

On the Effect of Question Ordering on Performance and Confidence in Computer Science Examinations

Brian Harrington, Jingyiran Li, Mohamed Moustafa, Nick Cheng, Marzieh Ahmadzadeh

Department of Computer and Mathematical Sciences, University of Toronto Scarborough



UNIVERSITY OF
TORONTO
SCARBOROUGH

Abstract

Most computer science examinations tend to start with the easiest questions and progress towards the more difficult material. Whether this is because of the highly scaffolded nature of the course, an attempt to ‘ease students in’, or simply by convention, is unclear. However, there is a great deal of data from the psychology literature to suggest that human perception of the difficulty or discomfort of a task is disproportionately affected by the last part of the task completed. Therefore, is it possible that by structuring our exams in an easy-to-hard fashion, we are causing students to perceive the test as more difficult than it actually is? Could changing the question order allow us to change students’ perception of their own achievement? What effect could this have on actual performance? This paper attempts to answer these questions by randomly assigning students to write exams ordered either easy-to-hard (referred as ‘Easy-Difficult’) or hard-to-easy (‘Difficult-Easy’), then ask them to predict their marks on per-question basis. We find that the question ordering has a small but not statistically significant effect on the performance, and virtually no effect on predicted marks when treating the entire class as one unstratified sample. However, the effect was significant for certain subgroups created via stratification. In particular, swapping the order of the questions appeared to hurt the performance of international students, and to raise both the performance and confidence of female students.

Main Objectives

1. Replicate and then extend previous study[1] which uses students estimation of their own marks on an introductory computer science final examination as a proxy for confidence and difficulty
2. Evaluate the effect of question ordering on students performance and confidence

Methodology

1. Add bonus question to final exam:

Predict your mark on each question of this examination. If your estimate is within 10% of your actual grade, you get this bonus mark.

2. Rank all exam questions in order of difficulty.
3. Create two versions of exam (Easy-Difficult, and Difficult-Easy) with same questions but opposite ordering.
4. Students randomly received one of two exams.

Results

- Strong correlation between exam performance, confidence and predictive ability ($\rho = -0.292, p < 0.001$ and $\rho = -0.293, p < 0.001$).
- Student perception is linked more strongly to previous mark history than to actual exam performance (Spearman $\rho = 0.702, p < 0.001$).
- No evidence of any impact of re-ordering the question for class overall (t test $p = 0.7174$).
- Difficult-Easy ordering had negative impact on performance of international students (not statistically significant)
- Difficult-Easy ordering had positive impact on female students, both in performance and confidence
- ANOVA results suggested that the interaction effect between gender and question ordering is statistically significant for over prediction ($p = 0.0135$) and prediction error magnitude ($p = 0.00415$), respectively.

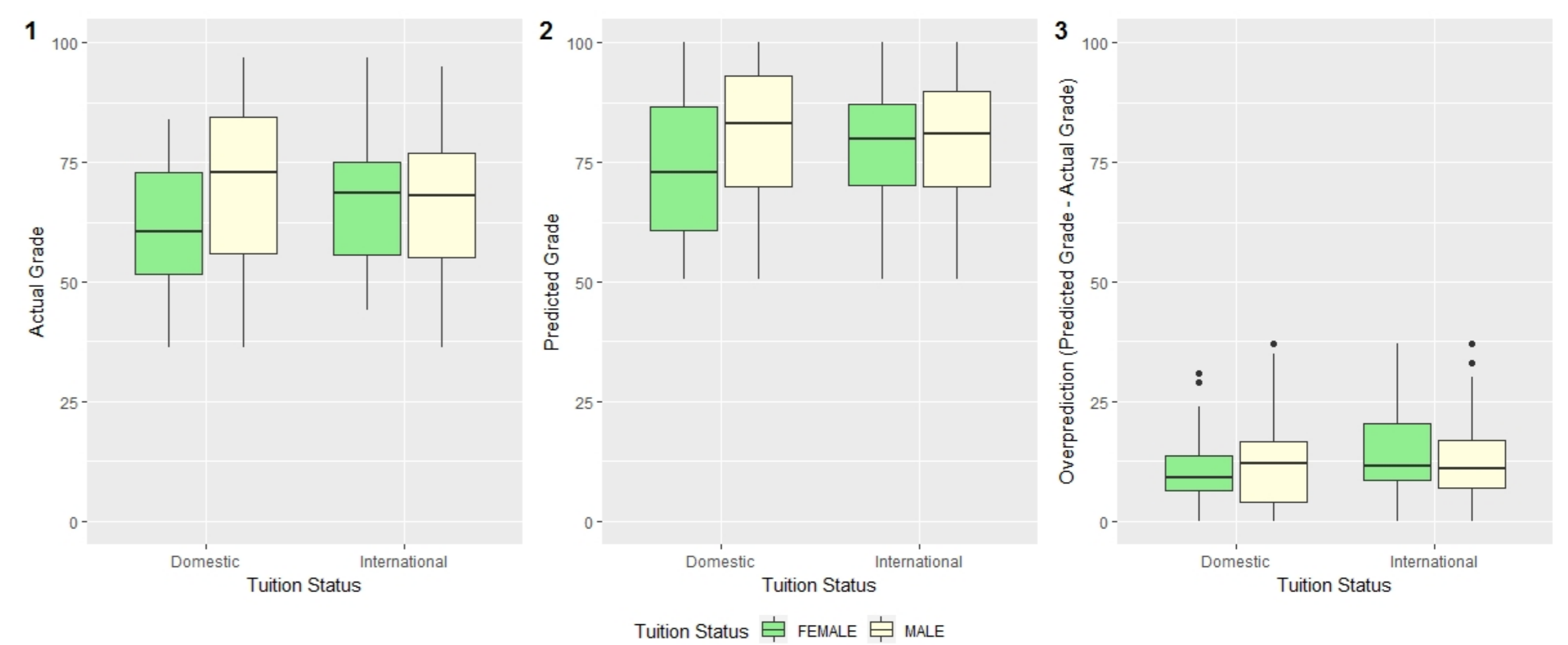


Figure 1: Exam Mark by Tuition Status and Gender

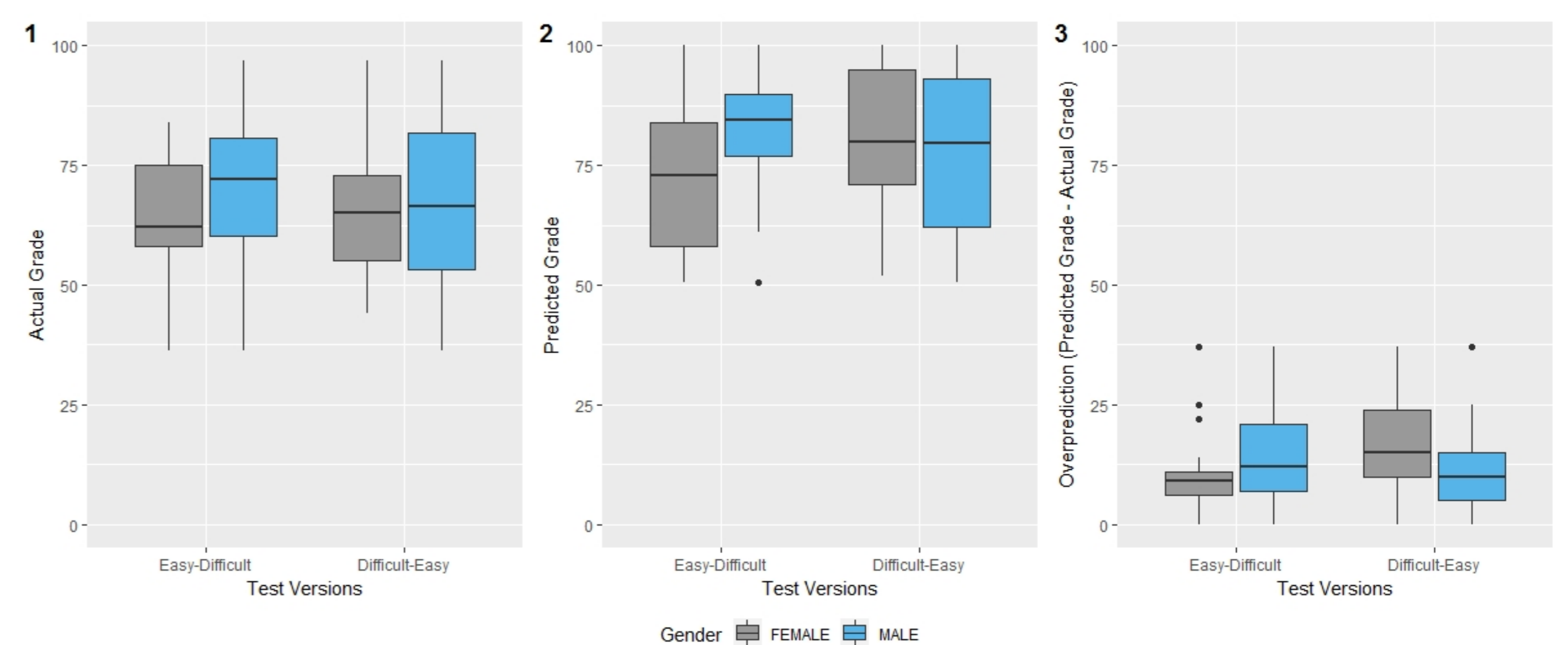


Figure 2: Exam Mark by Question Ordering and Gender

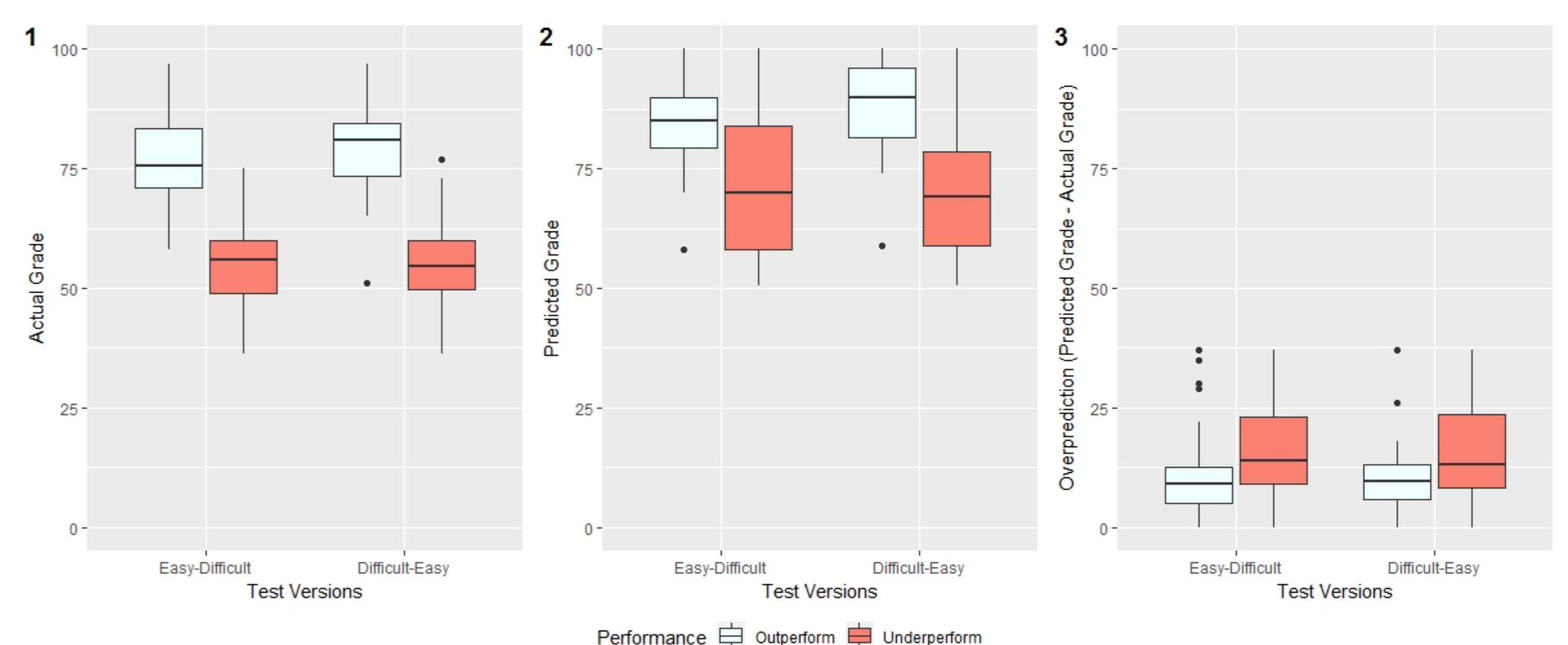


Figure 3: Exam Mark by Question Ordering and Term Test Performance

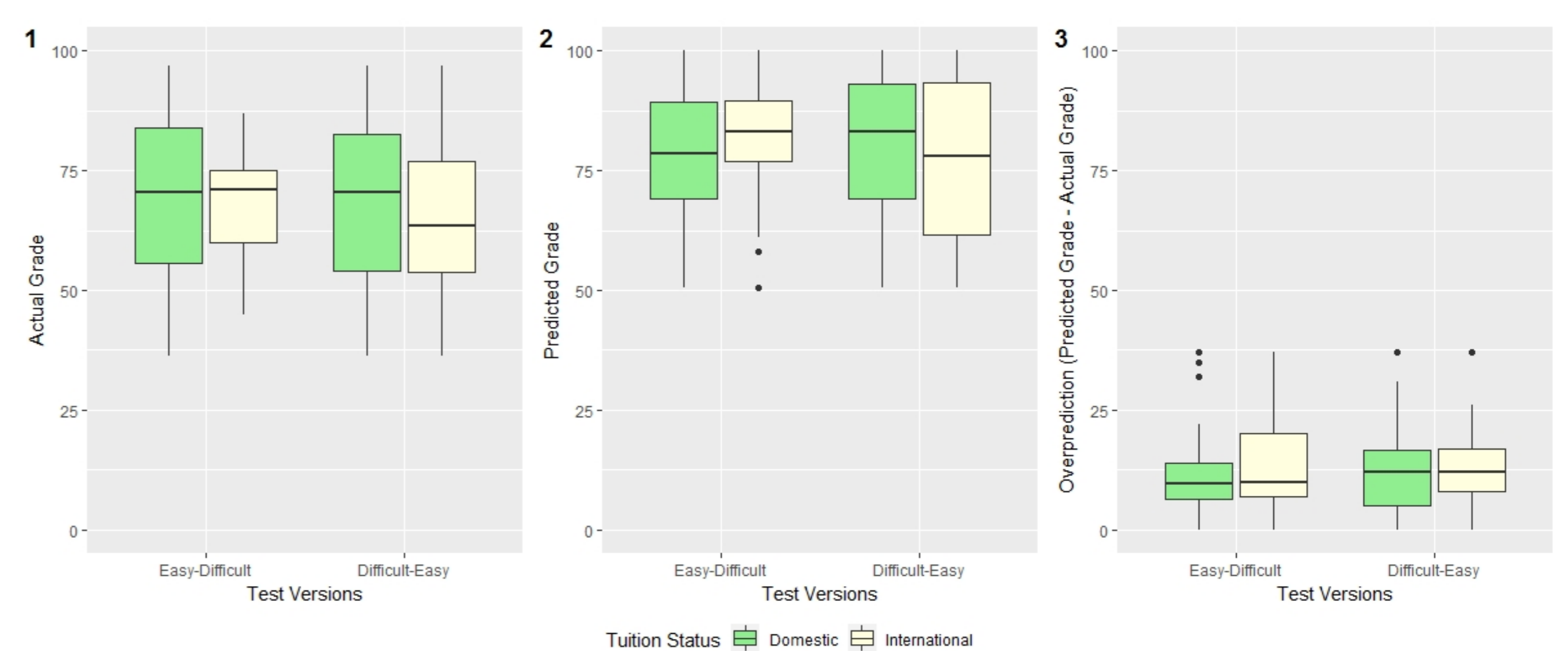


Figure 4: Exam Mark by Question Ordering and Tuition Status

Conclusions

- We were able to reproduce all of the findings in the areas replicated, aside from the relative overprediction of marks by male students.
- Dividing students by tuition status did not have an impact on its own, but the data shows that the gender confidence gap previously found does not hold equally across domestic and international students.
- Our initial assumption that we would see an impact of the peakend rule when re-ordering exam questions proved to be false for the student population as a whole. However, this was because the effect was only present in females, and actually reversed in males.

References

- [1] Brian Harrington, Shichong Peng, Xiaomeng Jin, and Minhaz Khan. Gender, confidence, and mark prediction in cs examinations. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, pages 230–235. ACM, 2018.