

Comparing Students' Performance in MAT112 and MAT113

Department Mathematics

1/5/2024

Contents

1	Introduction	3
2	Data Collection and Preparation	3
2.1	Design of Data Collection	3
2.2	Data Processing and Cleaning	3
2.3	Data Conversion	3
2.3.1	De-identification	4
2.3.2	Handling Sparse Categorical Variables	4
2.4	Summary Analytic Datasets	4
3	Influence of PPL and Exam	4
3.1	Unadjusted Descriptive Statistics	5
3.2	Correlation between Exam and PPL Scores	6
3.3	Regression Approach to Adjustments	6
3.4	Instructor Effect	7
3.5	Joint Influence of PPL and Instructor	8
3.6	Modeling approach with Simultaneous Adjustments	9
3.7	Summary	10
4	Comparisons with Secondary Demographic Factors	11
4.1	Effect of Different Majors	11
4.2	Potential ESL Effects	12
4.3	Grade Level	14
5	Question-specific Comparison - A Visual Inspection	14
5.1	Baseline Distributions Of Individual Questions	15
5.2	Instructor and Question-Specific Distributions	16
5.2.1	Instructor A	16
5.2.2	Instructor B	16
5.2.3	Instructor C	17
5.2.4	Instructor D	17
6	Conclusions	17

EXECUTIVE SUMMARY

This project used a semi-controlled experiment design in the data collection process to remove grading bias. Four instructors who taught **both** MAT112 **and** MAT113 in the fall semester of 2023 were invited to the study and their MAT112 and MAT113 classes were all included as the study sample. Students were recommended to MAT112 and MAT113 based on their PPL scores. However, students' enrollment into individual sections within MAT112 and MAT113 were considered to be random. All MAT112 and MAT113 sections used the same common exam. Each instructor was assigned a set of questions and graded them across all sections of MAT112 and MAT113. The Students' performance in these two courses was measured by the scores of the common exam.

In general, exam scores are influenced by various factors such as student preparedness (measured by PPL scores), instructors' strategies in teaching, and others. In this analysis, both visual and inferential methods were used to extract the true difference between mean exam scores in MAT112 and MAT113 with related effects caused due to other factors being removed. This true difference terms as the **net-gain** in the academic performance due to the one extra credit hour added to MAT112.

Some of the key findings are outlined in the following.

1. There was a solid evidence that MAT112 students outperformed MAT113 students with a **net-gain** of nearly 10.5 points (out of 100).
2. MAT112 students had significantly lower PPL scores than their peers in MAT113.
3. The instructor's effect influenced the exam grade significantly.
4. Among ESL students, MAT112 cohort still outperformed MAT113 cohort. However, when looking at MAT112 and MAT113 respectively, ESL students underperformed non-ESL students.
5. The degrees of influence of various factors on the exam were ordered by PLL score, Instructor effect, and ESL.
6. Other factors were also examined but not included in the regression models

Project Team (alphabetical order by last name):

Project Planning and Design: Michael Fisher, Allison Kolpas, Cheng Peng, Susan Turner.

Data Collection and Support: Michael Clemons, Michael Fisher, Carmen Laurelli, Sally Malarney, Andre Perkoski, Susan Turner.

Data Analysis and Reporting: Cheng Peng.

1 Introduction

The learning outcomes and the core coverage of both MAT112 and MAT113 are the same. MAT112 has one extra credit hour to cover basic algebra while MAT113 only does a brief review of the basic algebra. The goal of this analysis is to assess the “*net performance gain*” (*NG, to be elaborated later*) solely due to the addition of one extra credit hour to MAT112. The performance is measured with the transformed grade in a common exam given to participants in the entire study sample.

To extract the *The net performance gain* solely due to the addition of one credit hour to MAT112, all effects that have influences on the exam grade should be removed (or adjusted). The method of linear regression modeling can help achieve the goal. In fact, *The net performance gain is the regression coefficient of courseID.*

In general, omitting influential factors leads to biased estimates of the regression coefficients including *net performance gain*. To find the accurate *net performance gain*, we proceeded with the subsequent analysis by identifying possible influential factors to include in the model to correct the bias in the *net performance gain*. We call this process *model adjustment*.

2 Data Collection and Preparation

The project team decided to invite four colleagues who taught **both MAT112 and MAT113** in the fall semester of 2023 to help with data collection. All sections of MAT112 and MAT113 taught by these four instructors were included in this study. In the study sample, 278 students were enrolled in the study. Among them, 105 were enrolled in MAT112 and 173 in MAT113.

2.1 Design of Data Collection

The measurements used to assess the performance in these two courses are the grades of the common final exam. To control unnecessary bias, the following protocols were employed in the design of the experiment:

1. A common exam with show-your-work questions and multiple-choice questions (MCQ is only worth 3.4% of total points) was given to all MAT112 and MAT113 students. The completed exams were scanned and uploaded to *Crowdmark* - an online grading software system for instructors.
2. Each instructor graded a set of questions for all 278 participating students. Itemized marks were also recorded in the grade book. This defined the first primary data set for the analysis.
3. The second primary data set consisted of students’ math placement scores and a few demographic variables such as ESL and prior math classes.
4. The third data set was created based on class rosters that contain students’ majors and current grade levels (freshman, sophomore, junior, and senior).

2.2 Data Processing and Cleaning

Some standard data cleaning and processing procedures were performed to create the final analytic data sets. The major data steps taken in the process were data merging and data conversion.

Since 24 students in the sample did *not* have a math placement score, merging all three data sources to make a single analytic data set will lead to a reduced sample. We decided to create two analytic data sets: (1) merging all three data sources to create an analytic data set with 254 records; and (2) combining class rosters and exam data (278 records) for analysis without using placement scores.

2.3 Data Conversion

Data conversion in this analysis was based on two considerations: Privacy protection and technical requirements.

2.3.1 De-identification

The sole objective of this study is to assess the *net performance gain* of the one extra credit hour added to MAT112. However, a student's performance in the common exam could be affected by different factors including the instructor's teaching and student preparedness. Students' PPL scores were recorded after they completed the placement test. A proxy instructor ID was used to de-identify the information of participating instructors.

We used two different proxy variables for instructors on different occasions to increase the level of privacy protection. Furthermore, The exam scores were converted to a new scale so that the newly converted variable carried the same information as the original grades.

2.3.2 Handling Sparse Categorical Variables

Students' intended majors at the time of placement test and prior math classes are the two sparse categorical variables. These two variables cannot be used for analysis in their original forms due to sparse categories. We redefined these two variables based on their original distributions. The detailed definitions can be found in the RMarkdown source code.

2.4 Summary Analytic Datasets

The following is a list of variables and their descriptions.

1. ID: observation ID.
2. PlaceCtr: attempts of the placement test (PPL).
3. ESL: status of ESL: 1 = ESL student, 0 = nonESL student.
4. Qxxx: specific questions.
5. MCx: multiple choice questions.
6. Units: 4 for MAT112 and 3 for MAT113.
7. InstrID: instructor's proxy ID.
8. mpeScore: PPL score. *It will be simply called **PPL**.*
9. premath: highest level math course taken before MAT112/MAT113 (including those taken in high school and college).
10. majortype: aggregated into three major categories: pre-business, psychology, and other majors. *It will be simply called **Major**.*
11. CourseID: either MAT112 or MAT113 (MAT112 is the baseline). *net gain = average exam of MAT113 - average exam of MAT112 after other factors were adjusted.*
12. SecteNum: course section number.
13. examTotal: percentiles of the total exam scores. *It will be simply called **Exam**.*
14. MCtotal: total marks earned in MCQ.

These variable names were used to denote their corresponding effects on the response exam score in the rest of this report.

3 Influence of PPL and Exam

Some comparisons between MAT112 and MAT113 using both descriptive visualizations and analytic approaches were summarized in the following subsections.

3.1 Unadjusted Descriptive Statistics

The following table reports the means of PPL and the exam scores as well as 95% confidence intervals.

Table 1: Five-number-summaries and confidence intervals of mean PPL and exam scores of MAT112 and MAT113 respectively.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	N	LCI.mu.95	UCI.mu.95
PPL112	13.0	33.0	37.0	37.2	40.2	77.0	96	35.87	38.46
PPL113	9.0	35.0	49.0	48.0	59.0	88.0	158	15.24	50.60
Exam112	7.4	50.6	68.2	63.3	81.2	96.6	105	58.96	67.62
Exam113	11.9	39.2	58.0	55.6	72.7	96.6	173	52.54	58.69

The 95% confidence intervals of the mean PPL scores of MAT 112 and MAT113 were given, respectively, by [35.87, 38.46] and [15.24, 50.60]. The overlapped confidence intervals implied that the two means were not significantly different.

The 95% confidence intervals of the mean exam scores of MAT 112 and MAT113 were given, respectively, by [58.96, 67.62] and [52.54, 58.69]. This implied that the mean exam scores of MAT112 and MAT113 cohorts were significantly different. The mean exam score of MAT112 was higher than that of MAT113.

To see more detailed *distributional* information about the distributions of PPL and exam scores in MAT112 and MAT113, we plotted the density curves of these distributions in the same graphic panel for visual comparison.

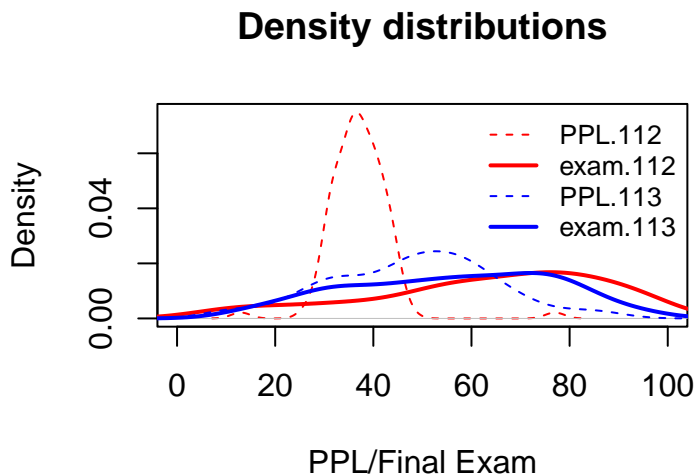


Figure 1: Distributions of PPL score and exam score between the two courses

The above figure indicates the difference between the distribution of the PPL and Exam in MAT112 and MAT113. *The farther the curve shifts to the right, the better the performance.* To be more specific

- The distributions of PLL scores (the dotted curves) of MAT112 and MAT113 students were different. MAT112 students had overall poorer PPL scores than their counterparts in MAT113.
- The performance of MAT112 students in the exam (solid curves) was significantly better than that of MAT113 students.

The 95% confidence interval of the difference between the mean PPL and exam scores of MAT112 and MAT113 students are given respectively in the following.

Table 2: 95% confidence intervals of mean PPL and exam scores between MAT112 and MAT113 students.

	LCI.95%	UCI.95%	MAT113-MAT112
PPL	7.915548	13.738461	10.82700
Exam	-12.599574	-1.085787	-6.84268

3.2 Correlation between Exam and PPL Scores

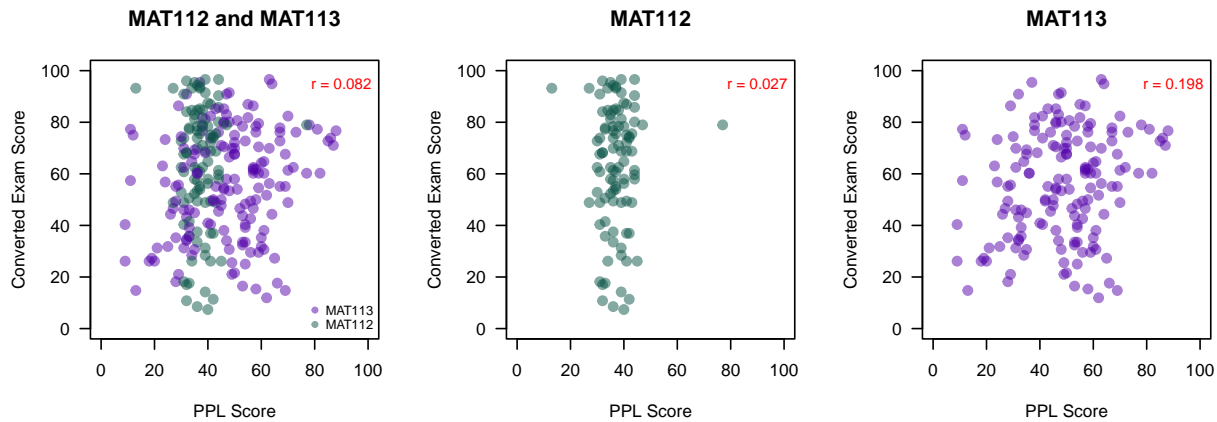


Figure 2: Correlation between PPL score and exam score.

The above figure shows a weak correlation between PPL and exam scores. The strength of the correlation of the two types of scores in the MAT112 cohort was much weaker than that of MAT113. Furthermore, the two scores were almost uncorrelated in MAT112. This observation implies that one extra credit hour played a key role in improving the performance of the common exam.

3.3 Regression Approach to Adjustments

We performed a simultaneous comparison of the performance for MAT112 and MAT113 by adjusting the baseline PPL score using the following regression model.

$$\text{Exam} = \beta_0 + \beta_1 \times \text{PPL} + \beta_2 \times \text{CourseID} + \epsilon.$$

In the above linear model, β_2 measures the “true” difference between the mean exam scores of MAT113 and MAT112, i.e., the **net performance gain** = *average exam score of MAT113 - average exam score of MAT112*, after the effect of PPL being adjusted. This **net performance gain** will be further adjusted as more influential factors are added to the subsequent models.

The above diagnostic plots do not indicate a serious violation of the model assumption. The regression results are summarized in the following table.

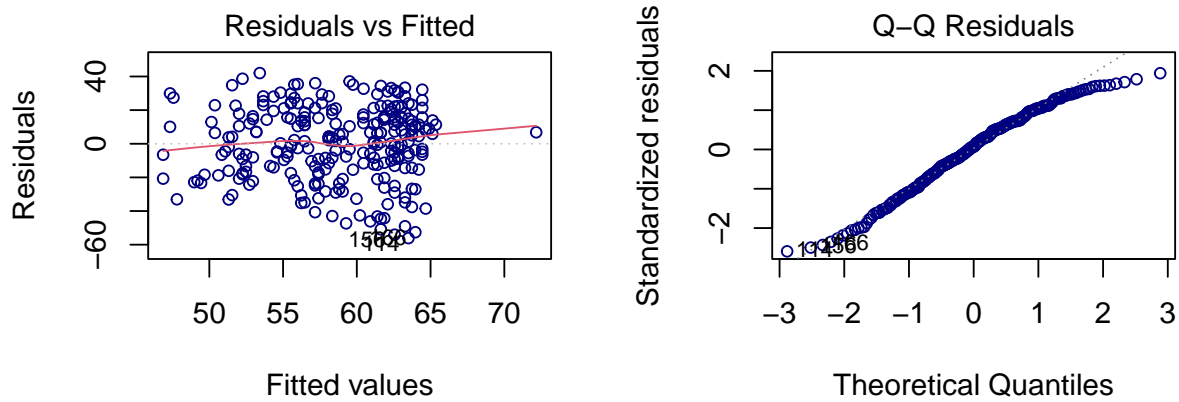


Figure 3: Diagnostic plots of linear regression model: No significant violations.

Table 3: Inferential results of regression model

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	54.1368715	4.3128814	12.552367	0.0000000
mpeScore	0.2340938	0.0994728	2.353346	0.0193773
CourseNumMAT113	-9.3772148	3.0148877	-3.110303	0.0020847

The above inferential table shows that

- The PPL scores were positively correlated to the exam scores.
- The negative coefficient -9.34 indicated that the mean exam score of MAT113 was 9.37 **lower than** that of MAT112 for any given PPL scores.

3.4 Instructor Effect

In general, individual instructor’s delivery methods influence exam scores to various degrees. We examined the distributions of the common exam grades in MAT112 and MAT113 taught by the same instructor in this subsection for potential differences to see whether to include InstrID in the model to estimate the “net performance gain” more accurately.

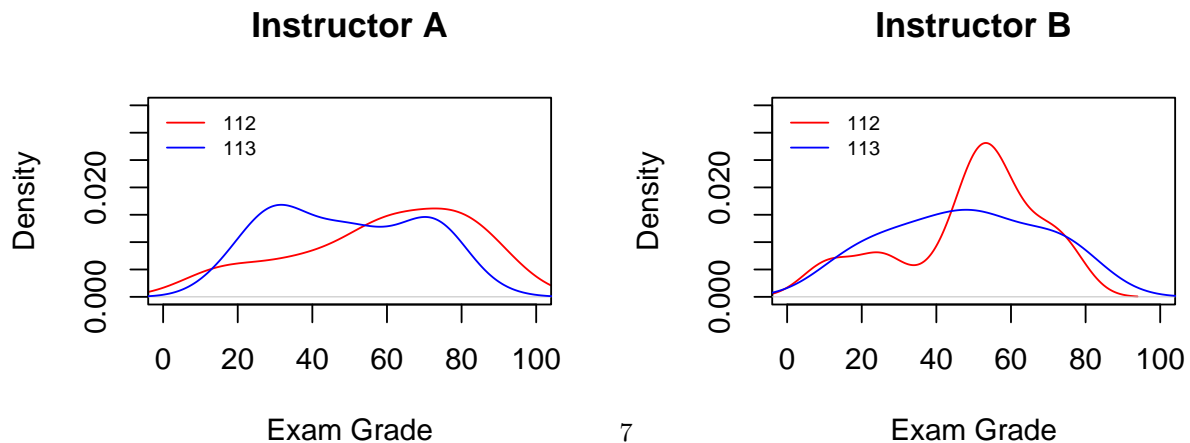


Figure 4: Difference by individual instructor: Instructors A and B

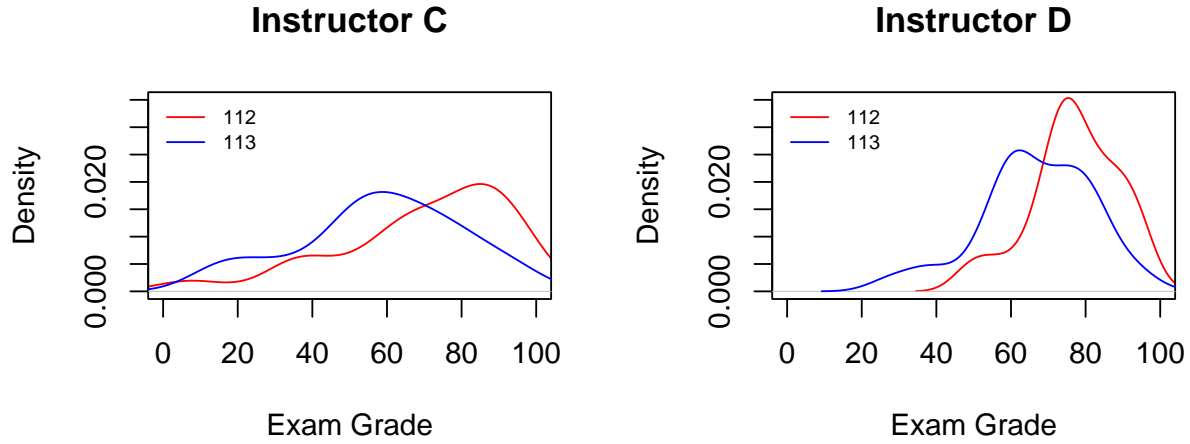


Figure 5: Difference by individual instructor: Instructors C and D

The above curves show a significant difference between the distributions of MAT112 and MAT113 for each individual instructor. To be more specific,

- The degrees of difference varied among instructors.
- Overall, MAT112 students outperformed MAT113 students.

The above implies that the effect of the instructor should be considered when comparing the performance in the exam of students in MAT112 and MAT113.

3.5 Joint Influence of PPL and Instructor

Adding corresponding PPL density curves to the above figures allows comparisons with joint adjustments of PPL and Instructor. The next few plots represent the distributions of exam grades in MAT112 and MAT113 with considerations of the influences of PPL scores and instructors.

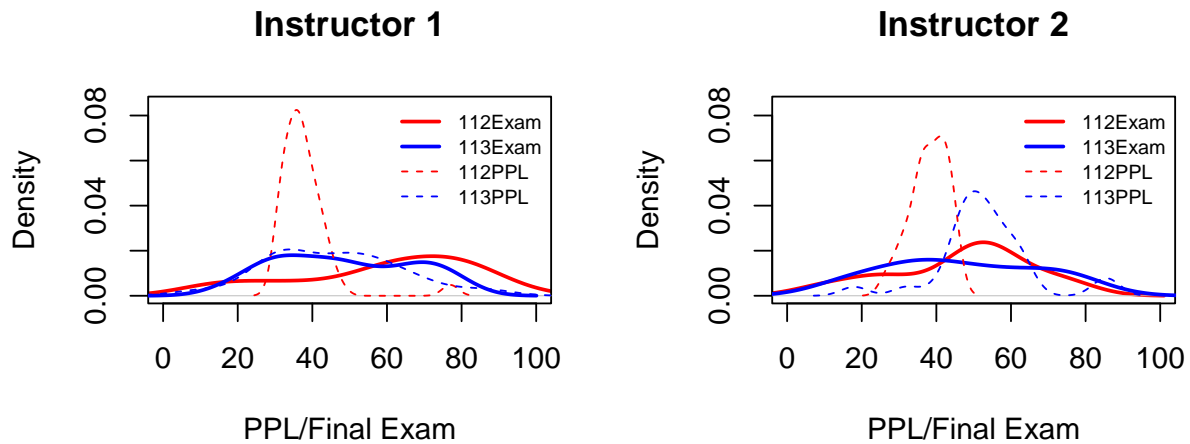


Figure 6: Difference by individual instructor: Instructors 1 and 2

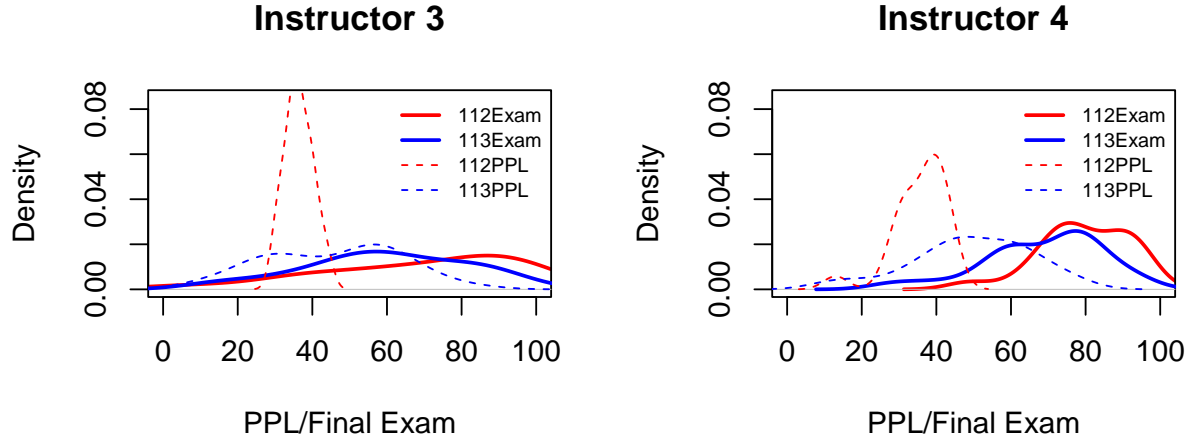


Figure 7: Difference by individual instructor: Instructors 1 and 2

The above plots provide an adjusted pair-wise comparison between PPL grade and the exam grade adjusted by the effect of the instructor. *This project does not intend to perform any comparison among instructors.*

The patterns that are common to all instructors:

- (1) MAT112 students had poor PPL scores but better exam scores.
- (2) MAT113 students had better PPL scores but poorer exam scores. The difference in shapes of the solid density curves (particularly the degrees of shift to the right) indicated instructors' influence.

Therefore, both PPL and the influence of the instructor should be adjusted simultaneously when comparing the performance of MAT112 and MAT113 students in the common exam.

3.6 Modeling approach with Simultaneous Adjustments

Since the PPL score and instructor effect jointly contribute to the performance of the exam, we use a linear regression model to help compare the performance in MAT112 and MAT113 with simultaneous adjustment of PPL and InstrID using the following regression model.

$$\text{Exam} = \beta_0 + \beta_1 \text{PPL-score} + \beta_2 \text{InstrID}_2 + \beta_3 \text{InstrID}_3 + \beta_4 \text{CourseID}_4 + \beta_5 \text{CourseID} + \epsilon$$

The above model will generate three indicator variables:

$$(\text{InstrID})_2 = \begin{cases} 1 & \text{if 2nd inst,} \\ 0 & \text{if not 2nd inst} \end{cases}, (\text{InstrID})_3 = \begin{cases} 1 & \text{if 3rd inst,} \\ 0 & \text{if not 3rd instr} \end{cases}, (\text{InstrID})_4 = \begin{cases} 1 & \text{if 4th inst,} \\ 0 & \text{if not 4th instr} \end{cases}.$$

The next residual and QQ plots do not show any serious violation of the assumption of linear regression. The results of the regression model are summarized in the following table.

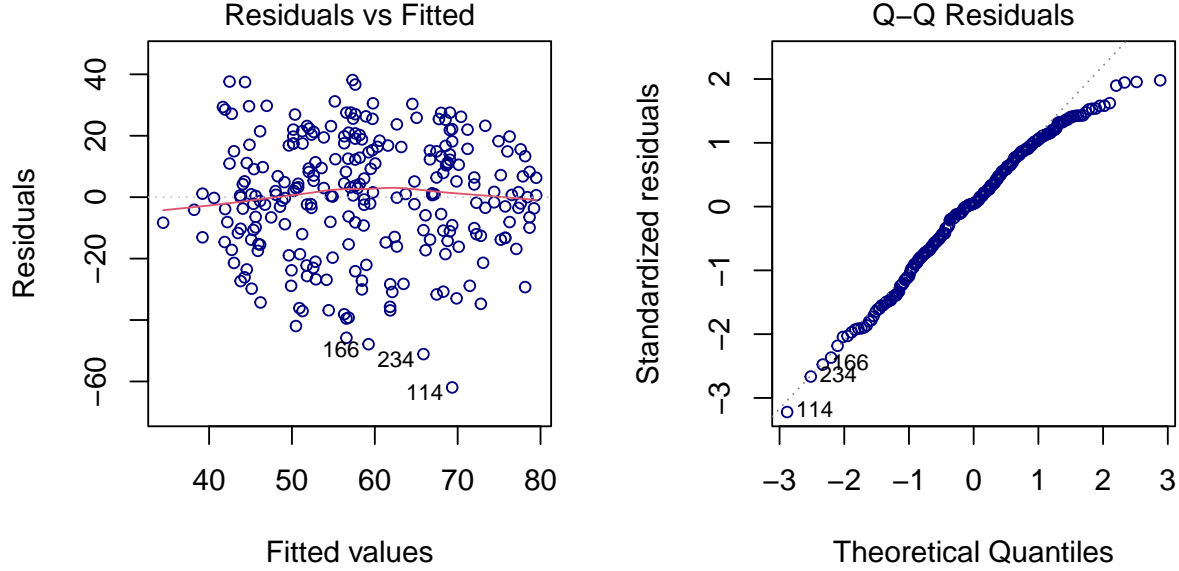


Figure 8: Diagnostic plots of linear regression model: No significant violations.

Table 4: Inferential results of regression model

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	48.0175945	4.1127284	11.675362	0.0000000
mpeScore	0.2673127	0.0899312	2.972413	0.0032455
factor(InstrID)2	-7.1601811	3.5189521	-2.034748	0.0429400
factor(InstrID)3	10.6275359	3.4302225	3.098206	0.0021712
factor(InstrID)4	19.6987262	3.1709209	6.212304	0.0000000
CourseNumMAT113	-11.2214768	2.7127709	-4.136537	0.0000483

The above shows that both PPL and Instructor effects are significant as expected. After additional adjustments with the instructor effect, the *net performance gain* increased from the previous 9.34 to 11.22.

3.7 Summary

There is solid evidence to support the hypothesis that MAT112 students outperformed MAT113 students after the potential effects of PPL and instructor being adjusted.

As observed earlier, PPL has a weak positive correlation with performance in the exam. within each instructor, one mark increment in the average PPL led to a 0.23 increase in the average exam for every instructor.

The individual instructors' contribution to the difference of averages between MAT112 and MAT113 differs vastly. The comparison between instructors is not part of the goal of this study.

4 Comparisons with Secondary Demographic Factors

Several other demographic variables are also available in the data. We performed some visual inspections to make necessary further adjustments to this section.

4.1 Effect of Different Majors

The sample had 53% of students from pre-business programs and 30% from psychology programs. We classified all participating students into three categories: pre-business, psychology, and other. The following density curves represent the distributions of PPL and exam scores by majors.

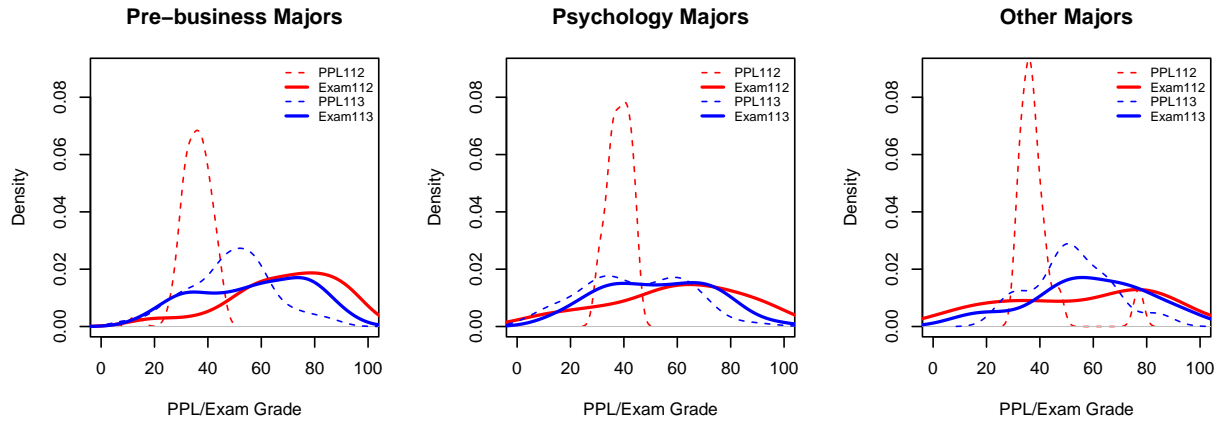


Figure 9: Potential Difference adjusted by majors and PPL scores

The shapes of the density curves of the exam grades of the two major cohorts were similar to each other. In other words, the mean exam scores among the three categories of majors were not significantly different. Therefore, students from the three major categories will not contribute to the *net performance gain*.

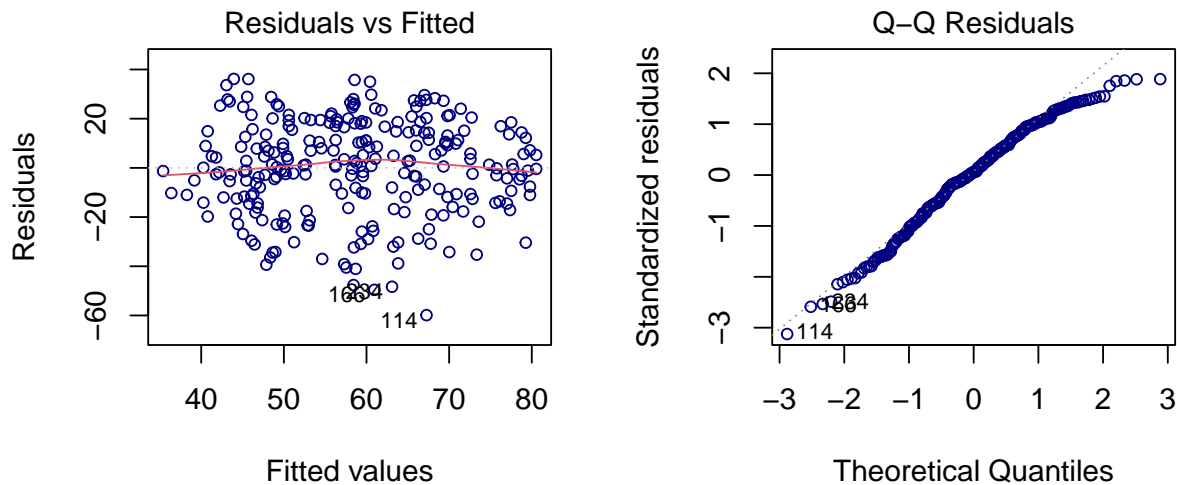


Figure 10: Diagnostic plots of linear regression model: No significant violations.

The above residual diagnostic plots do not indicate serious violations of assumptions.

Table 5: Inferential results of regression model: adjusted with PPL, Instr., and Major

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	50.3654929	4.9917531	10.0897403	0.0000000
mpeScore	0.2521266	0.0906324	2.7818589	0.0058236
factor(InstrID)2	-6.1548344	3.5798542	-1.7192975	0.0868181
factor(InstrID)3	12.2163938	3.5565279	3.4349214	0.0006955
factor(InstrID)4	19.9575119	3.3006787	6.0464873	0.0000000
majorypeprebus	-0.8827623	3.5435672	-0.2491168	0.8034783
majorypepsych	-5.4041901	3.8852703	-1.3909431	0.1654995
CourseNumMAT113	-11.4843369	2.7146948	-4.2304340	0.0000329

The above regression results also confirmed major-type to be insignificant ($p > 0.1$).

4.2 Potential ESL Effects

There are 24 ESL students in this study (about 10% of the total participants). The next two figures show the same type of distributional information.

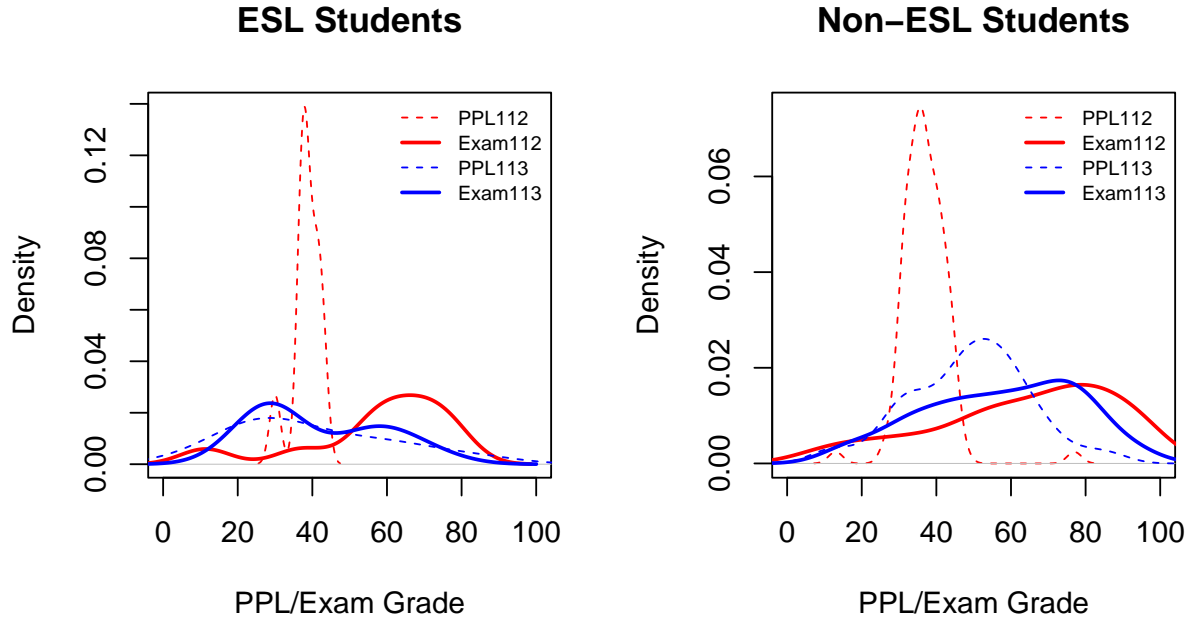


Figure 11: Potential Difference adjusted by ESL and PPL scores

For ESL students, 60% of MAT112 students earned 60 or above while 21% of MAT113 students earned 60 or above. The mean PPL scores of both MAT112 and MAT113 are close to each other but the MAT112 cohort had a significantly small variance compared to the MAT113 cohort. *The distributions of PPL and exam grades for both MAT112 and MAT113 were essentially the same as the overall distributions as depicted in Figure 1.*

Table 6: Five-number-summary of exam grades of students in MAT112 and MAT113 respectively.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
ESL112	11.363636	54.54545	62.50000	57.72727	70.45455	78.40909
ESL113	21.022727	28.69318	31.81818	41.72078	56.25000	72.72727
nonESL112	7.386364	49.00568	68.46591	63.43155	82.52841	96.59091
nonESL113	11.931818	42.18750	60.22727	57.38242	74.57386	96.59091

ESL students in MAT112 outperformed their counterparts in MAT113. The same pattern was observed in nonESL students. However, Among MAT112 students, nonESL students outperformed ESL students, the same pattern was also observed in MAT113. These observations could affect the actual *net performance gain*. Therefore, ESL should be added to the previous regression model.

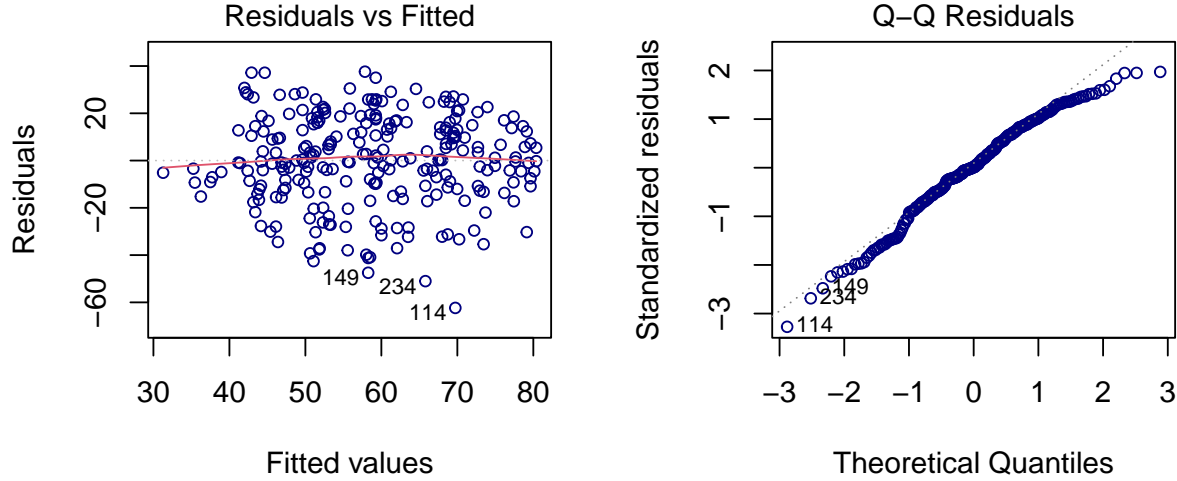


Figure 12: Diagnostic plots of linear regression model: No significant violations.

No significant violations of model assumptions were found from the above residual diagnostic plots. We summarize the regression results in the following table

Table 7: Inferential results of regression model: adjusted with PPL, Instr., ESL

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	50.312759	4.1843609	12.024001	0.0000000
mpeScore	0.248542	0.0894213	2.779449	0.0058639
factor(InstrID)2	-8.187107	3.5117401	-2.331353	0.0205399
factor(InstrID)3	9.480932	3.4311007	2.763233	0.0061540
factor(InstrID)4	19.130924	3.1497615	6.073769	0.0000000
ESLYes	-10.128765	4.2175954	-2.401550	0.0170655
CourseNumMAT113	-11.145949	2.6872504	-4.147715	0.0000462

The above table shows the statistical significance of ESL. After being adjusted with ESL, the *net performance gain* increased slightly from 11.22 to 11.48. The small change in *net performance gain* is due to the small number of ESL students in the study.

4.3 Grade Level

The following 2-way table shows the distribution of enrollments MAT112 and MAT113 by grade levels.

	MAT112	MAT113
Freshman	92	91
Junior	1	4
Senior	1	2
Sophomore	2	61

Since almost all MAT112 students were freshmen, grade level will *not* be included in the model as an adjusting variable. The following figures provide various visual comparisons of distributions between MAT112 and MAT113 students.

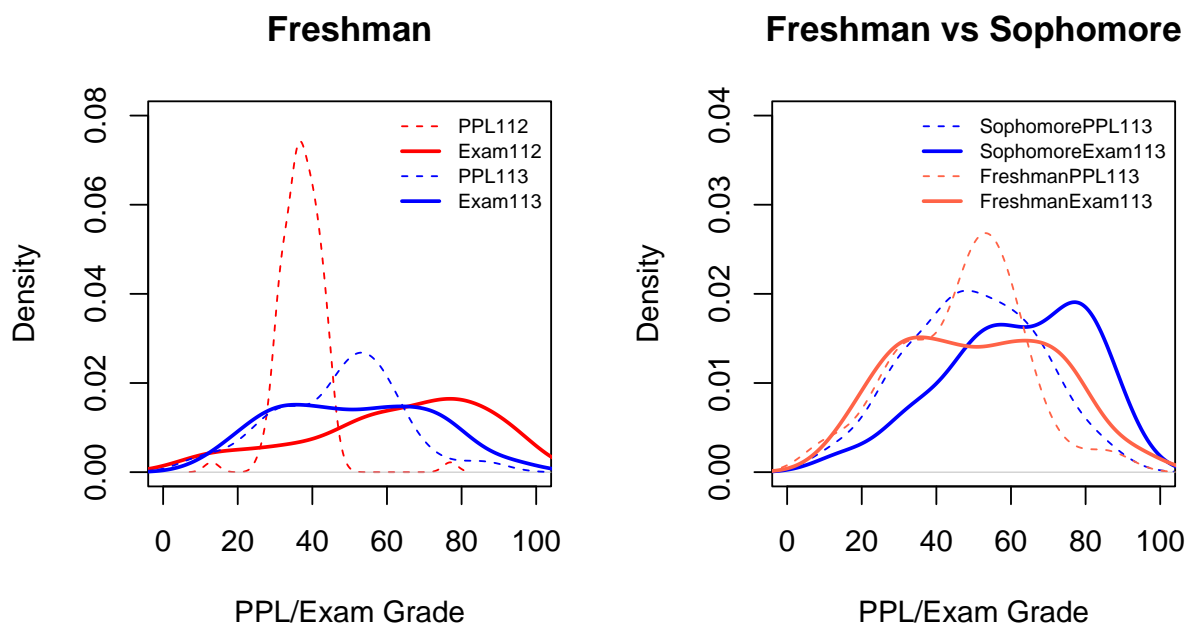


Figure 13: Potential Difference adjusted by grade level (freshman and sophomore) and PPL scores

We can see from the above figure that MAT112 students outperformed MAT113 students.

5 Question-specific Comparison - A Visual Inspection

The question-specific sub-scores were also converted to the corresponding percentiles in all comparisons. Since this analysis is not critical to address the sole goal of the project, we will *not* expand the analysis to the same granular level as we did in the previous few sections unless team MAT112/113 wants to explore potential pedagogical strategies to teach specific topics.

5.1 Baseline Distributions Of Individual Questions

To provide a starting point for potential exploration, we provide only some graphic representations of the distributions of the converted exam grades for individual questions for every instructor. To avoid cross-comparison among instructors, we defined the baseline distributions of each question based on all 278 participating students. Individual instructors' question-specific distributions were based on all MAT112 and MAT113 students they taught. Therefore, individual instructors can compare their own distribution with the common baseline distributions.

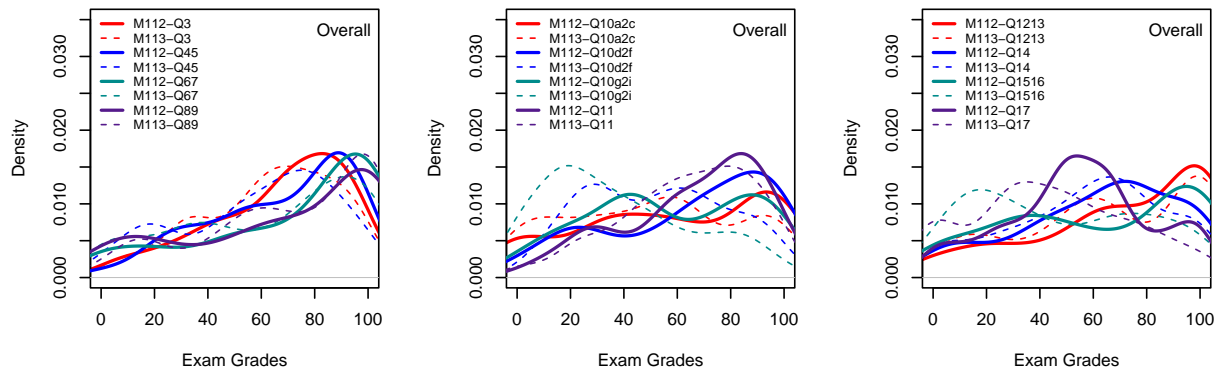


Figure 14: Topic-specific difference between MAT112 and MAT113: Overall

To help read the above density curves, we pick the density distributions of Q89 in MAT112 and MAT113 from the left panel of the above figure to explain the distributional difference between two exam grades.

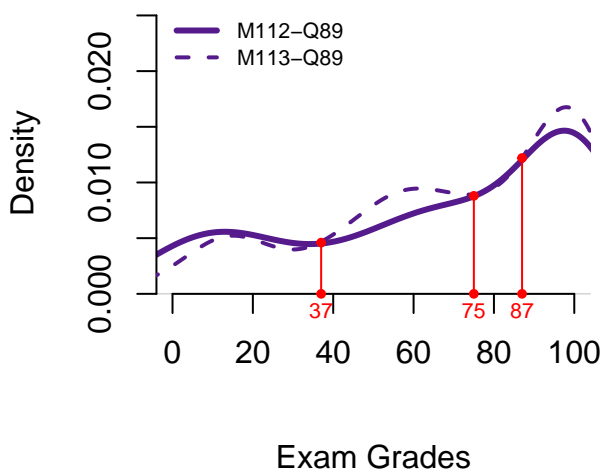


Figure 15: Topic-specific difference between MAT112 and MAT113: Questions 8 and 9.

As stated earlier, the grades Q89 were converted to the percentile. The above figure shows that,

- the proportion of students who earned 37 or less in MAT112 was higher than that in MAT113
- the proportion of students who earned 87 or more in MAT112 was lower than that in MAT113
- the proportion of students who earned between 75 and 87 in MAT112 was about the same as that in MAT113
- the proportion of students who earned between 37 and 85 in MAT112 was lower than that in MAT113

The density curves of grades of individual questions in the above figures and the following instructor-specific density distributions can be similarly interpreted.

5.2 Instructor and Question-Specific Distributions

The next four figures show the density curves of distributions of grades of individual questions earned by MAT112 and MAT113 students from individual instructors.

5.2.1 Instructor A

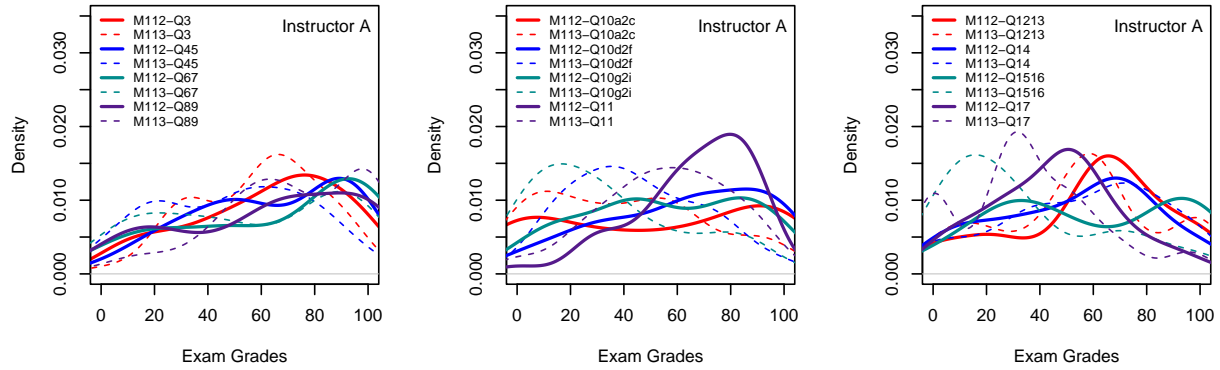


Figure 16: Topic-specific difference between MAT112 and MAT113: Instructor A

5.2.2 Instructor B

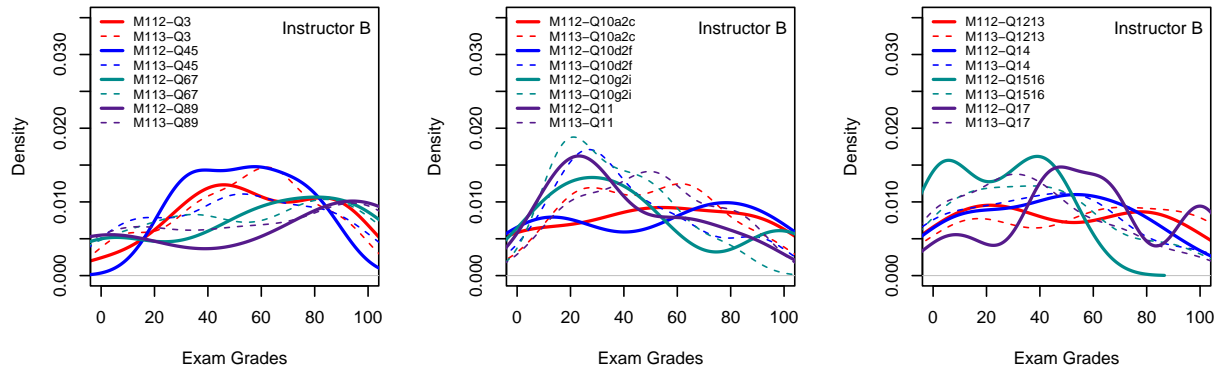


Figure 17: Topic-specific difference between MAT112 and MAT113: Instructor B

5.2.3 Instructor C

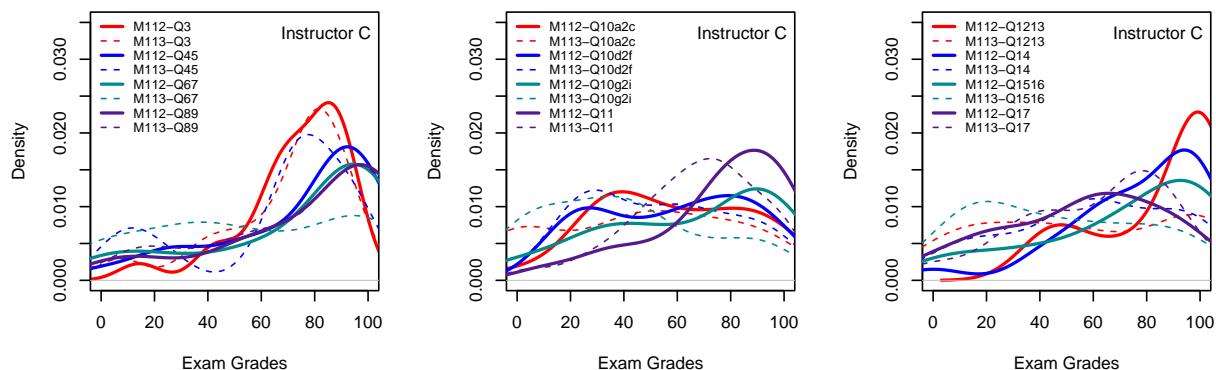


Figure 18: Topic-specific difference between MAT112 and MAT113: Instructor C

5.2.4 Instructor D

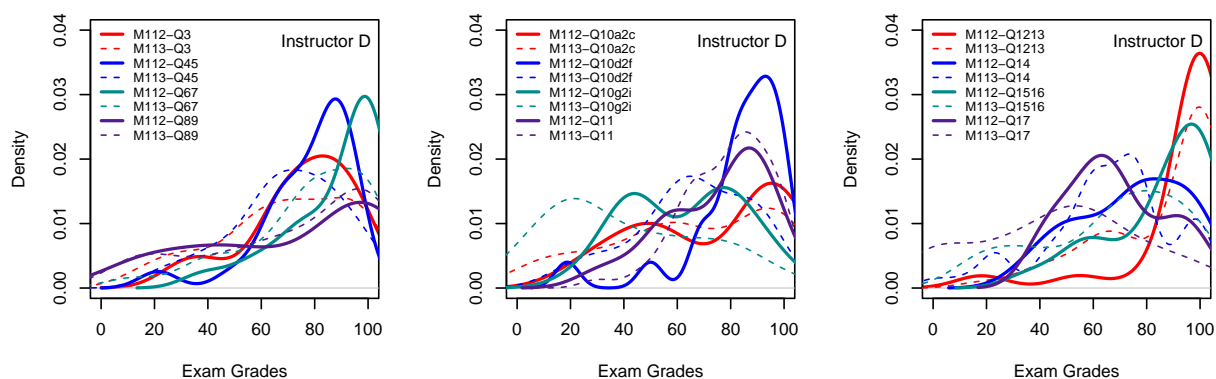


Figure 19: Topic-specific difference between MAT112 and MAT113: Instructor D

6 Conclusions

As stated at the beginning of this report, the goal of this project is to assess the *net performance gain* of the one extra credit added to MAT112. The measurement used for this assessment was the exam grade of the common exam. Since various factors may affect the exam grade, we identified several key factors and used them to adjust the analysis to extract the net difference between exam grades in MAT112 and MAT113 (i.e., **net performance gain**).

Some routine data processing and variable transformation/conversion were performed to remove some potentially sensitive information. The issue of sparse categories for some categorical variables was also handled to improve data quality.

Descriptive methods and visual tools were used to identify potential factors that influence the exam scores. We also used linear regression models to extract the true difference of the mean exam scores between MAT112

and MAT113 (also called *net-gain*) with adjustments due to other influential factors such as placement scores (PPL) and the instructors.

Key findings are summarized in the following.

- MAT112 students outperformed significantly MAT113 students. The *net performance gain* was about 11 points on a 100-point scale after the model was adjusted by factors such as PPL, instructor effect, etc.
- The math placement scores were positively correlated with the final exam scores ($p < 0.05$). The strength of the correlation was weak (the correlation coefficient $r \approx 0.082$).
- The average PPL score of MAT112 was significantly lower than that of MAT113 students, but the average exam score of MAT112 was significantly higher than that of MAT113.
- Instructor effect was significant in that it influenced the *net performance gain* substantially.
- ESL students(native language is not English) performed poorly than non-ESL students.
- Other factors such as major and grade level had insignificant influence on the performance in the exam.
- Question-specific comparisons were not systematically performed but the relevant density curves of grade distributions were provided for potential further exploration.