The Future of Engineering Education
2024 Annual Conference & Exposition
Oregon Convention Center
Portland, OR . June 23 - 26, 2024
ASEE
Paper ID #44171

Evaluation of the Effect of Anonymous Grading on Student Performance on High-Stakes Assessments

Dr. Neha B. Raikar, University of Maryland, Baltimore County

Dr. Raikar is a Senior Lecturer at the University of Maryland, Baltimore County, in the Chemical, Biochemical, and Environmental Engineering department. She has taught both undergraduate and graduate-level courses. Dr. Raikar also has 3 years of industry experience from working at Unilever Research in the Netherlands.

Dr. Nilanjan Banerjee

Nilanjan Banerjee is an Associate Professor at University of Maryland, Baltimore County. He is an expert in mobile and sensor systems with focus on designing end-to-end cyber-physical systems with applications to physical rehabilitation, physiological mon

Evaluation of the effect of anonymous grading on student performance on high-stakes assessments

Abstract

This study investigates the impact of anonymous grading on student performance in assessments within engineering courses. Traditional grading methods, often influenced by implicit biases, can negatively affect student outcomes and increase anxiety, thus undermining fairness. This paper aims to decouple student identity from their work by implementing an anonymous grading system using barcodes, potentially reducing bias and enhancing academic integrity. The system was piloted in undergraduate chemical engineering courses, providing initial evidence of its viability. Through a comprehensive analysis comparing student outcomes under traditional and anonymous grading methods, the study seeks to empirically validate the effectiveness of anonymous grading in improving student performance and psychological well-being, contributing to the development of more equitable educational practices.

Introduction

Academic evaluation has traditionally been dominated by exams and quizzes. While widely used, these conventional approaches have come under scrutiny for their potential to perpetuate implicit biases. Among these, the halo and horn effects [1][2] stand out, where an instructor's overall impression of a student can skew the grading of individual pieces of work, either favorably or unfavorably. This phenomenon is not merely an academic concern; it has tangible impacts on student outcomes, contributing to significant grade discrepancies that can alter the trajectory of a student's academic and professional future.

Moreover, the psychological impact of these traditional assessment methods on students cannot be overstated. The anxiety associated with how an instructor perceives their work can affect students' performance, often irrespective of their actual academic capabilities or understanding of the subject matter. This anxiety is not just a by-product of the high stakes involved but is linked to the fear of subjective bias in grading practices. Such a climate of fear and uncertainty can stifle learning, discourage risk-taking in intellectual pursuits, and ultimately undermine the educational process.

Recognizing these challenges, our study proposes anonymous grading as a countermeasure to mitigate the effects of implicit bias and alleviate student anxiety associated with academic assessments. This proposition is rooted in the hypothesis that anonymizing submissions can effectively decouple student identity from the work being assessed, thus minimizing the influence of preconceived notions or biases on grading decisions [3].

This research builds upon our preliminary findings published in a work-in-progress paper [4], where we explored the feasibility and initial impacts of implementing anonymous grading in academic settings. The paper led to the development of a tool that leverages barcode technology to maintain the anonymity of student submissions throughout the grading process. This tool was piloted in two undergraduate-level chemical engineering courses, offering a real-world context for our investigation and providing initial evidence supporting the viability of anonymous grading.

Our study's foundation is built upon this innovative approach to grading, aiming to expand the scope of our investigation to thoroughly assess the impact of anonymous grading on both student performance and psychological well-being. Through a comprehensive analysis of student outcomes in courses utilizing this system, compared to those adhering to traditional grading methods, we seek to provide empirical evidence supporting our hypothesis. This evidence is crucial not only for validating the effectiveness of anonymous grading but also for informing future educational policies and practices, potentially leading to the widespread adoption of such systems in academic institutions worldwide.

In sum, the intersection of traditional grading practices, implicit bias, and student anxiety presents a complex challenge to the integrity and fairness of academic assessments. Our study aims to address these issues head-on, offering a novel solution by implementing anonymous grading. By exploring this approach, we hope to contribute to creating a more just and equitable educational environment where student success is determined by ability and effort rather than perceptions and biases.

Related Works

From our analysis of the anonymous grading literature, the focus is on its application in peer assessments [5],[6] and the evaluation of student papers [7], with a body of work exploring its effectiveness across various disciplines, including medicine [8],[9]. However, there is a noticeable gap in research specifically addressing the utility of anonymous grading for in-person examinations or quizzes within the engineering field, which is the focus of this work. Even though learning management systems like Blackboard are frequently used, the support for anonymous grading is limited to online or electronically submitted exams.

Tools such as the Akindi bubble sheet system provide a mechanism for anonymous grading for multiple-choice assessments, demonstrating the feasibility of anonymizing certain types of assessments [10]. Additionally, auto-graders have been developed for programming assignments, offering anonymity but requiring strict adherence to submission guidelines to function correctly. Despite these advancements, the literature indicates a lack in tools specifically designed for anonymizing grading in traditional in-class paper exams and quizzes, which remain a staple of academic evaluation on many campuses. This absence suggests a broader challenge in extending the principles of anonymous grading to all facets of academic assessment, particularly in environments where traditional examination formats prevail. As such, developing innovative tools or methodologies that bridge this gap could significantly impact educational fairness and objectivity, especially in disciplines like engineering, where in-person assessments play a crucial role in student evaluation.

Workflow of the Implementation

There are 3 distinct elements to the implementation of our system and the development of the tool to administer anonymous in-person engineering exams, illustrated in Figure 1.

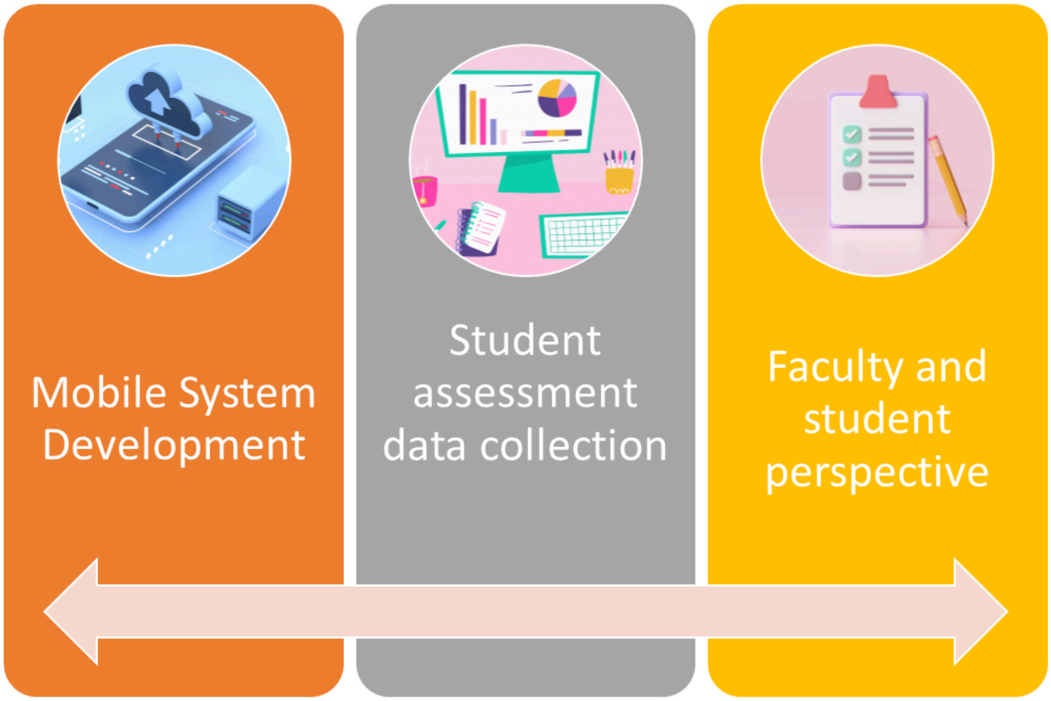
Mobile System Development
Student assessment data collection
Faculty and student perspective

Figure 1: Workflow of the Implementation

Objective 1: Development and Implementation of an Anonymous Grading Tool

The primary goal of this work is to devise an accessible tool to administer anonymous exams seamlessly. By integrating barcode technology, we successfully dissociate student identities from their submissions, ensuring anonymity in grading. This tool's deployment in real-world educational settings provided valuable insights into the practical aspects of implementing anonymous grading systems. The web application development approach is described below.

1. Class Roster Input
    - The application allows instructors to upload or input the class roster, including student names and any other relevant information.
2. Alphanumeric Code Generation.

- Upon inputting the class roster, the system automatically generates a unique alphanumeric code for each student. This code serves as the identifier for the student in the grading process, ensuring anonymity.

3. Barcode Generation.
   - Along with the alphanumeric code, a corresponding barcode is generated for each student. This barcode represents the student's unique code in a scannable format.
4. Database Storage.
   - The mappings between student names, alphanumeric codes, and barcodes are securely stored in a backend Amazon Web Service (AWS) DynamoDB database. Access to this mapping is restricted to authorized users only, maintaining the integrity of the anonymous grading process.
5. Printing of Barcodes and Name pages.
   - The web application provides functionality for printing two pages. The first page has the student's name and any information the instructor wishes to provide. The second page has just barcodes (no names) and exam-related information. These two pages are added to the remaining exam pages.
6. Separation of Identity and Responses.
   - The first page with the student's name is used to hand out the right exam to the corresponding student during the exam. The page with the student's name is collected after the exams are distributed, leaving only the barcode page attached for grading purposes.

Objective 2: Comprehensive Impact Analysis of Anonymous Grading

Our methodology encompasses a detailed examination of student performance across multiple exams, comparing results from traditional grading methods against those utilizing our anonymous grading tool. This approach involves random assignment of students to control and test groups, with periodic rotations to expose all participants to both grading scenarios. Such a design enriched our dataset, facilitating a more nuanced analysis of the effects of anonymous grading.

Table 1: Control/Test methodology used for various exams in Class A. The group shaded green will receive the test anonymously, and the other group non-anonymously.

Exam 1	***Exam 2***	***Exam 3***	***Final Exam***
Group A	Group B	Group A	Group B
Group B	Group A	Group B	Group A

This approach helps examine the exam's effect on student performance so that we can compare the control and test groups for the same exam.

Evaluation

We evaluated two mid-sized undergraduate classes for this paper. The demographics for the two classes are shown below. We focus on the first 3 ethnicities and Class A for the remainder of the analyses. These ethnicities had a larger number of samples than the others. Figure 2 shows that the demographics are similar in the two classes considered.

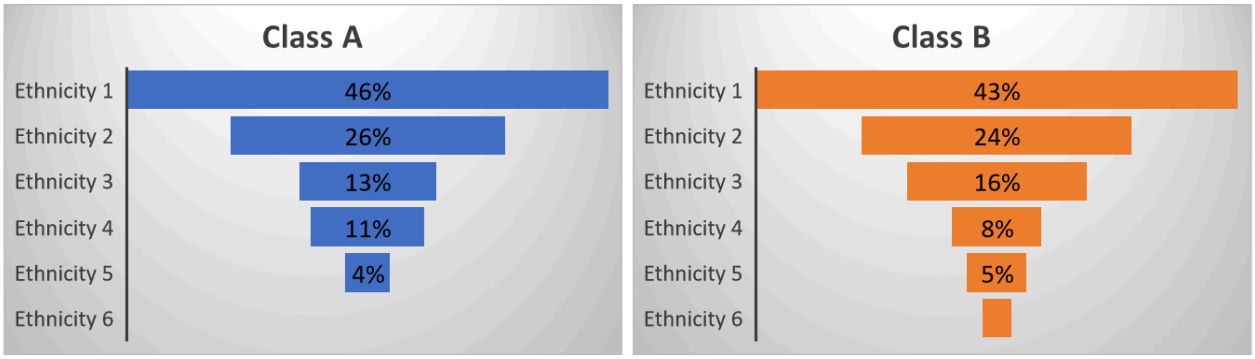

Figure 2: Demographics of the two courses being assessed in this work.

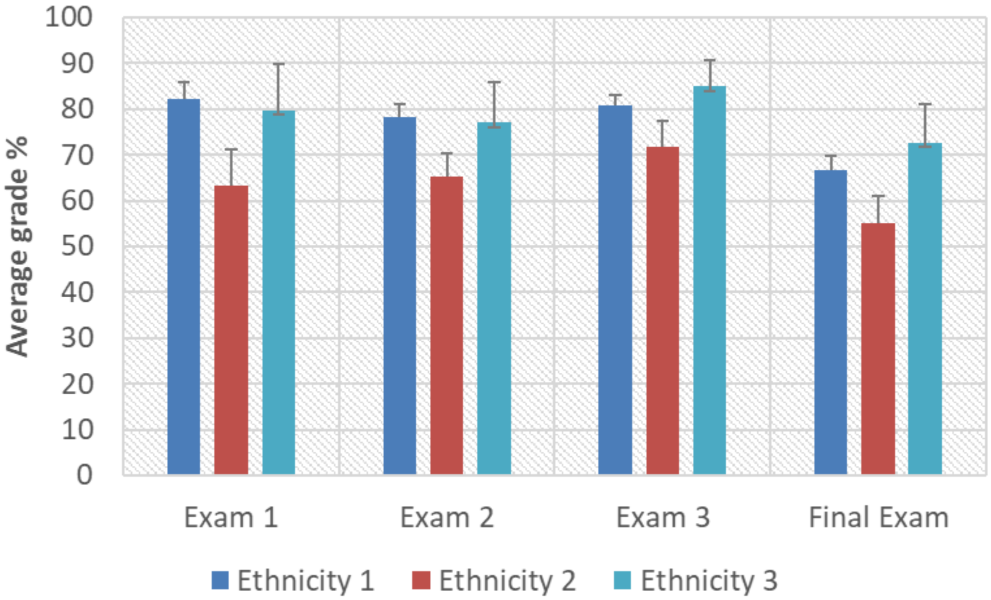

Figure 3: The average grade by ethnicity for the 4 exams considered for Class A. The error bars represent the standard error.

From Figure 3, it is evident that there is a difference in the average grade earned by the three ethnicities shown. The standard deviation and, consequently, the standard error was higher for ethnicities 2 and 3. This indicates that there is a difference in the performance of these two ethnicities. In addition, ethnicities 1 and 3 consistently outperformed ethnicity 2. Another important observation is that the final exam had a lower exam average than the 3 midterm examinations.

With the above baseline, we performed testing to evaluate the efficacy of anonymous grading. The students in the class were divided into two groups; Group A was graded anonymously on the first exam and then was graded without anonymous grading on the second exam. The pattern alternated for the two groups, as illustrated in Table 1. Figure 4 shows the average grade for group A. The blue bars represent anonymous exams, while the red bars indicate non-anonymous exams. As noted earlier, the final exam had a lower average score, which is reflected across the 3 ethnicities shown. Figure 4 also shows that anonymizing the exam leads to performance improvement for Ethnicity 2. Ethnicities 1 and 3 showed no difference.

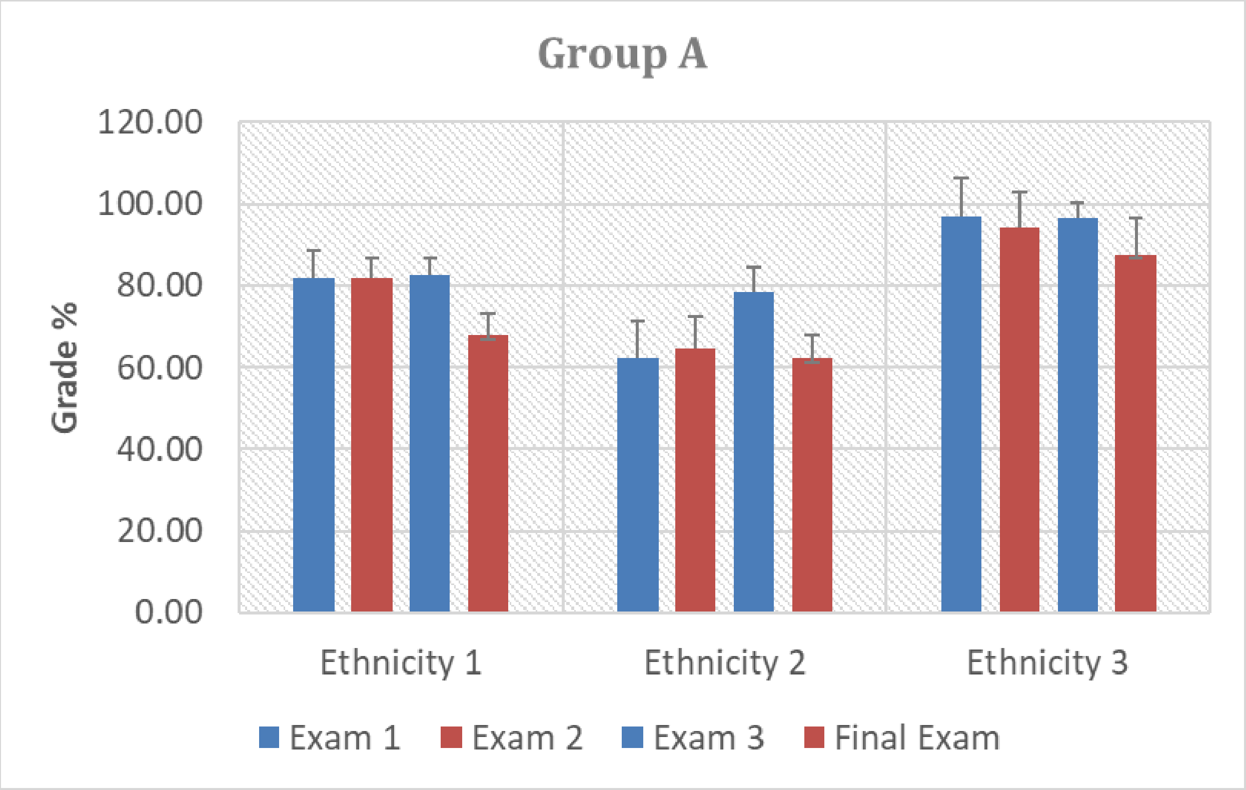

Figure 4: The average grade by ethnicity for the 4 exams considered for Group A in Class A. The error bars represent the standard error. Group A started with anonymous exams and then switched.

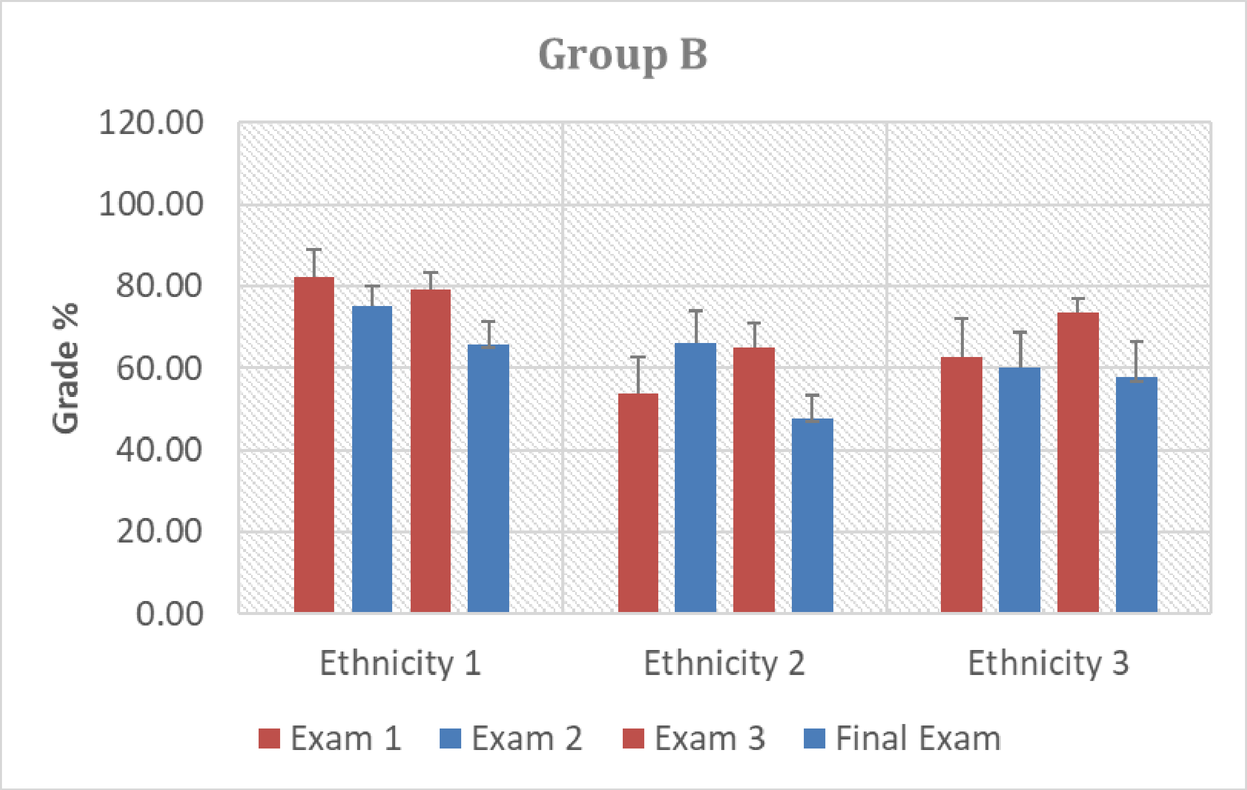

Figure 5: The average grade by ethnicity for the 4 exams considered for Group B in Class A. The error bars represent the standard error. Group B started with non-anonymous exams and then switched.

Figure 5 shows the average grade for group B, which is the group that started with non-anonymous exams. Once again, blue bars represent anonymous exams, while the red bars indicate non-anonymous exams. Ethnicity 2 also seems to show performance improvement due to anonymous grading in the case of group B. On the contrary, ethnicities 1 and 3 might show a drop in performance.

Ethnicity 3 had a higher average for group A than group B. Since the two groups were randomly created, it appears that higher performing students ended in group A for ethnicity 3. This is highlighted in Figure 6 which shows the score difference between Group A and Group B. Group A is seen to consistently outperform Group B for all ethnicities. This is from an unintentional sampling bias, which was difficult to determine a priori, and also from a small sample size.

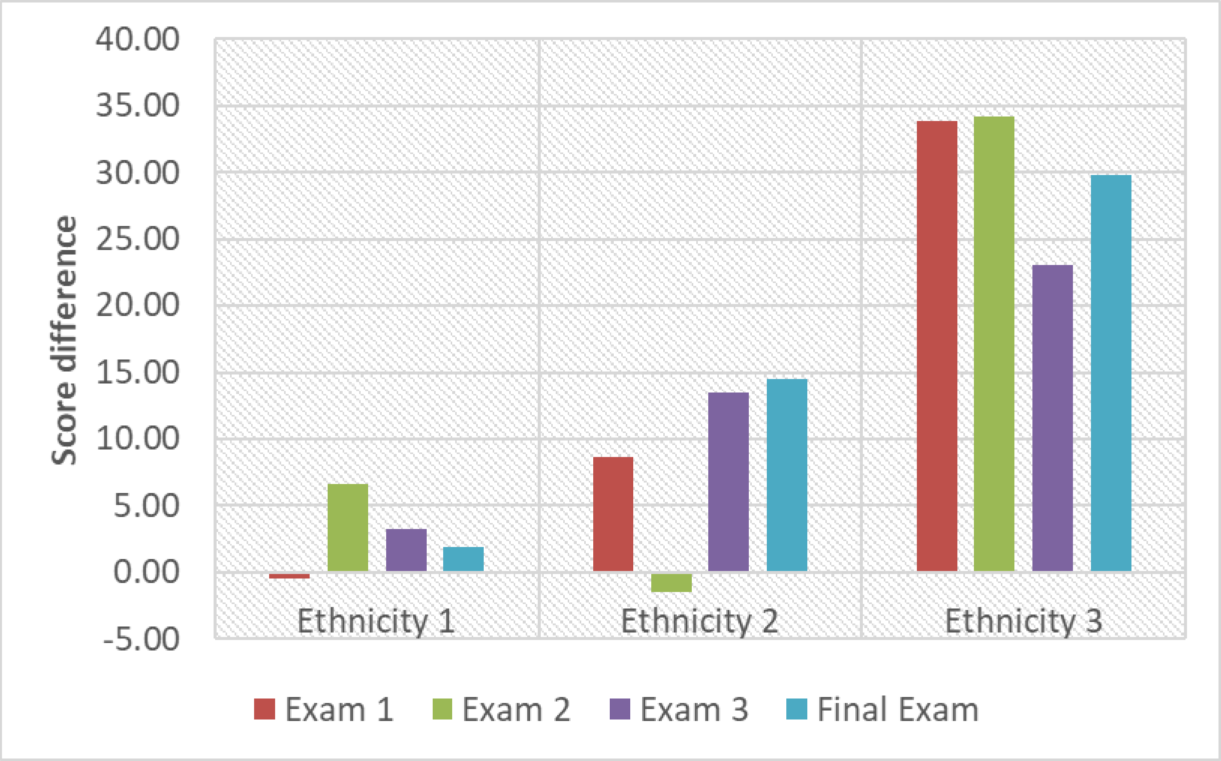

Figure 6: Group A and B score differences for the 3 ethnicities. Positive differences indicate a higher score for Group A compared to Group B.

Another consideration is to evaluate performance by gender. For this work, based on gender reporting, we only consider two classes, Gender 1 and Gender 2. Figure 7 shows that Gender 2 always outperforms Gender 1. Like with the ethnicities, it appears that Group A shows higher performance than Group B for Gender 1. For Gender 2, the performance of group A is higher or at par with group B. This leads us to believe that anonymous grading did not improve gender-based performance. However, this could also be due to the sampling bias.

The analysis from the second class showed similar trends; hence, we do not elaborate on the results from the second class here.

Summary: We draw two conclusions from our evaluation. First, we observe that anonymous grading can lead to better grades for certain ethnicities (Ethnicity 2). Secondly, we observe that anonymous grading does not lead to better grades when considering gender.

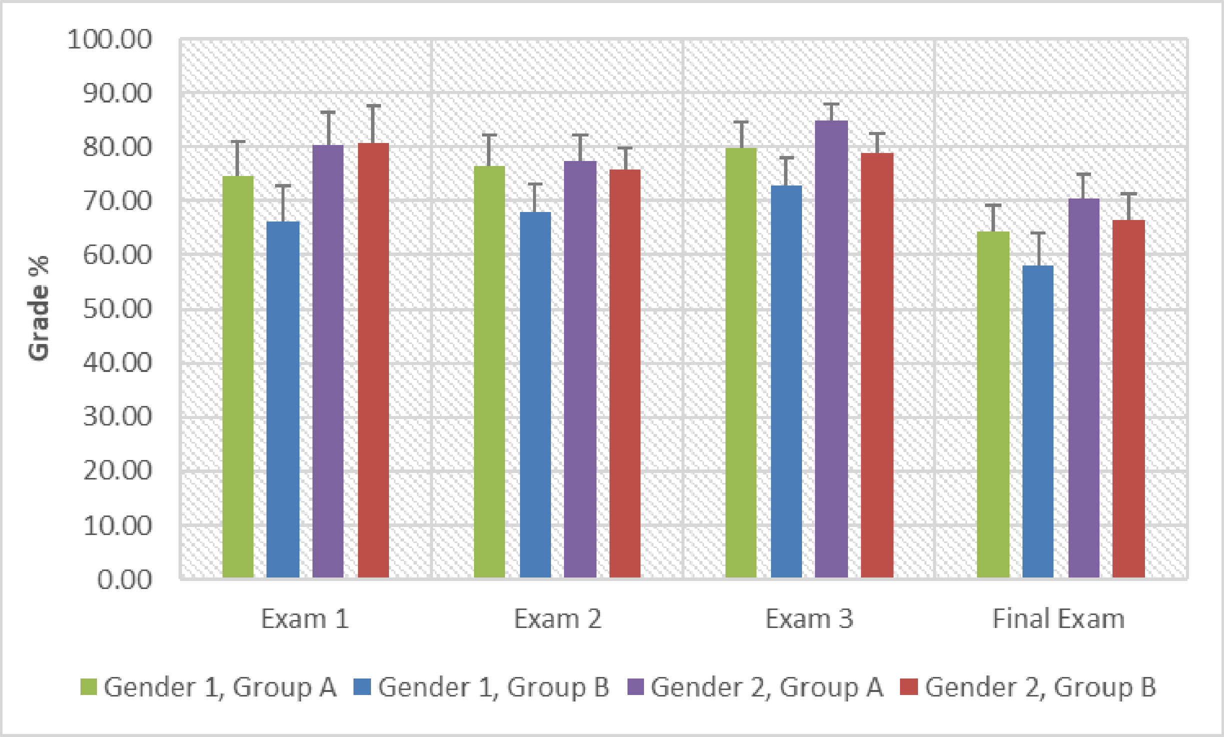

Figure 7: Grade differences between two genders considered for the two groups.

Potential Pitfalls

We have found three main limitations of our study.

1) Writing style: The assessments we are considering are handwritten in-person submissions. Some identifiers, like handwriting or cursive font, can reveal the student's identity. This will be more pronounced if multiple exams are taken by the student or if the student has the same instructor in multiple classes.
2) Sample size: The classes we considered had around 40 students. After splitting the class into control and test groups, the number of students in some demographics was low.
3) Nature of the course taught: We employed this technique for courses in the chemical engineering major. A wider outreach with more courses in other majors will provide a comprehensive look at the problem.

Future Considerations

We are pursuing several avenues of future research in this project. We outline some of our future work below.

Improvement of the Anonymous Grading tool

- Security and Privacy: Ensuring the security of student data and the integrity of the anonymous grading process is paramount. The application should implement robust security measures, including encrypted database storage and secure access controls.
- Ease of Use: The interface should be user-friendly, allowing instructors to easily upload class rosters, generate codes and barcodes, and print stickers without extensive technical knowledge.
- Integration with Existing Systems: Ideally, the application should be capable of integrating with existing Learning Management Systems (LMS) to streamline the process of managing class rosters and grades.
- Scalability: The system should be scalable and capable of handling classes of varying sizes, from small seminars to large lecture courses.
- Flexibility: While designed for in-person exams, the system should be adaptable to various assessment formats, including quizzes, midterms, and finals, across different disciplines.

Evaluation of Tool Efficacy and Acceptance

A critical aspect of our research focuses on evaluating the tool's effectiveness in reducing bias and its usability. However, we have not completed this part of the assessment. To achieve this goal, surveys will be crafted to collect both qualitative and quantitative data from students, assessing their perceptions of the anonymous grading process, their interaction with the tool, and its perceived impact on their performance and anxiety levels. We would also like to enlist other instructors to use our tool and provide feedback on the usability and efficacy of the platform.

Conclusions

The findings from this study underscore the potential of anonymous grading to mitigate implicit biases in academic evaluations, particularly in high-stakes engineering assessments. For certain demographics, anonymous grading has been shown to improve grades, illustrating its capacity to foster a more equitable evaluation process. However, the impact of anonymous grading on performance based on gender was minimal. While the study confirms the viability of anonymous grading in enhancing fairness, it also highlights the need for further research to refine the system and fully understand its implications across diverse educational contexts. Pursuing such innovations in grading practices promises to advance educational equity and ensure student success more accurately reflects ability and effort.

References

1. T. M. Addy et al., *What Inclusive Instructors Do: Principles and Practices for Excellence in College Teaching*. Stylus Publishing, LLC, 2021.
2. J. M. Malouff, A. J. Emmerton, and N. S. Schutte, "The risk of a halo bias as a reason to keep students anonymous during grading," *Teaching of Psychology*, vol. 40, no. 3, pp. 233-237, 2013.
3. L. R. Southgate, "Rethinking Anonymous Grading," *Ethic Theory Moral Prac*, 2023. [Online]. Available: https://doi-org.proxy-bc.researchport.umd.edu/10.1007/s10677-023-10415-y
4. N. B. Raikar and N. Banerjee, "Using anonymous grading for high-stakes assessments to reduce performance discrepancies across student demographics," in *2023 ASEE Annual Conference & Exposition*, Baltimore, Maryland, 2023, doi: 10.18260/1-2--42859.
5. A. Satyanarayana, R. Lansiquot, and C. Rosalia, "Using Prescriptive Data Analytics to Reduce Grading Bias and Foster Student Success," in *2019 IEEE Frontiers in Education Conference (FIE)*, IEEE, 2019.
6. J. K. Dorsey and J. A. Colliver, "Effect of anonymous test grading on passing rates as related to gender and race," *Academic Medicine*, 1995.
7. M. Gusev, M. Kostoska, and S. Ristov, "A new e-Testing platform with grading strategy on essays," in *2017 IEEE Global Engineering Education Conference (EDUCON)*, IEEE, 2017.
8. M. Kobayashi, "Does anonymity matter? Examining quality of online peer assessment and students' attitudes," *Australasian Journal of Educational Technology*, vol. 36, no. 1, pp. 98-110, 2020.
9. E. Panadero and M. Alqassab, "An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation, and peer grading," *Assessment & Evaluation in Higher Education*, 2019.
10. Akindi, [Online]. Available: https://akindi.com/