Science from Louisiana State University in 2020. As her interests in Computer Science grew, she decided to continue her education in the Division of Computer Science and Engineering at Louisiana State University. She will receive her Master of Science in Computer Science in May 2022 and plans to work as a Data Scientist upon graduation.