# The Impact of External Factors on Multiple Choice Exams

Robert Dodge, Industrial Engineering
Mentor: Dr. Michael Clough
School of Computing, Informatics, and Decision Systems Engineering

## Background

In a perfect world, exams would be unbiased and hold each student to the same academic standard. However, there are many factors outside of the questions asked that may impact how a student performs on any given exam. The purpose of this project is to examine if differences in construction between exams, without altering the questions asked, will produce statistically significant differences in exam scores.

## Methodology

This project examined the effect of two factors commonly present in exam construction.

1. "Answer not here" being given as a possible answer choice.
2. The order in which questions are presented.

To collect the data required for this project, the Spring 2021 IEE 380 class was used as a sample group. Over the course of this class, data was collected from 7 quizzes administered online through the Canvas platform. These quizzes were open note, open book, and did not require a lockdown browser. The scores collected from these quizzes were compared to the scores from the previous semester.

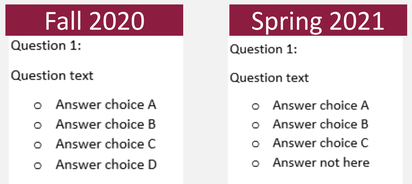


Figure 1. Comparison of questions given to the Fall 2020 class compared to question given to Spring 2021 class

## Results

| Quiz Scores | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Quiz 1 | Quiz 2 | Quiz 3 | Quiz 4 | Quiz 5 | Quiz 6 | Quiz 7 |
| Factor Present | None | None | Order | Order | "Answer not here" | Both | "Answer not here" |
| Average Fall 2020 Scores (Control) | 8.01 | 7.86 | 7.87 | 8.13 | 7.14 | 8.18 | 6.61 |
| Average Spring 2021 Scores | 8.10 | 7.64 | 7.86 | 8.91 | 7.50 | 8.23 | 6.34 |

Figure 2. Average scores for all 7 quizzes from both Fall 2020 and Spring 2021.

### Data Analysis

F-tests

$$H_0: \sigma_1 = \sigma_2$$
$$H_1: \sigma_1 \neq \sigma_2$$

t-tests

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

| F-test Results | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Quiz 1 | Quiz 2 | Quiz 3 | Quiz 4 | Quiz 5 | Quiz 6 | Quiz 7 |
| F Statistic | 0.5997 | 0.8930 | 1.2736 | 0.8259 | 0.9674 | 1.023 | 1.070 |
| P-Value | 0.0001 | 0.3800 | 0.0610 | 0.1530 | 0.8108 | 0.8535 | 0.5942 |
| Outcome | Reject Null | Fail to Reject | Fail to Reject | Fail to Reject | Fail to Reject | Fail to Reject | Fail to Reject |

Figure 3. Results for a statistical test on the difference of variances for Fall 2020 and Spring 2021 quiz scores. All tests were performed with an α level of .05.

| T-test Results | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Quiz 1 | Quiz 2 | Quiz 3 | Quiz 4 | Quiz 5 | Quiz 6 | Quiz 7 |
| t Statistic | 0.5083 | -1.0498 | -0.0779 | 4.5697 | 1.5228 | 0.2567 | -1.0536 |
| P-Value | 0.6116 | 0.2943 | 0.9379 | 0 | 0.1284 | 0.7975 | 0.2925 |
| Outcome | Fail to Reject | Fail to Reject | Fail to Reject | Reject Null | Fail to Reject | Fail to Reject | Fail to Reject |

Figure 4. Results for a statistical test on the difference of means for Fall 2020 and Spring 2021 quiz scores. All tests were performed with an α level of .05.

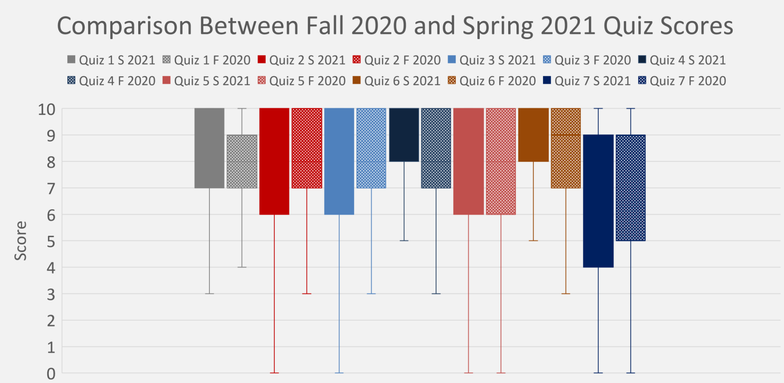


Figure 5. Boxplot showing the distributions of quiz scores for both Fall 2020 and Spring 2021 semesters. Quizzes are color coded and control quizzes are denoted with a checker pattern.

## Discussion

The results from this experiment are inconclusive. The collected data lacks evidence to suggest that there is a difference in student performance based on any of the factors that was examined. Quiz 4 did indicate a significant difference, however the lack of agreement from the other tests suggests that this may be due to random noise, not the factor in question.

There were several shortcomings with the methodology of this experiment that could have impacted the results.

1. Data was collected through online Canvas quizzes.
2. Not every question in the quizzes was multiple choice.
3. The data used as a control was from a different class.
4. Quizzes were compared on overall performance, rather than on a question-by-question basis.

Repeating this experiment in the future with a more refined methodology may yield different, more accurate results.

## Acknowledgements

I would like to thank Dr. Michael Clough for his support in both designing the experiment as well as providing the data from the previous semester's classes to be used as a control.

## References

Perlini, A. H., Lind, D. L., & Zumbo, B. D. (1998). Context effects on examinations: The effects of time, item order and item difficulty. *Canadian Psychology/Psychologie Canadienne*, *39*(4), 299-307. doi:http://dx.doi.org.ezproxy1.lib.asu.edu/10.1037/h0086821

Laffitte Jr., R. G. (1984). Effects of Item Order on Achievement Test Scores And Students' Perception of Test Difficulty. Teaching of Psychology, 11(4), 212–213. https://doi-org.ezproxy1.lib.asu.edu/10.1177/009862838401100405

FURI

Ira A. Fulton Schools of Engineering
Arizona State University