A Comparison of ALEKS PPL and WCU Placement Test

Department of Mathematics

3/16/2021

Executive Summary

We have conducted both descriptive and inferential analyses from different perspectives to compare the two placement tests.

The direct comparison of the scoring distributions indicates that the two placement tests are significantly different (Pearson chi-square test yields a p-value ≈ 0). This means that the same score from different placement tests does not imply the same level of mathematical proficiency.

The direct comparison of the overall grade distributions of ALEKS and WCUPT students in their "first math course" indicates that there is no significant difference between the two placement tests (Table 5) in terms of DFW rates. However, when comparing the grade distribution at the course level, the same grade distributions corresponding to ALEKS and WCUPT students are significantly different. This implies that the behaviors of the two placement scoring systems are different.

The comparison of the grade distributions based on a subset of data with instructors teaching the same courses to BOTH ALEKS and WCUPT students also reveals the same pattern: No significant difference between ALEKS and WCUPT was found but the significant difference was found at individual instructor's courses.

A descriptive analysis of the DFW rates of the involved instructors in the entire data showed a wide variation. This is an area that requires a team effort to reduce the variation of DFW rates by tuning the scoring system of the placement tests (either ALEKS or WCUPT) and adjusting instructors' grade distributions so the department as a team to have a consistent and "more fair" scoring practices.

An inferential logistic regression was built to explore the potential impact of the placement tests on the odds of earning a grade of C or better. The impact of the placement type on the odds of success is dependent on the test score. The results indicate that ALEKS and WCUPT are NOT significantly different. However, descriptive statistics showed that ALEKS is marginally better than WCUPT in terms of odds of success.

Project Team Members: Brian Bowen, Peter Glidden, Cheng Peng, and Susan Turner
Data management, analysis, and project report: Cheng Peng
Data Collection: Nilima Inamdar, WCU's Institutional Research

Contents

1 Objectives

Comparing the discriminatory performance of the two competing placement instruments: ALEKS PPL vs WCU homegrown placement.

2 The Formulation of Analytical Problems

The following terms will be used throughout this report.

- **ALEKS** - ALEKS PPL placement test
- **WCUPT** - WCU homegrown placement test
- **WCUPT students** - Students who took WCU homegrown placement test
- **ALEKS students** - students who took ALEKS PPL.
- **Study population** - all students in the study.
- **Math Grades** - grades earned from the "first" math class after they took either one of the placement tests.
- **DFW** - a grade in either D range, F or W.
- **success** - passing a course with a grade C or better
- **failure** - failing a course with a grade of DFW
- **odds of success** - the probability of passing a course with a grade C or better divided by the probability of failing the same course with a DFW grade. Caution: the odds are not the same thing as the probability.
- **OR** - the ratio of the odds of success between two groups of subjects with the same characteristics. In this report, subjects are students with the same placement score and in the same class taught by the same instructor.

2.1 The Nature of the Two Placement Instruments

Two tests are designed for the same purpose – placing students to appropriate mathematics classes. The two tests use different grading scales. The ALEKS PPL scores are converted to a 5-point scale, the scale used in WCU MAT, to make a placement decision.

2.2 Conversion of the Two Scoring Systems

The cut-offs initial ALEKS PPL scores used for converting to a 5-point-scale are determined by WCU's mathematics department. This analysis does not intend to discuss the determination of optimal cut-off scores. Instead, we use the existing conversion cut-offs to see the potential discrepancy between the two placement instruments.

2.3 The Rationale and Base of Proposed Analyses

If the two instruments do not have a significant difference, we will expect to see a significant difference in following grade distributions

2.3.1 Direct Comparison of Placement Scores

The distribution of the scores based on WCU MAT is expected not to be significantly different from that based on the converted ALEKS PPL scores. This hypothesis assumes that incoming student populations do not change significantly before and after using the ALEKS PPL.

2.3.2 Indirect comparison - Performance in First Mathematics Courses

Whether a placement test correctly places students to appropriate courses can be assessed by students' performances in corresponding courses taken immediately after the placement test. We will look into the

grade distributions of these "first courses" to see the potential discrepancy between ALEKS PPL or WCU MAT.

2.3.3 D/F/W Rate Analysis

As a special case of 2.3.2, we can dichotomize the course letter grades into D/F/W and non-D/F/W. The comparison of the grade distributions associated with two placement tests is equivalent to comparing the D/F/W rates associated with ALKES PPL and WCU MAT, respectively.

3 Data Sources and Limitations

The comparison is based on WCU's institutional research database.

3.1 Two data sources were extracted from WCU's PeopleSoft.

Both data sources were provided by WCU's Office of Institutional Research. Both data sources are at the individual level with all personally identifiable information being removed to protect individual identities and privacy.

- Student Math Course Grade Data

The course grades of the "first math course(s)" immediately after one of the placement tests with the key stratification variables course ID and proxy faculty ID. The several other variables in the data are relevant to this specific comparison between the two placement tests.

- Placement Score Data

The raw 5-point-scale scores of WCU MAT and the converted 5-point-scale scores from ALEKS PPL together with the date of the placement test including multiple attempts.

- Data Collection Timeframe:

Three semesters (Summer 2019, Fall 2019, or Spring 2020)immediately before and two semesters (Summer 2020 or Fall 2020) after launching the ALEKS PPL.

3.2 Limitations – Unobservable Confounding Factors

There several limitations

- ALEKS PPL Scoring Conversion

The ALEKS PPL adopts a 1-100 scoring system. Currently, WCU converts the ALEKS PPL default scoring system to a 5-point scoring system that was used in WCU MAT based on ALEKS's suggested cut-off scores. This conversion between the two different scoring systems associated with ALEKS PPL and ECU MAT could confound the discrepancy between the two placement tests.

- Pandemic Impact on Learning and Teaching

The ALEKS PPL started in April 2020 when the University moved almost all face-to-face courses to remote instruction due to the pandemic. The impact of the change of instruction modality on instructors' teaching, students' learning, and particularly the instructors' grading remains unknown based on the available data[2,3]. This is an uncontrollable confounding factor that could potentially impact the outcome of the comparison of the two placement tests.

4 Descriptive Comparisons

The analyses will be conducted in the following order: detect discrepancy and identify the goodness of the individual placement tests under study. We will define success measures in the subsequent math courses

such as rates of A, B or better, C or better, and D/F/W rates associated with the two placement tests, respectively. Descriptive statistics, visualizations, and inferential statistics.

4.1 Description of the analytical data set

4.1.1 Student inclusion criteria

- students took one of the two placement tests, and
- also took the "first math course(s)" between Fall 2019 and Fall 2020.

Note: if a student took two math courses immediately after the placement test, both math courses will be considered as "first math courses".

4.1.2 Variables to include in the study:

- Proxy Student ID – used the primary key to join data the two data sets
- Response variable: Math Grades – for indirect analysis
- Placement type – derived based on the test dates
- Placement test scores – used in direct analysis
- Math course ID – to be as stratification variables and predictor variable as well in generalized linear regression models as an adjusted variable.
- Proxy Instructor ID – to be used as a stratification and adjusted variable in both descriptive and inferential modeling.
- Term ID – will be used as an adjusted variable.

4.1.3 Final Analytic Data

The final analytic data was merged from math grade data and placement test data. Some data management and cleansing were performed.

- **Math grade data**
  - All students who took placement tests but have not yet taken any math courses were dropped from the final data set.
  - All records having a course taken beyond the data collection time were excluded from the data set.
  - For records associated with students who took math courses in different semesters, we keep the earliest record and drop the rest of the records.
- **Placement score data**
  - A significant number of WCUPT students scored 0. Categories 0 and 1 in the WCUPT scoring system were combined and relabeled as 1.
  - 62 students attempted the placement test twice. The portion of that group of students is about 1.5% of the data set. We only kept the record with the highest score.
  - 64 records have grades in the forms of "IP", "NG", "Z", and "AU". These grades consist of about 1.5% total study population. They were included in the data but excluded in some of the descriptive statistics and inferential models.

```
total = length(unique(combined.data$SID))
num.courses = length(unique(combined.data$Course))
freq.Transf = table(combined.data$AdmitType)[2]
freq.First = table(combined.data$AdmitType)[1]
freq.WCU = table(combined.data$PlaceType)[2]
freq.ALEKS = table(combined.data$PlaceType)[1]
freq.count = cbind(Student.count = total, number.of.course = num.courses,
                   number.of.transfers = freq.Transf, number.of.first.year = freq.First,
```

```
                    WCU = freq.WCU, ALEKS = freq.ALEKS)
row.names(freq.count) = ""
kable(freq.count, caption = "Frequency Summary")
```

Table 1: Frequency Summary

Student.count	number.of.course	number.of.transfers	number.of.first.year	WCU	ALEKS
4146	14	281	3930	2425	1786

4.2 Direct Comparison

Direct comparison using the grade distributions resulted from the ALKES PPL and WCU MAT respectively. This analysis is based on the idea outlined in 2.3.1. This is a starting point

4.2.1 Visual Analysis

The following group bar-plot displays the frequency distributions of the placement scores by WCU home-grown test and the ALEKS. Assuming that the incoming students are at the same level of preparedness for college, we would expect the distributions of the placement test scores are supposed to be the same regardless of the type of the test if the two scoring systems are equivalent.

```
## relative frequency distributions of the placement scores by placement type
counts <- table(combined.data$PlaceType, combined.data$PlaceScore)
counts0 <- rbind(ALEKS =counts[1,]/sum(counts[1,]), WCU = counts[2,]/sum(counts[2,]) )
mycols = c("darkblue","skyblue")
barplot(counts0, main="Placement Score Distribution by Placement Type",
        xlab="Placement Score",
        col = mycols,
        cex.main = 0.8,
        col.main = "navy",
        beside=TRUE)
 legend(x = "topright", legend = rownames(counts), fill = mycols,
        bty = "n", y.intersp = 1, cex = 0.8)
```

The above relative frequency bar plot shows that the distributions of placement scores based on WCU's home-grown placement and the ALEKS are significantly different.

4.2.2 Descriptive analysis

The following plot displays the discrepancy between the distributions of the two placement tests.

```
ALEKS =counts[1,]/sum(counts[1,])
WCU = counts[2,]/sum(counts[2,])
plot(0:5, 100*WCU, type = "b", lwd = 2, col = "navy",
     ylim=c(0,40),
     xlab="Placement Test Score",
     ylab="Relative Frequency (%)",
     axes = FALSE,
     main = "Relative Frequency Distribution of Placement Score \n  WCU v.s. ALEKS",
     cex.main = 0.8,
     col.main = "navy")
axis(1, pos = 0)
axis(2, pos = -0.4)
lines(0:5, 100*ALEKS, type = "b", lwd =2, col ="darkred")
```

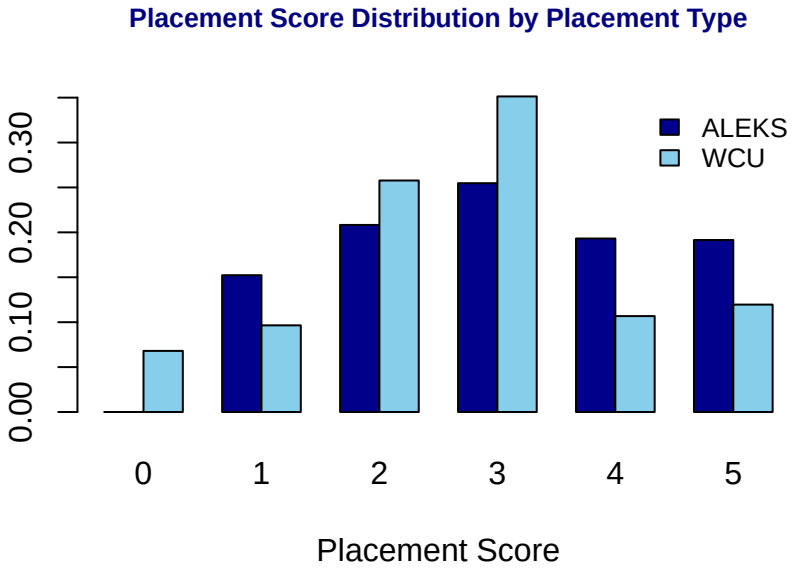


Figure 1: Probability distribution histogram of the scores of the two placement tests

```
legend("topright", c("WCUPT", "ALEKS"), col=c("navy", "darkred"), lwd=rep(2,2),
       cex = 0.7, bty="n")
text(0:5, 100*ALEKS+3*c(1,1,-1,1,1,1), as.character(round(100*ALEKS,1)),
     col="darkred", cex = 0.7)
text(0:5, 100*WCU+3*c(1,-1,1,1,-1,1), as.character(round(100*WCU,1)),
     col="navy", cex = 0.7)
```

The relative frequency of the converted scores from each of the placement tests are summarized in the following table.

```
kable(counts0,caption = "Relative frequency distributions of the
      placement scores associated with ALEKS and WCUPT")
```

Table 2: Relative frequency distributions of the placement scores associated with ALEKS and WCUPT

	0	1	2	3	4	5
ALEKS	0.0000000	0.1522956	0.2082867	0.2547592	0.1931691	0.1914894
WCU	0.0680412	0.0964948	0.2577320	0.3513402	0.1068041	0.1195876

Of all students who took ALEKS, 36% of them scored 2 or less and 64% of them scored 3 or higher. For students who took WCU's homegrown placement test, 42% of them scored 2 or less and 58% of them scored 3 or higher. 38.4% of students who took ALEKS and scored either 4 and 5 while only 22.7% of students who took the WCU homegrown placement test scored either 4 or 5.

4.2.3 Inferential analysis

generalized linear regression models such as the ordinal logistic regression model will be used to assess the potential difference between the two placement tests.

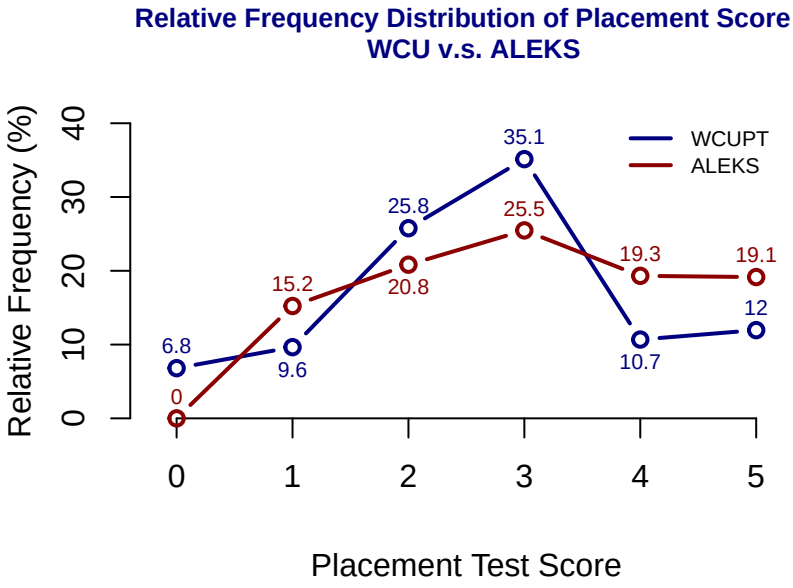


Figure 2: Line plot of the relative frequency distribution of the two placement test scores

```
par(mar = c(0, 2, 0, 2))
dat = as.data.frame(cbind(combined.data$PlaceType, combined.data$PlaceScore))
names(dat) = c("PlaceType", "PlaceScore")
mosaic(~ PlaceType + PlaceScore,
       direction = c("v", "h"),
       data = dat,
       shade = TRUE)
```

The above figure shows that the distributions of the scores of the two placement tests are statistically different with a p-value < 0.0001 based on the Pearson chi-squared test.

In summary, the direct comparison between the scores of the two placement tests significantly different. The ALKES tends to produce more 4s and 5s while WCU's homegrown placement test (WCUPT) produces more low scores. Therefore, the two scoring scales are not consistent. This implies that some adjustments are needed to make the two scoring systems comparable. Otherwise, the ways of placing students to appropriate courses using different placement test scoring systems should be different.

4.3 Indirect Comparison

Indirect comparison by looking at the performance in the subsequent mathematics class (focus on the first mathematics class taken immediately after the placement tests.

4.3.1 Descriptive Approach

We break down the data to course level and compare the above-mentioned "success rates" associated with the two placement tests.

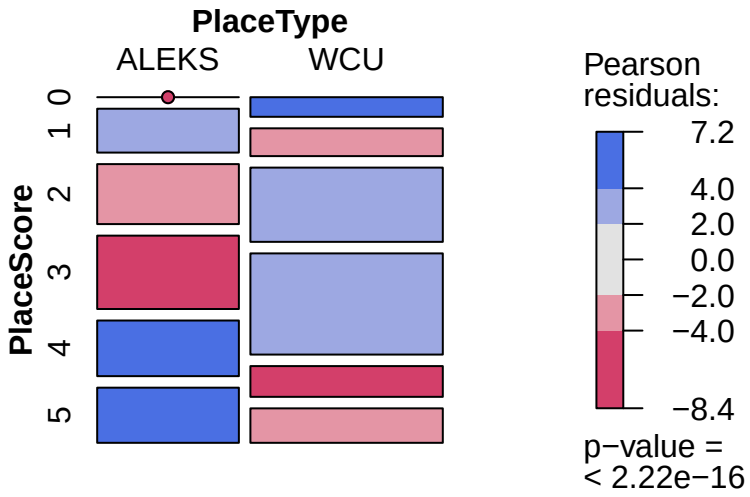


Figure 3: Pearson Chi-square test and the graphic representation of the residuals and relative frequencies of each of the placement test scores

```
WCU.id = which(combined.data$PlaceType=="WCU")
ALEKS.id = which(combined.data$PlaceType=="ALEKS")
course.WCU = combined.data$Course[WCU.id]
Score.WCU = combined.data$PlaceScore[WCU.id]
WCU=table(course.WCU,Score.WCU)
course.ALEKS = combined.data$Course[ALEKS.id]
Score.ALEKS = combined.data$PlaceScore[ALEKS.id]
ALEKS=table(course.ALEKS,Score.ALEKS)
W1 = (WCU[,1] + WCU[,2])/sum(WCU[,1] + WCU[,2])
W2 = WCU[,3]/sum(WCU[,3])
W3 = WCU[,4]/sum(WCU[,4])
W4 = WCU[,5]/sum(WCU[,5])
W5 = WCU[,6]/sum(WCU[,6])
##
A1 = ALEKS[,1]/sum(ALEKS[,1])
A2 = ALEKS[,2]/sum(ALEKS[,2])
A3 = ALEKS[,3]/sum(ALEKS[,3])
A4 = ALEKS[,4]/sum(ALEKS[,4])
A5 = ALEKS[,5]/sum(ALEKS[,5])
course.place.dist = 100*round(cbind(W1=W1, A1=A1,W2=W2,A2=A2,W3=W3,
                          A3=A3, W4=W4,A4=A4,W5=W5,A5=A5),4)
kable(course.place.dist, caption = "The comparison of placement distributions
      for individual placement score by the placement types ")
```

Table 3: The comparison of placement distributions for individual placement score by the placement types

	W1	A1	W2	A2	W3	A3	W4	A4	W5	A5
MAT-101	0.25	0.74	14.72	20.43	9.74	15.38	9.27	10.43	5.17	8.77
MAT-103	0.50	0.37	20.00	15.32	11.27	5.93	6.95	6.38	6.55	4.39
MAT-104	0.00	0.00	3.20	3.76	1.06	2.20	0.77	0.58	0.34	0.88
MAT-113	0.00	0.37	0.64	0.27	39.55	38.02	6.56	14.20	5.86	5.85
MAT-115	0.00	0.00	0.00	0.00	1.06	2.64	0.00	1.16	0.00	2.63
MAT-121	0.25	0.37	32.32	29.84	15.02	12.97	22.01	19.42	15.86	14.91
MAT-125	0.00	0.00	0.48	1.08	5.28	5.93	8.88	8.99	14.83	11.11
MAT-131	0.25	0.00	0.48	0.00	14.67	11.43	16.22	14.20	2.41	2.63
MAT-143	0.25	0.00	0.64	0.00	0.47	0.22	25.10	18.26	17.59	14.33
MAT-145	0.00	0.00	0.16	0.00	0.00	0.44	0.77	1.74	2.76	3.80
MAT-151	0.25	0.00	0.80	0.00	1.64	0.44	1.16	1.16	5.52	2.34
MAT-161	0.25	0.00	0.16	0.00	0.00	0.22	1.93	0.58	22.76	26.90
MAT-Q20	42.36	28.31	2.24	0.81	0.12	1.54	0.00	0.00	0.00	0.29
MAT-Q30	55.64	69.85	24.16	28.49	0.12	2.64	0.39	2.90	0.34	1.17

The above table shows how students were placed into different courses with a given placement score. The uppercase letter in the column represents the type of the placement tests (A = ALEKS, W = WCU home-grown test) and the number represents the score of the replacement test. Since ALEKS does not generate a score, WCUPT scores 0 and 1 were combined as labeled as 1.

We can see from the following figure that almost all students who scored 1 in both placement tests were placed into MAT-Q20 or MAT-Q30. The following Figures explore whether the same score obtained from different placement tests placed students into different MAT courses in the same way.

```
#col.id = rownames(course.place.dist)
col.id = c("101", "103", "104", "113", "115", "121", "125", "131", "143",
           "145", "151", "161", "Q20", "Q30")
par(mfrow=c(1,2))
plot(1:14, A2, type="l", col="navy", xlab="", ylab="Percent (%)",
     ylim=c(-0.1, 0.6),
     main ="Placement behavior of indicidual scores \n
     from WCU and ALEKS: scores 2 and 3",
     cex.main=0.8, col.main="navy", axes = FALSE)
axis(1, pos=-0.05, at=1:14, labels = col.id, las=2, cex.lab =0.6, col.lab="navy")
axis(2)
points(1:14, A2, pch = 20, col = "navy")
lines(1:14, W2, type = "l", col = "darkred")
points(1:14, W2, pch = 20, col = "darkred")
lines(1:14, A3, type = "l", col = "blue")
points(1:14, A3, pch = 20, col = "blue")
lines(1:14, W3, type = "l", col = "purple")
points(1:14, W3, pch = 20, col = "purple")
legend("topright", c("ALEKS score: 2", "WCU score: 2", "ALEKS score: 3",
                     "WCU score: 3"), lty=rep(1,4),
       col=c("navy", "darkred", "blue", "purple"), pch=rep(20, 4),
       bty="n", cex = 0.7)
##
plot(1:14, A4, type="l", col="navy", xlab="", ylab="Percent (%)",
      ylim=c(-0.1, 0.6),
     main ="Placement behavior of indicidual scores \n
```

```
     from WCU and ALEKS: scores 4 and 5",
     cex.main=0.7, col.main="navy", axes = FALSE)
axis(1, pos=-0.05, at=1:14, labels = col.id, las=2, cex.lab =0.6, col.lab="navy")
axis(2)
points(1:14, A4, pch=20, col = "navy")
lines(1:14, W4, type = "l", col = "darkred")
points(1:14, W4, pch=20, col = "darkred")
lines(1:14, A5, type = "l", col = "blue")
points(1:14, A5, pch=20, col = "blue")
lines(1:14, W5, type = "l", col = "purple")
points(1:14, W5, pch=20, col = "purple")
legend("topright", c("ALEKS score: 4", "WCU score: 4", "ALEKS score: 5",
                     "WCU score: 5"), lty=rep(1,4),
                     col=c("navy", "darkred","blue", "purple"),
                     pch=rep(20, 4), bty="n", cex = 0.7)
```

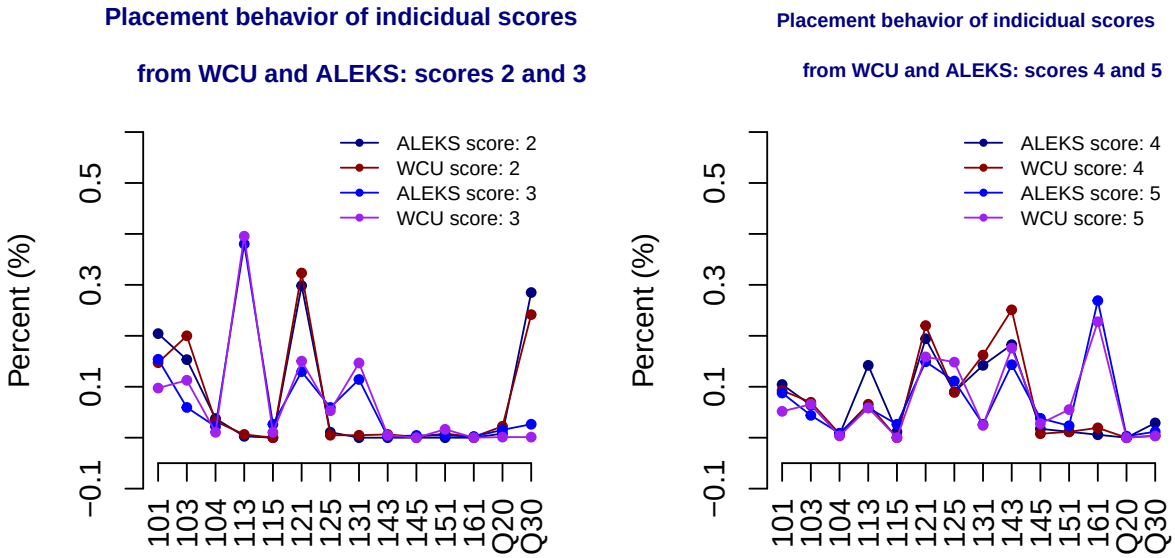


Figure 4: The percentage of students placed into different courses with a given placement test score.

Students who earned a score of 2 or 3 from either of the placement tests seemed to be arbitrarily placed into MAT-101 and MAT-103. Students who scored 4 or 5 seemed to be placed into similar courses consistently. Overall, students with the same test score (from either ALEKS or WCUPT) were placed into similar courses except for MAT-121 and MAT-131 that had a slightly different pattern. We will revisit these two courses in the stratification analysis.

We can see from the following table that ALEKS placed more students in MAT-12X courses: MAT-125, MAT-143, MAT-145, and MAT-161. This is because the ALEKS tends to produce higher placement scores as pointing out earlier. The numbers in the table are row percentages of students who took the corresponding placement test were placed in different suggested mathematics courses based on the placement scores. For example, ALEKS placed about 5.3% of students in MAT-161 while WCU's homegrown test placed about 3.0% of students in the same MAT-161.

```
course.cts0 =table(combined.data$PlaceType, combined.data$Course)
course.cts = 100*round(rbind(course.cts0[1,]/sum(course.cts0[1,]),
                             course.cts0[2,]/sum(course.cts0[2,])),3)
column.id = c("101", "103", "104", "113", "115", "121", "125", "131",
              "143", "145", "151", "161", "Q20", "Q30")
colnames(course.cts) = column.id
rownames(course.cts) = c("ALEKS", "WCU")
kable(course.cts, caption="Two-way contingency table: placement
      test type vs first math course")
```

Table 4: Two-way contingency table: placement test type vs first math course

	101	103	104	113	115	121	125	131	143	145	151	161	Q20	Q30
ALEKS	12.0	6.8	1.6	13.7	1.4	16.2	5.6	6.2	6.3	1.2	0.8	5.3	4.9	18.0
WCU	8.9	10.7	1.3	15.5	0.4	17.9	4.7	7.3	5.2	0.5	1.6	3.0	7.6	15.5

In the next section, we will present some visual comparisons to see whether the the type of placement tests had an impact on students' performance (course grade) in their "first math" course.

4.3.2 Stratified Analysis

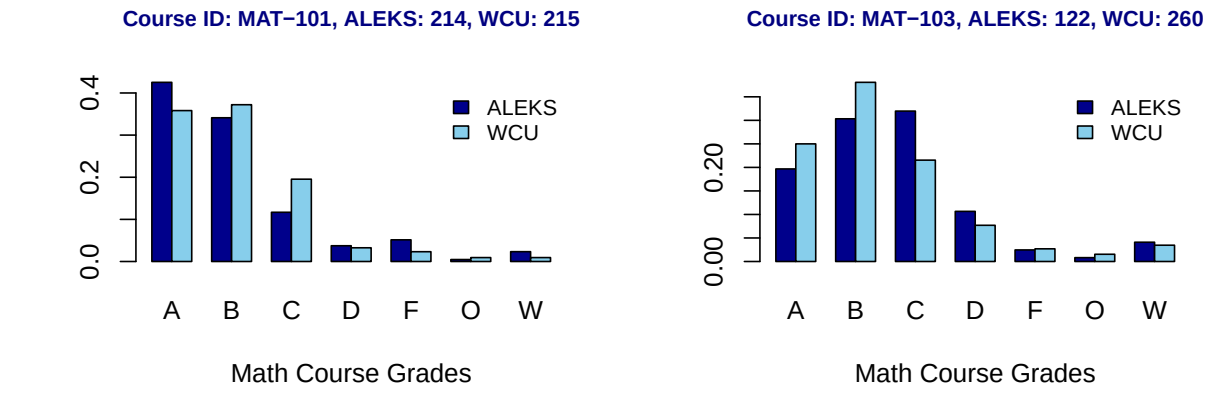
We will to two types of descriptive stratification analyses: course level and instructor level.

4.3.2.1 Course Level Analysis Several group bar-plots are based on the relative frequency distribution of the math course grades earned by students who took ALEKS and WCUPT.

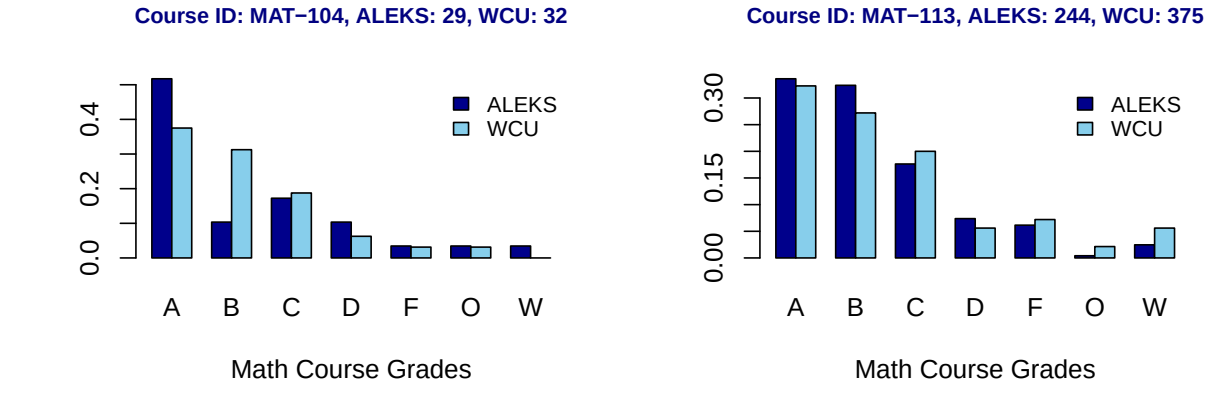
```
#course.id = "MAT-101"
course.comparison = function(course.id){
  course.data = combined.data[which(combined.data$Course==course.id),]
  grade0 = course.data$Grade
  other.id = grade0 %in% c("AU", "NG", "IP", "Z")
  A.id = grade0 %in% c("A", "A+", "A-")
  B.id = grade0 %in% c("B", "B+", "B-")
  C.id = grade0 %in% c("C", "C+", "C-")
  D.id = grade0 %in% c("D", "D+", "D-")
  F.id = grade0 %in% c("F")
  W.id = grade0 %in% c("W")
  ##
  course.data$Grade[A.id] = "A"
  course.data$Grade[B.id] = "B"
  course.data$Grade[C.id] = "C"
  course.data$Grade[D.id] = "D"
  course.data$Grade[F.id] = "F"
  course.data$Grade[W.id] = "W"
  course.data$Grade[other.id ] = "O"
  ##
  counts <- table(course.data$PlaceType, course.data$Grade)
  counts0 <- rbind(ALEKS =counts[1,]/sum(counts[1,]), WCU = counts[2,]/sum(counts[2,]) )
  mycols = c("darkblue","skyblue")
  barplot(counts0, main=paste("Course ID: ",course.id, ", ALEKS: ", sum(counts[1,]),
                              ", WCU: ", sum(counts[2,]), sep = "" ),
        xlab="Math Course Grades",
```

```
        col = mycols,
        cex.main = 0.8,
        col.main = "navy",
        beside=TRUE)
   legend(x = "topright", legend = rownames(counts), fill = mycols,
        bty = "n", y.intersp = 1, cex = 0.8)
  }
```

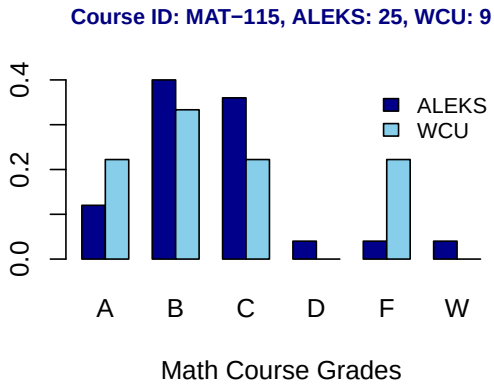
```
par(mfrow=c(1,2))
course.comparison(course.id = "MAT-101")
course.comparison(course.id = "MAT-103")
```

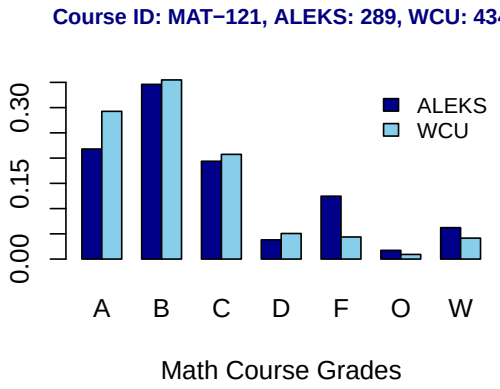



```
par(mfrow=c(1,2))
course.comparison(course.id = "MAT-104")
course.comparison(course.id = "MAT-113")
```

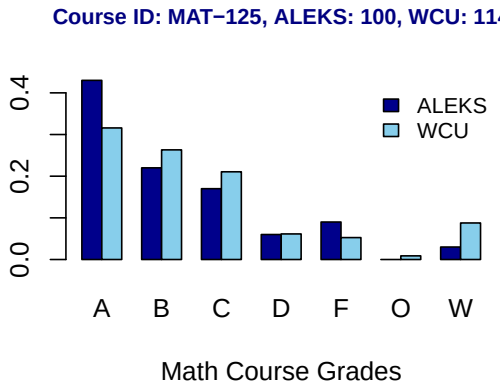


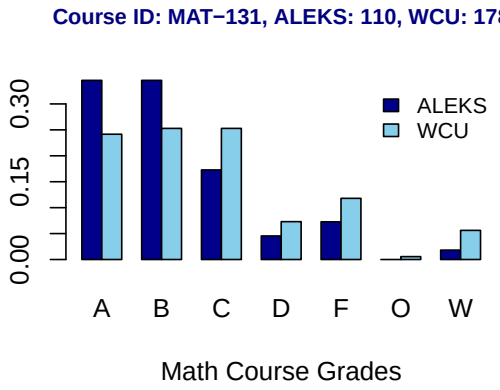

```
par(mfrow=c(1,2))
course.comparison(course.id = "MAT-115")
course.comparison(course.id = "MAT-121")
```



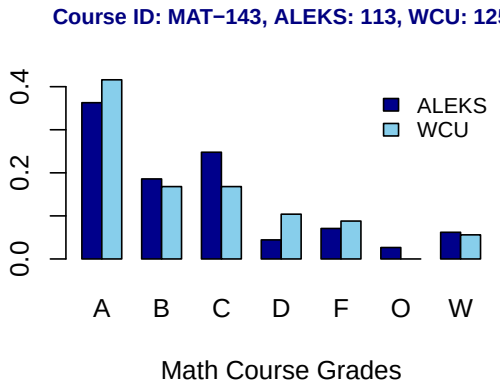


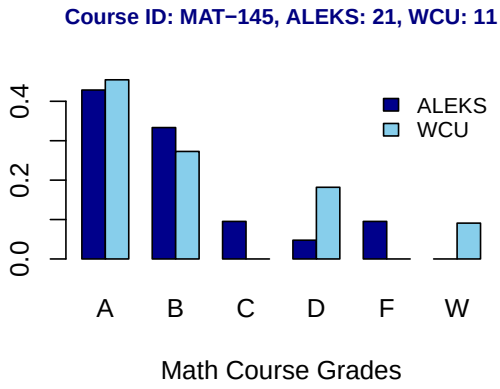

```
par(mfrow=c(1,2))
course.comparison(course.id = "MAT-125")
course.comparison(course.id = "MAT-131")
```



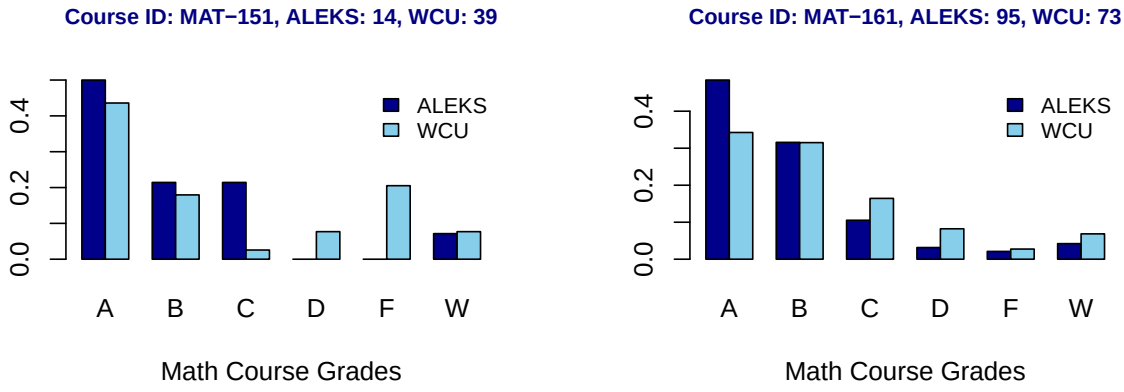



```
par(mfrow=c(1,2))
course.comparison(course.id = "MAT-143")
course.comparison(course.id = "MAT-145")
```





```
par(mfrow=c(1,2))
course.comparison(course.id = "MAT-151")
course.comparison(course.id = "MAT-161")
```




```
par(mfrow=c(1,2))
course.comparison(course.id = "MAT-Q20")
course.comparison(course.id = "MAT-Q30")
```

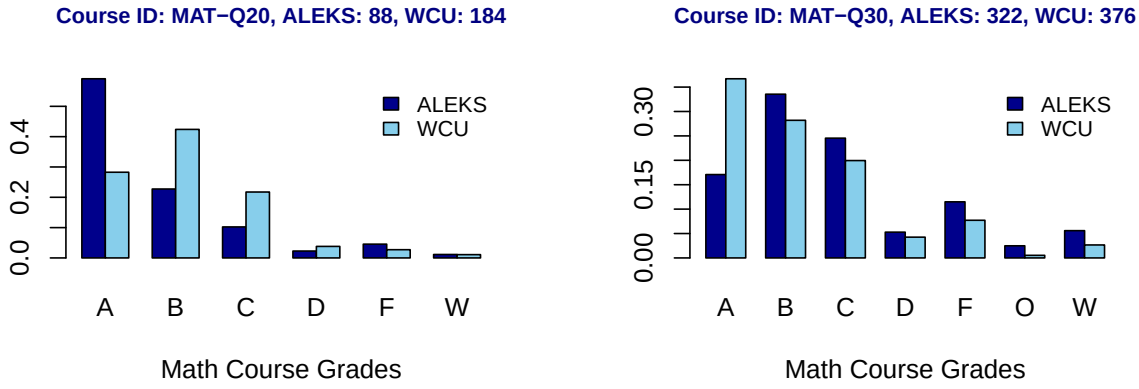


Figure 5: Comparing math grade distributions between ALKES and WCUPT students across all courses

We can observe some patterns from the above grade distributions across the courses. First of all, the ALEKS generated more DFW in MAT-101, MAT-103, MAT-104, MAT-121, MAT-Q30 while WCU's homegrown placement test generated more DFW in the rest of the other courses. Since MAT-101, MAT-103, MAT-121, and MAT-Q30 have multiple parallel sections, the ALEKS seemed to generate more DFWs than WCUPT did.

```
success.data0 = combined.data
  grade0 = combined.data$Grade
  other.id = grade0 %in% c("AU", "NG", "IP", "Z")
  success.data = success.data0[!other.id,]
  grade1 = success.data$Grade
  ##
  D.id = grade1 %in% c("D", "D+", "D-")
  F.id = grade1 %in% c("F")
```

```
W.id = grade1 %in% c("W")
A.id = grade1 %in% c("A", "A+", "A-")
B.id = grade1 %in% c("B", "B+", "B-")
C.id = grade1 %in% c("C", "C+", "C-")
##
success.data$Grade.new[D.id] = "D"
success.data$Grade.new[A.id] = "A"
success.data$Grade.new[B.id] = "B"
success.data$Grade.new[C.id] = "C"
success.data$Grade.new[F.id] = "F"
success.data$Grade.new[W.id] = "W"
success.counts = table(success.data$PlaceType, success.data$Grade.new)
success.rate <- round(100*rbind(ALEKS =success.counts[1,]/sum(success.counts[1,]),
                      WCU = success.counts[2,]/sum(success.counts[2,])),2)
kable(success.rate, caption = "Overall grade distribution by placement test type")
```

Table 5: Overall grade distribution by placement test type

	A	B	C	D	F	W
ALEKS	32.22	31.20	19.48	5.27	7.76	4.08
WCU	32.14	31.68	20.36	5.79	5.95	4.08

The above frequency table reveals a surprising piece of information about the overall grade distribution: students placed in different courses performed almost equally well regardless of which placement test they had taken. The only exception is that the DFW rates in students who took ALEKS are about 1.5% higher than those who the WCUPT.

The following figure also shows the same information that was summarized in the above table.

```
mycols = c("darkblue","skyblue")
barplot(success.rate, main=paste("Overal Grade Distribution \n  ALEKS: ", sum(success.counts[1,]),
                            ", WCU: ", sum(success.counts[2,]), sep = "" ),
      xlab="Math Course Grades",
      col = mycols,
      cex.main = 0.8,
      col.main = "navy",
      beside=TRUE)
 legend(x = "topright", legend = rownames(counts), fill = mycols,
      bty = "n", y.intersp = 1, cex = 0.8)
```

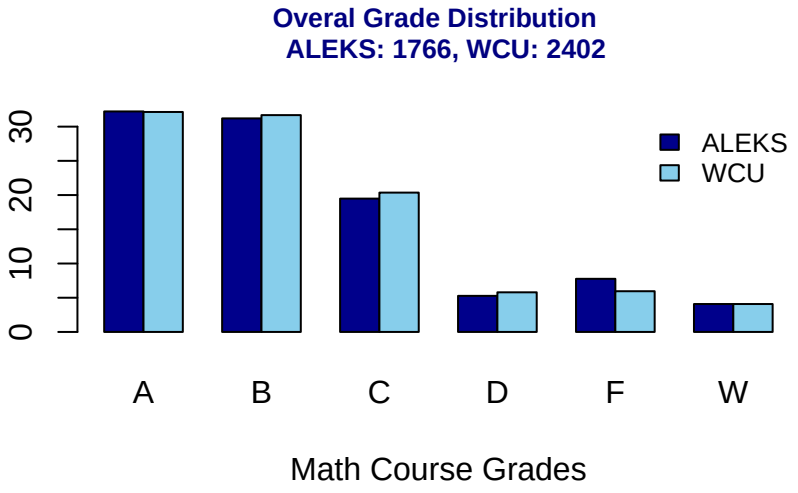


Figure 6: The distributions of math grades of ALEKS and WCUPT students - an overall comparison

Stratification by Instructors

Students who took their "first" math course in Summer 2019, Fall 2019, or Spring 2020 were placed into appropriate courses with placement scores from WCUPT while those who took their "first" math course in Summer 2020 or Fall 2020 were placed into the appropriate courses with ALEKS scores.

In this section, we select several instructors who taught the same courses to both ALEKS and WCUPT students to define a controlled subset of data. We will compare the distribution of math grades earned by both ALEKS and WCUPT students directly to see whether there is a potential difference. If there is a difference between the two distributions, we could link that difference to the type of the placement tests since we have removed as much as possible confounding from the controlled data.

The list of instructors who taught the same course to both cohorts of students are listed below.

```
COURSE    FID
MAT-Q30   7464, 3637, 3287, 2028, 1562
MAT-Q20   1562
MAT-161   2343, 5691
MAT-151   1613
MAT-131   3480
MAT-125   1337
MAT-121   3287, 4587, 5206, 6124, 6232
MAT-113   7957, 8201, 9502
MAT-104   7464
MAT-103   2343, 8220
MAT-101   1331, 6232
```

The next table provides a group comparison of the performance of the math courses taught by the selected instructors listed above.

```
FID0 = c(7464, 3637, 3287, 2028, 1562, 1562, 2343, 5691,
         1613, 3480, 1337, 3287, 4587, 5206, 6124, 6232,
         7957, 8201, 9502, 7464, 2343, 8220, 1331, 6232)
COURSE0 =c("MAT-Q30", "MAT-Q20", "MAT-161", "MAT-151",
           "MAT-131", "MAT-125", "MAT-121", "MAT-113",
```

```
          "MAT-104", "MAT-103", "MAT-101")
FID = combined.data$FID
fid.id = FID%in%FID0
FID.subdata =combined.data[fid.id,]
course =FID.subdata$Course
course.id = course%in%COURSE0
FID.COURSE = FID.subdata[course.id,]
##
  grade0 = FID.COURSE$Grade
  other.id = grade0 %in% c("AU", "NG", "IP", "Z")
  success.data = FID.COURSE[!other.id,]
  grade1 = success.data$Grade
  ##
  D.id = grade1 %in% c("D", "D+", "D-")
  F.id = grade1 %in% c("F")
  W.id = grade1 %in% c("W")
  A.id = grade1 %in% c("A", "A+", "A-")
  B.id = grade1 %in% c("B", "B+", "B-")
  C.id = grade1 %in% c("C", "C+", "C-")
  ##
  #success.data$Grade.new[D.id] = "D"
  success.data$Grade.new[A.id] = "A"
  success.data$Grade.new[B.id] = "B"
  success.data$Grade.new[C.id] = "C"
  success.data$Grade.new[D.id] = "D"
  success.data$Grade.new[F.id] = "F"
  success.data$Grade.new[W.id] = "W"
  success.counts = table(success.data$PlaceType, success.data$Grade.new)
  success.rate <- round(100*rbind(ALEKS =success.counts[1,]/sum(success.counts[1,]),
                       WCU = success.counts[2,]/sum(success.counts[2,])),2)
  kable(success.rate, caption = "The comparison of the grades of the `first`
        math course of students taught by the above-selected instructors")
```

Table 6: The comparison of the grades of the `first` math course of students taught by the above-selected instructors

	A	B	C	D	F	W
ALEKS	35.12	31.17	19.15	5.35	6.38	2.82
WCU	32.67	31.87	19.12	5.93	6.59	3.81

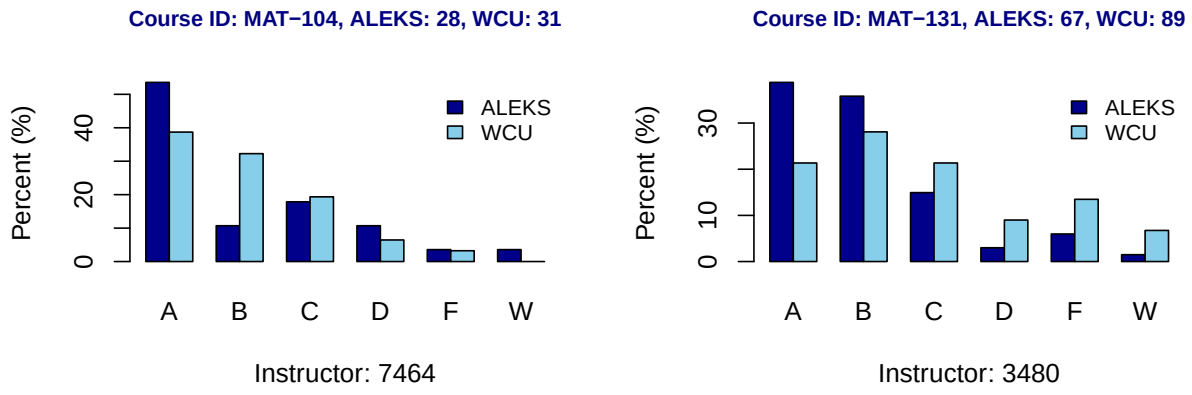
The above summary table of the math grade distribution of the courses taught by this specific group of instructors indicates that the distribution associated with WCUPT students is consistent with the same distribution of the entire study population. But the distribution of math grades of ALEKS students in this cohort is different from that of overall ALEKS students. Particularly, the rates of "F" and "W" grades in ALEKS students are lower than those of the WCUPT students.

```
# R function to create the comparison bar-plots for the select MAT courses
Inst.WCU.ALEKS = function(class.id, inst.id){
  grade0 = FID.COURSE$Grade
  other.id = grade0 %in% c("AU", "NG", "IP", "Z")
  success.data0 = FID.COURSE[!other.id,]
  CID = which(success.data0$Course==class.id)
  success.data = success.data0[CID,]
```

```
  grade1 = success.data$Grade
    ##
  D.id = grade1 %in% c("D", "D+", "D-")
  F.id = grade1 %in% c("F")
  W.id = grade1 %in% c("W")
  A.id = grade1 %in% c("A", "A+", "A-")
  B.id = grade1 %in% c("B", "B+", "B-")
  C.id = grade1 %in% c("C", "C+", "C-")
  ##
  #success.data$Grade.new[D.id] = "D"
  success.data$Grade.new[A.id] = "A"
  success.data$Grade.new[B.id] = "B"
  success.data$Grade.new[C.id] = "C"
  success.data$Grade.new[D.id] = "D"
  success.data$Grade.new[F.id] = "F"
  success.data$Grade.new[W.id] = "W"
  success.counts = table(success.data$PlaceType, success.data$Grade.new)
  success.rate <- round(100*rbind(ALEKS =success.counts[1,]/sum(success.counts[1,]),
                        WCU = success.counts[2,]/sum(success.counts[2,])),2)
  mycols = c("darkblue","skyblue")
  barplot(success.rate, main=paste("Course ID: ",class.id, ", ALEKS: ", sum(success.counts[1,]),
                                ", WCU: ", sum(success.counts[2,]), sep = ""),
        xlab=paste("Instructor: ",inst.id, sep="" ),
        ylab= "Percent (%)",
        col = mycols,
        cex.main = 0.8,
        col.main = "navy",
        beside=TRUE)
   legend(x = "topright", legend = rownames(success.counts), fill = mycols,
        bty = "n", y.intersp = 1, cex = 0.8)
}
```

The next few figures compare the grade distributions of ALEKS and WCUPT students in MAT-151, MAT-131, MAT-125, and MAT-104 taught by the same instructors.

```
par(mfrow=c(1,2))
Inst.WCU.ALEKS(class.id="MAT-104", inst.id = "7464")
Inst.WCU.ALEKS(class.id="MAT-131", inst.id = "3480")
```



```
par(mfrow=c(1,2))
Inst.WCU.ALEKS(class.id="MAT-125", inst.id ="1337")
Inst.WCU.ALEKS(class.id="MAT-151", inst.id ="1613")
```

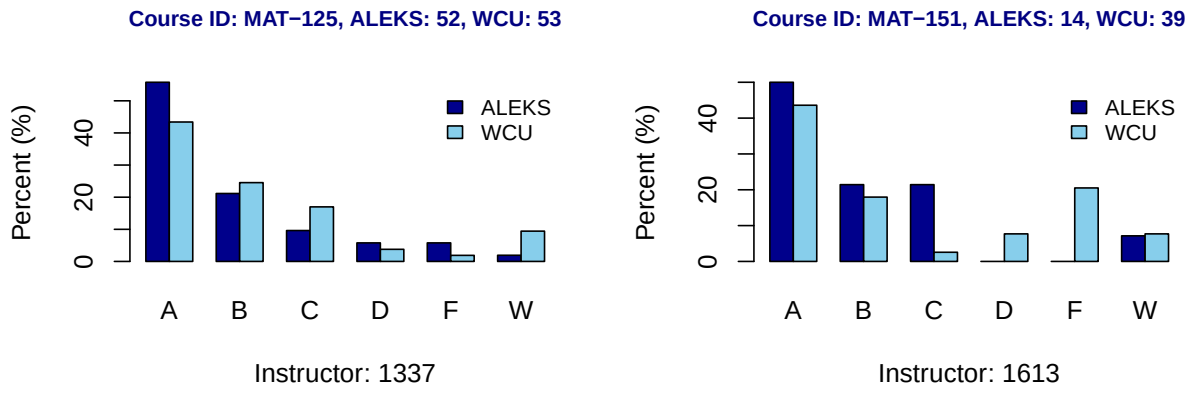
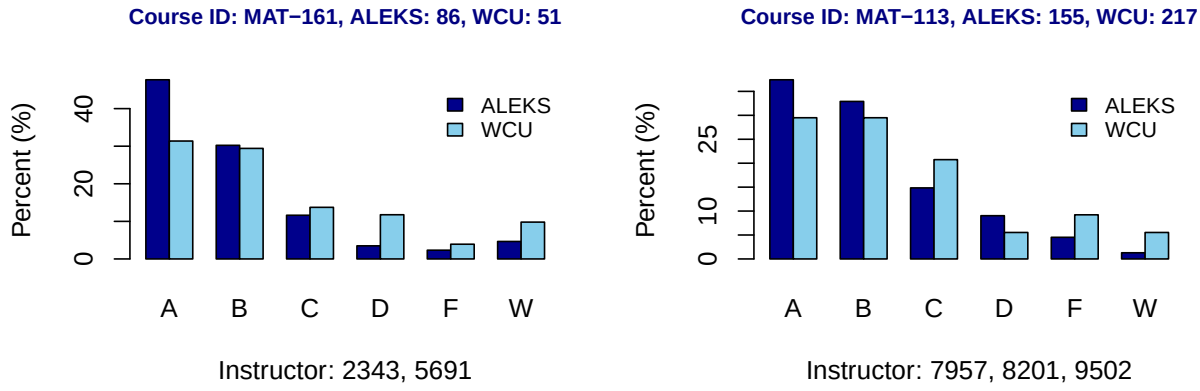


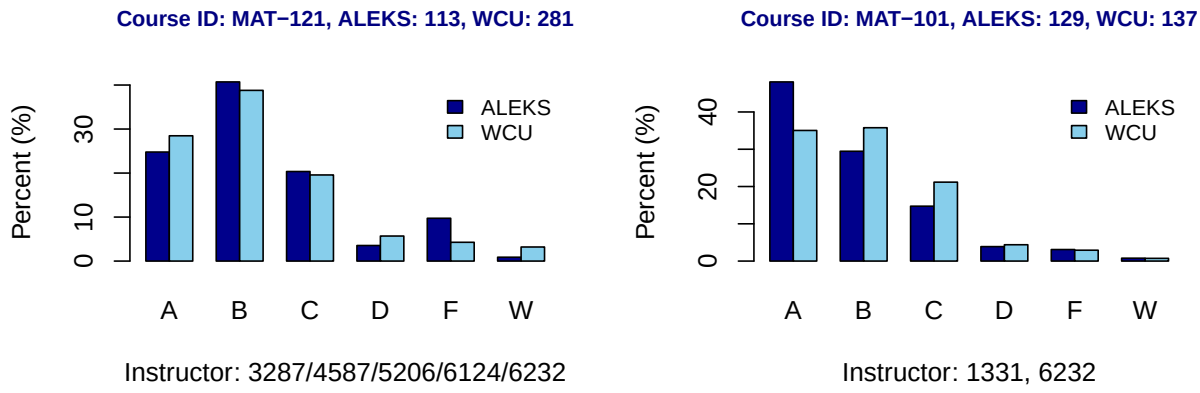
Figure 7: The same course taught by the same instructor.

We can see that each of the above four instructors taught the same course to both ALEKS and WCUPT students and the resulting distributions of the math grades are significantly **different** . The following few bar-plots show the distribution of math grades of courses taught by a group of instructors.

```
par(mfrow=c(1,2))
Inst.WCU.ALEKS(class.id="MAT-161", inst.id ="2343, 5691")
Inst.WCU.ALEKS(class.id="MAT-113", inst.id ="7957, 8201, 9502")
```




```
par(mfrow=c(1,2))
Inst.WCU.ALEKS(class.id="MAT-121", inst.id ="3287/4587/5206/6124/6232")
Inst.WCU.ALEKS(class.id="MAT-101", inst.id ="1331, 6232")
```




```
par(mfrow=c(1,2))
Inst.WCU.ALEKS(class.id="MAT-Q30", inst.id = "7464/3637/3287/2028/1562")
Inst.WCU.ALEKS(class.id="MAT-103", inst.id = "2343, 8220")
```

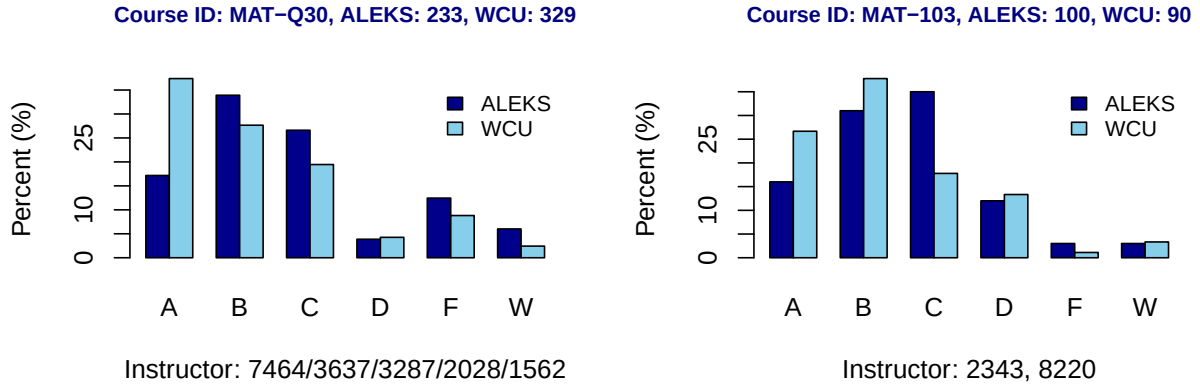


Figure 8: The same course taught by the same group of instructors respectively to both ALEKS and WCUPT students

From the above figures, we can see that the distributions of math grades of ALEKS and WCUPT students are different across all MAT courses. In most of the courses, ALEKS students had more chances to earn an A grade. The percentage of W grade in ACUPT students is higher than that in ALEKS students. These observed patterns are inconsistent with the overall grade distributions for ALEKS and WCUPT students summarized in Table 6 – this phenomenon is the so-called Simpson's paradox.

4.4 Instructors' Grade Distribution

In this subsection, we create a frequency distribution table of grade distributions of all courses taught by individual instructors during this study period. This adds another dimension to look at the confounding factor that impacts the success rate.

Among various grade rates, the DFW rates are of particular interest. We create a probability distribution histogram based on individual instructor's DFW rates. The histogram shows a wide variation in DFW rates. This is a significant observation. Since a high DFW rate could be caused by several potential reasons: instructors' expectations of students or placement test inappropriately placed students' into courses. It is an

area the faculty team can collaborate to work together to reduce the variation of the DFW rates by adjusting expectations for students and tuning the scoring systems of the placement test.

```
course.data0 = combined.data
grade0 = course.data0$Grade
other.id = grade0 %in% c("AU", "NG", "IP", "Z")
course.data = course.data0[!other.id,]
grade1 = course.data$Grade
A.id = grade1 %in% c("A", "A+", "A-")
B.id = grade1 %in% c("B", "B+", "B-")
C.id = grade1 %in% c("C", "C+", "C-")
D.id = grade1 %in% c("D", "D+", "D-","F", "W")
#F.id = grade1 %in% c("F")
#W.id = grade1 %in% c("W")

##
course.data$Grade[A.id] = "A"
course.data$Grade[B.id] = "B"
course.data$Grade[C.id] = "C"
course.data$Grade[D.id] = "DFW"
#course.data$Grade[F.id] = "F"
#course.data$Grade[W.id] = "W"
freq.table = as.matrix(table(course.data$FID, course.data$Grade))
rel.freq = 100*round(prop.table(freq.table, 1),4)
summary.table = cbind(freq.table, rel.freq)
#
col.grade.type = c("A:count",  "B:count", "C:count", "DFW:count",
                   "A: %", "B: %", "C: %", "DFW: %")
row.fid=as.character(1:40)
colnames(summary.table) = col.grade.type
rownames(summary.table) = row.fid
final.summary.table = cbind(Instr.id = 1:40, summary.table)
kable(final.summary.table, caption="Individual instructor's course grade
      distribution: between Fall 2019 and Fall 2020")
```

Table 7: Individual instructor's course grade distribution: between Fall 2019 and Fall 2020

Instr.id	A:count	B:count	C:count	DFW:count	A: %	B: %	C: %	DFW: %
1	8	4	2	1	53.33	26.67	13.33	6.67
2	67	70	34	16	35.83	37.43	18.18	8.56
3	73	40	19	18	48.67	26.67	12.67	12.00
4	9	14	7	4	26.47	41.18	20.59	11.76
5	45	49	32	14	32.14	35.00	22.86	10.00
6	24	10	4	15	45.28	18.87	7.55	28.30
7	25	29	21	20	26.32	30.53	22.11	21.05
8	21	25	24	24	22.34	26.60	25.53	25.53
9	40	49	35	26	26.67	32.67	23.33	17.33
10	12	7	6	1	46.15	26.92	23.08	3.85
11	3	9	2	2	18.75	56.25	12.50	12.50
12	29	35	5	14	34.94	42.17	6.02	16.87
13	23	38	28	26	20.00	33.04	24.35	22.61
14	66	73	40	27	32.04	35.44	19.42	13.11
15	31	59	40	41	18.13	34.50	23.39	23.98
16	25	17	20	16	32.05	21.79	25.64	20.51

Instr.id	A:count	B:count	C:count	DFW:count	A: %	B: %	C: %	DFW: %
17	57	62	47	62	25.00	27.19	20.61	27.19
18	10	18	12	6	21.74	39.13	26.09	13.04
19	43	57	43	28	25.15	33.33	25.15	16.37
20	11	15	11	10	23.40	31.91	23.40	21.28
21	26	35	15	12	29.55	39.77	17.05	13.64
22	19	31	20	7	24.68	40.26	25.97	9.09
23	42	27	18	16	40.78	26.21	17.48	15.53
24	19	17	24	4	29.69	26.56	37.50	6.25
25	39	36	18	11	37.50	34.62	17.31	10.58
26	69	72	43	40	30.80	32.14	19.20	17.86
27	49	70	39	17	28.00	40.00	22.29	9.71
28	51	35	42	21	34.23	23.49	28.19	14.09
29	7	9	8	8	21.88	28.12	25.00	25.00
30	22	21	14	13	31.43	30.00	20.00	18.57
31	47	41	19	25	35.61	31.06	14.39	18.94
32	130	38	15	17	65.00	19.00	7.50	8.50
33	30	39	26	20	26.09	33.91	22.61	17.39
34	32	34	17	34	27.35	29.06	14.53	29.06
35	9	3	4	5	42.86	14.29	19.05	23.81
36	48	36	32	19	35.56	26.67	23.70	14.07
37	1	3	5	7	6.25	18.75	31.25	43.75
38	40	40	16	23	33.61	33.61	13.45	19.33
39	24	14	13	10	39.34	22.95	21.31	16.39
40	15	31	13	2	24.59	50.82	21.31	3.28

```
DFW = final.summary.table[,9]
boundary = seq(min(DFW), max(DFW), length = 10)
lab = as.character(round(boundary,1))
hist(DFW,breaks = boundary, probability = TRUE, axes = FALSE,
     ylab = "percentage (%)", xlab = "DFW Rates (%)",
     border = "navy", col = "skyblue")
axis(1, at = boundary, labels = lab)
axis(2)
abline(h=c(0,0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05),
       col="darkseagreen3", lty =3)
```

5 Modeling "Success Rates"

In this subsection, we use a logistic regression model to explore the association between the performance in the "first" math classes and the types of placement tests as well as the placement scores.

5.1 Design of Experiment

To obtain a data set for building an inferential model, we choose a subset of instructors who taught the same courses to both ALEKS and WCUPT students to reduce the potential confounding effect. A subset of data contains 2457 students were selected in this section for building an association model to identify the potential factors that impact the "success rate". In this subset, 14.6% of students received DFW grades.

We define "success" to be of earning a grade of C or better and "failure" to be of earning a DFW grade. We use the subset of data used in stratification analysis based on instructors. The following model will explore the factors such as placement scores, test types, instructors, and courses. The primary variables of interest are placement test type and scores

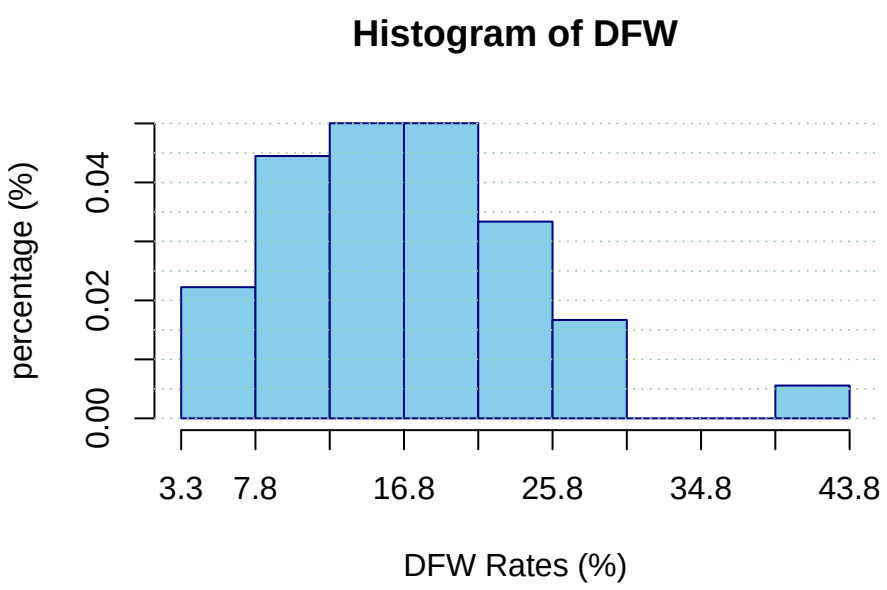


Figure 9: The distribution of individual instructor's DFW rates

5.2 Logistic Regression for Modeling "Success Rate"

Since the two placement tests have different scoring systems, the scoring conversion did not result in comparable distribution between the scores of ALEKS and WCUPT. This implies that the placement test and scores have natural cross-over interaction effects. This interaction will be included in the final model. The way of placing students into an appropriate course based on the score and the placement test also potentially impacts students' performance in the class.

However, the actual grade distribution of Instructors' courses may be dependent on the overall performance of the class - this may offset the potential difference in the discriminatory power of the two placement tests. To detect the meaningful and significant interactions, we initially include all possible interactions and perform a standard variable selection procedure to identify the final model with the summary statistics in the following table.

```
logit.data = FID.COURSE[!FID.COURSE$Grade%in%c("IP", "NG", "Z"),]
pass.id = logit.data$Grade %in% c("A", "A-","B+", "B", "B-", "C+", "C", "C-")
fail.id = logit.data$Grade %in% c("D", "D+", "D-", "F", "W")

logit.data$PlaceScore[which(logit.data$PlaceScore==0)] = 1
success = rep("pass", length(logit.data$Grade))
success[fail.id] = "fail"
logit.data$sucess = success
logit.data = as.data.frame(logit.data[-which(logit.data$FID=="1613"),])
```

```
model.01 = glm(as.factor(sucess) ~ PlaceType + PlaceType*factor(PlaceScore) + Course +
                 factor(FID), family = binomial(link = logit), data = logit.data)
##
model.coef.stats = summary(model.01)$coef
odds.ratio = exp(coef(model.01))
out.stats = cbind(round(model.coef.stats,4), odds.ratio = round(odds.ratio,3))
kable(out.stats,caption = "Summary Statistics in Odds Ratios")
```

Table 8: Summary Statistics in Odds Ratios

	Estimate	Std. Error	z value	Pr(>|z|)	odds.ratio
(Intercept)	1.5228	0.4037	3.7723	0.0002	4.585
PlaceTypeWCU	0.7723	0.2820	2.7380	0.0062	2.165
factor(PlaceScore)2	0.8844	0.3077	2.8742	0.0041	2.421
factor(PlaceScore)3	1.2017	0.3580	3.3569	0.0008	3.326
factor(PlaceScore)4	1.4031	0.3943	3.5583	0.0004	4.068
factor(PlaceScore)5	3.1732	0.5469	5.8016	0.0000	23.883
CourseMAT-103	-1.5287	0.7746	-1.9735	0.0484	0.217
CourseMAT-104	-0.3265	0.6194	-0.5271	0.5981	0.721
CourseMAT-113	-0.3277	0.4466	-0.7338	0.4631	0.721
CourseMAT-121	-0.3464	0.3630	-0.9544	0.3399	0.707
CourseMAT-125	-1.3067	0.5726	-2.2819	0.0225	0.271
CourseMAT-131	-0.5088	0.5352	-0.9507	0.3418	0.601
CourseMAT-161	-2.3953	0.7126	-3.3612	0.0008	0.091
CourseMAT-Q20	1.2671	0.6254	2.0261	0.0428	3.550
CourseMAT-Q30	0.0607	0.4115	0.1475	0.8827	1.063
factor(FID)1337	-0.1126	0.5559	-0.2025	0.8395	0.894
factor(FID)1562	0.5695	0.5962	0.9552	0.3395	1.767
factor(FID)2028	-1.1681	0.4366	-2.6756	0.0075	0.311
factor(FID)2343	0.2166	0.7249	0.2988	0.7651	1.242
factor(FID)3287	-0.4724	0.3855	-1.2254	0.2204	0.624
factor(FID)3480	-1.1191	0.5274	-2.1221	0.0338	0.327
factor(FID)3637	-1.2770	0.3930	-3.2493	0.0012	0.279
factor(FID)4587	-1.3199	0.5579	-2.3660	0.0180	0.267
factor(FID)5206	-0.4555	0.5154	-0.8837	0.3769	0.634
factor(FID)5691	-0.2314	0.6235	-0.3711	0.7105	0.793
factor(FID)6124	-0.1304	0.5288	-0.2467	0.8052	0.878
factor(FID)6232	-0.3256	0.4056	-0.8028	0.4221	0.722
factor(FID)7464	-0.3989	0.4684	-0.8518	0.3943	0.671
factor(FID)7957	-0.7668	0.5517	-1.3900	0.1645	0.464
factor(FID)8201	-0.1740	0.4620	-0.3767	0.7064	0.840
factor(FID)8220	0.6319	0.8244	0.7665	0.4434	1.881
factor(FID)9502	-0.7894	0.4786	-1.6494	0.0991	0.454
PlaceTypeWCU:factor(PlaceScore)2	-0.7090	0.3669	-1.9321	0.0533	0.492
PlaceTypeWCU:factor(PlaceScore)3	-1.1187	0.3542	-3.1582	0.0016	0.327
PlaceTypeWCU:factor(PlaceScore)4	-0.4387	0.5050	-0.8687	0.3850	0.645
PlaceTypeWCU:factor(PlaceScore)5	-1.4780	0.5199	-2.8426	0.0045	0.228

The above summarizes the factors that impact the odds of passing a class with a grade of C or better. There are several observations worth mentioning.

- *The main effect of the placement test*: The WCUPT is better than the ALEKS in terms of the odds of passing a course with a grade of C or better in general (p-value = 0.0062, OR = 2.165).
- *The main effect of the placement test scores*: in general, a higher placement score results in better odds of passing a class with a grade of C or better regardless of the placement type. Interestingly, there is no discriminatory effect between scores 3 and 4. The following interaction effect between the placement test and the scores of the placement test.
- *The interaction effect between the placement test type and the placement scores*: Since the interaction effect is significant, we need to look into the actual odds of "success" of the WCUPT test is dependent on the individual placement test score.

- *Heterogeneity of the odds of success across the course*: In this logistic regression model, the baseline course is MAT-101. The odds ratios (last column of the table) represent the odds of "success" of the corresponding courses divided by the odds of "success" of the baseline course (MAT-101). If the odds ratio of a course is equal to 1, then there is no difference between the course and the baseline course in terms of the odds of success. If the odds ratio of a course is less than 1, the odds of success of the course is less than that of the baseline course, and vice versa. We can see that the odds of success of different courses are different among the selected courses.
- *The heterogeneity of the odds of success among the selected instructors*: The baseline instructor ID is 1331. The odds ratios of success of taking a course with an instructor versus instructor 1331 are given in the last column. The signs of the Estimate of the regression coefficients are log odds ratios (in column 2). The statistical significance levels are given in the 4th column of the table. An instructor is associated with a positive estimate implying that taking a course with the instructor tended to have a higher odds of success than that of taking the course with instructor 1331. A negative estimate of the log-odds can be interpreted in a similar way.
- **A closer look at the interaction effect between the placement type and the test score**: Since the placement type, test score, and their interaction are the major contributors to the odds of success. We fit a logistic regression model with only placement type and the test score to probe into the interaction effect of the two variables on the odds of success.

```
model.02 = glm(as.factor(sucess) ~ PlaceType + PlaceType*factor(PlaceScore),
               family = binomial(link = logit), data = logit.data)
##
model.coef.stats.2 = summary(model.02)$coef
odds.ratio.2 = exp(coef(model.02))
out.stats.2 = cbind(round(model.coef.stats.2,4), odds.ratio.2 = round(odds.ratio.2,3))
kable(out.stats.2,caption = "Summary Statistics in Odds Ratios")
```

Table 9: Summary Statistics in Odds Ratios

	Estimate	Std. Error	z value	Pr(>|z|)	odds.ratio.2
(Intercept)	1.3408	0.1674	8.0078	0.0000	3.822
PlaceTypeWCU	0.4703	0.2495	1.8851	0.0594	1.601
factor(PlaceScore)2	0.3174	0.2472	1.2840	0.1991	1.374
factor(PlaceScore)3	0.4127	0.2395	1.7231	0.0849	1.511
factor(PlaceScore)4	0.6122	0.2918	2.0982	0.0359	1.844
factor(PlaceScore)5	1.1941	0.3333	3.5831	0.0003	3.301
PlaceTypeWCU:factor(PlaceScore)2	-0.3458	0.3409	-1.0144	0.3104	0.708
PlaceTypeWCU:factor(PlaceScore)3	-0.8944	0.3239	-2.7617	0.0058	0.409
PlaceTypeWCU:factor(PlaceScore)4	-0.1721	0.4795	-0.3589	0.7197	0.842
PlaceTypeWCU:factor(PlaceScore)5	-0.8403	0.4742	-1.7719	0.0764	0.432

```
y10 = 1.3408 + 0
y20 = 1.3408 + 0.3174
y30 = 1.3408 + 0.4127
y40 = 1.3408 + 0.6122
y50 = 1.3408 + 1.1941
y0.avg = mean(c(y10, y20, y30, y40, y40))
##
y01 = 1.3408 + 0.4703 + 0
y02 = 1.3408 + 0.4703 - 0.3458 + 0.3174
y03 = 1.3408 + 0.4703 - 0.8944 + 0.4127
y04 = 1.3408 + 0.4703 - 0.1721 + 0.6122
```

```
y05 = 1.3408 + 0.4703 - 0.8403 + 1.1941
y1.avg = mean(c(y01, y02, y03, y04, y05))
##
plot(1:2, c(y10, y01), type="l", lwd = 2, col = 1, xlab= "Placement type",
     ylab = "log odds of success", ylim=c(1, 4.5), axes= FALSE,
     main = "Interaction Between Placement Type and Placement Score",
     cex.main = 0.8, col.main = "navy")
axis(1, pos=1, at=c(1,2), label=c("ALEKS", "WCUPT"))
axis(2)
points(1:2,c(y10, y01), pch=20, col=1 )
#
lines(1:2, c(y20, y02),  type="l", lwd=2, col=2 )
points(1:2,c(y20, y02), pch=20, col=2 )
#
lines(1:2, c(y30, y03),  type="l", lwd=2, col=3 )
points(1:2,c(y30, y03), pch=20, col=3 )
#
lines(1:2, c(y40, y04),  type="l", lwd=2, col=4 )
points(1:2,c(y40, y04), pch=20, col=4 )
#
lines(1:2, c(y50, y05),  type="l", lwd=2, col=5 )
points(1:2,c(y50, y05), pch=20, col=5 )
##
legend("topright", c("score =1", "score =2", "score =3", "score =4", "score =5"),
       lwd=rep(2,5), col=1:5, cex=0.8, bty="n")
```

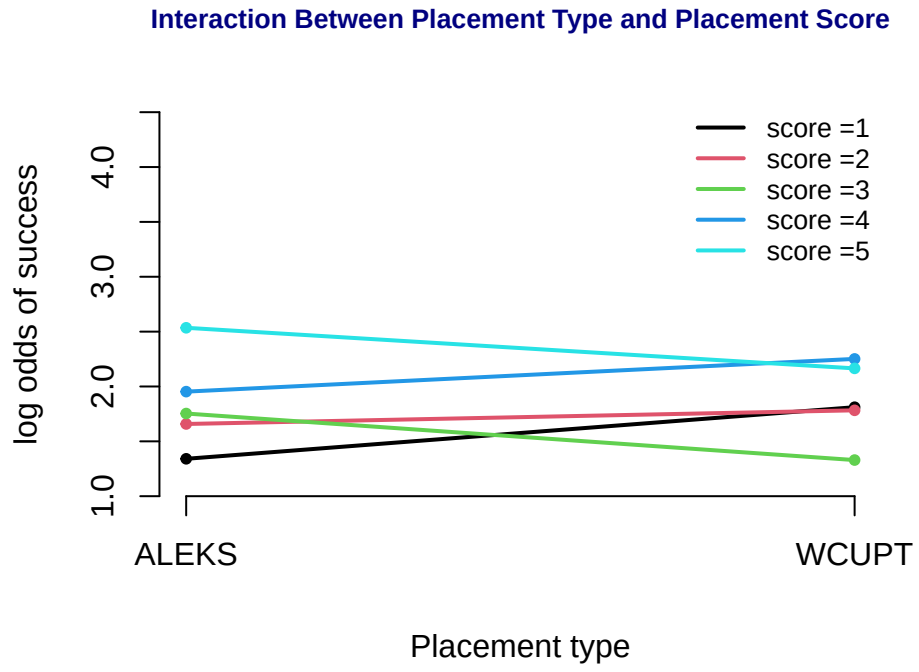


Figure 10: Interaction effect between the placement type and the placement scores.

```
##
#lines(1:2, c(y0.avg, y1.avg),  type="l", lwd=2, col=6, lty=2 )
#points(1:2,c(y0.avg, y1.avg), pch=20, col=6 )
```

We can see from Figure 10 that there are complicated cross-over interaction effects between the placement type and the placement scores. These patterns can be used to tune the scoring systems in the future to reduce the DFW rates.

6 Summary and Discussions

We have conducted both descriptive and inferential analyses from different perspectives to compare the two placement tests.

The direct comparison of the distributions indicates that the score distributions of two placement tests are significantly different (Pearson chi-square test yields a p-value ≈ 0). This means that the same score from different placement tests does not imply the same level of mathematical proficiency. Therefore, the placement methods should reflect this discrepancy between the two distributions of the corresponding placement test scores.

We also compare the students' performance in the "first" mathematics course after they took the placement test. The direct comparison of the overall grade distributions of ALEKS and WCUPT students indicates that there is no significant difference between the two placement tests (Table 5). However, the distributions of math grades associated with ALKES and WCUPT at each course level are significantly different. This seemingly inconsistent phenomenon is called Simpson Paradox.

To reduce the variation due to the potential difference of the grade assignments of instructors who did not teach the same courses to BOTH ALEKS and WCUPT students, we select a group of instructors who taught the same course to BOTH ALEKS and WCUPT students. The resulting overall frequency distributions of the math grades of ALEKS and WCUPT students are NOT significantly different from this "controlled" data. However, the distributions of math grades in the same course taught by the instructor(s) are also different. The results from this controlled data could be considered as the systematic difference between the two placement tests.

Using the aforementioned controlled data, we performed a logistic regression analysis to assess the association between the students' success and the type of the placement test adjusted by other factors including placement test scores, course types, and instructors, and the necessary interaction effects between the factors. The result indicates no significant difference between the two placement tests, but the first-order interaction analysis shows a marginal benefit of ALEKS in terms of the odds of success.

Since this is not a random sample taken from a population, we can still justify the validity of the logistic regression and the output inferential statistics in the following:

- Variation of grade assignment at the instructor level.
- Variation of placing students into courses due to the variation of the two different scoring systems.

Both of the above variations bring uncertainty to the "success" of passing a course with a grade of C or better.

The analysis of DFW rates shows a wide variation of the DFW rates from each instructor. This is the area the instructors to work with to reduce the DFW rates by (1) adjusting the placement scoring system and the way of placing students into appropriate courses; (2) adjusting the course grade distribution to "more fair" grades that consistent with other colleagues'.

References

1. Dorans, N. (2008). The Practice of comparing scores on different tests. R&D Connections, Educational Testing Services. https://www.ets.org/Media/Research/pdf/RD_Connections6.pdf (last accessed on 1/25/2021)
2. Flaherty, C. (2020). Grading for a pandemic, Insight Higher Educations, https://www.insidehighered.com/news/2020/04/23/how-lenient-or-not-should-professors-be-students-right-now (last accessed on 1/25/2021)
3. Gonzalez T, de la Rubia MA, Hincz KP, Comas-Lopez M, Subirats L, Fort S, et al. (2020) Influence of COVID-19 confinement on students' performance in higher education. PLoS ONE 15(10): e0239490. doi:10.1371/journal.pone.0239490

Appendix: SAS Code

```
/*************************************************************
  Creating An Analytic Data Set for Placement Analysis
  Author: C. Peng
  Date: 1/26/2021
  Last update: 3/16/2021
  Tasks: 1. Clean both Math grade and placement score data
            sets.
         2. Merge two clean data sets
         3. Perform some exploratory analysis to inspect
            potential irregular features of the variables.
         4. Some frequency analysis for data quality assuance
*************************************************************/

LIBNAME assess "C:\Assessment";

/*******************************
******  First Math Courses   ****
*******************************/
/** drop the the empty columns **/
DATA assess.FinalMathScore;
SET assess.Mathscore;
DROP F9 F10 F11;
RUN;

/** Checking data types **/
PROC CONTENTS DATA = assess.FinalMathScore;
RUN;

/** Whether a student took the placement but have not taken a math course **/
DATA assess.StudentWithFirstMath;
SET assess.FinalMathScore;
FirstMath = "YES";
IF FacultyID = " " THEN FirstMath = "NO";
IF FirstMath = "YES";     /* have taken first math course */
IF PlacementFlag = "Y";   /* have attented a placement test */
IF ASPFlag = " " THEN ASPFlag = "T"; /* Missing value are new transfer  */
RUN;

PROC SORT DATA = assess.StudentWithFirstMath;
BY StudentID AdmitTerm;
```

```
RUN;

/** Create some auxiliary variables to check whether there
    were students who took two math courses in different
    semesters.    **/
DATA assess.MultipleFirstMath;
SET assess.StudentWithFirstMath;
Constant = 1;
BY StudentID AdmitTerm;
RETAIN count total;
IF FIRST.StudentID THEN
   DO;
      COUNT = 0;
      Total = 0;
   END;
   COUNT = COUNT + 1;
   total = total + Constant;
   avg = total/count;
   OUTPUT;
RUN;

/** students with at least 1 math course after the placement test **/
DATA assess.multipleMath1;
SET assess.MultipleFirstMath;
 AdmitTerm0 = AdmitTerm;
 StudentID0 = StudentID;
 Course0 = Course;
 CourseGrade0 = CourseGrade;
 total0 = total;
KEEP AdmitTerm0 StudentID0 Course0 CourseGrade0 StudentID total0
AdmitType ASPFlag PlacementFlag FirstMath FacultyID;
RUN;


PROC SORT DATA = assess.multipleMath1;
BY StudentID;
RUN;


/** students with at least 2 math course after the placement test **/
DATA assess.multipleMath2;
SET assess.MultipleFirstMath;
IF total > 1;
KEEP AdmitTerm StudentID Course CourseGrade FacultyID total;
RUN;

/** Check whether two courses were taken in the same semester     **/
/** Some students took their "first math" course in the same semester **/
DATA assess.TwoCourseSameSemester;
MERGE assess.multipleMath1 assess.multipleMath2;
BY StudentID;
IF TOTAL NE .;
IF AdmitTerm NE AdmitTerm0;
RUN;
```

```
/** Final Data Set for the Math Courses **/
DATA assess.FinalMathCourseData;
SET assess.MultipleFirstMath;
RUN;



/***********************************
*****     Placement Tests         ***
***********************************/

/** make a copy of clean data set **/
DATA assess.FinalPlacementScore;
SET assess.PlacementScore;
RUN;

/** Exclude students who took the placement test before the 11/17/2018 **/
DATA assess.PlacementType;
SET assess.FinalPlacementScore;
* YRMO = put(PlaceTestDate,yymmd7.);
IF PlaceTestDate > '17nov2018'd;
RUN;

/** ALEKS vs WCU: ALEKS implemented on 4/16/2020 **/
DATA assess.PlaceTypeFlag;
SET assess.PlacementType;
PlaceType = "WCU  ";
IF PlaceTestDate > '16apr2020'd THEN PlaceType = "ALEKS";
RUN;

/** Some students took the placement test multiple times.
    Keep only the highest score**/
PROC SORT DATA = assess.PlaceTypeFlag;
BY StudentID PlaceScore;
RUN;

DATA assess.PlaceAttempts;
SET assess.PlaceTypeFlag;
BY StudentID AdmitTerm;
RETAIN count;
IF FIRST.StudentID THEN
   DO;
      COUNT = 0;
   END;
   COUNT = COUNT + 1;
   OUTPUT;
RUN;

/** Final Placement Data Set **/
DATA assess.FinalPlaceAttemptScore;
SET assess.PlaceAttempts;
BY StudentID PlaceScore;
IF LAST.StudentID;
```

```
RUN;


PROC SORT DATA = assess.FinalPlaceAttemptScore;
BY StudentID;
RUN;


/*****************************************
*****  Combining the two data sets     ******
*****************************************/

/** Combining placement scores and math course grades  **/
DATA assess.MathPlaceCombined;
MERGE assess.multipleMath1 assess.FinalPlaceAttemptScore;
BY StudentID;
IF total0 NE .;
IF PlaceScore NE .;
IF AdmitType = "F" THEN AdmitType = "FIRST ";
IF AdmitType = "T" THEN AdmitType = "TRANSF";
DROP ASPFlag AdmitTerm0 total0 FirstMath PlacementFlag;
RUN;

/** Sorting data **/
PROC SORT DATA = assess.MathPlaceCombined;
BY FacultyID AdmitTerm Course0;
RUN;

/***** Some frequency analysis ****/

PROC SORT DATA = assess.MathPlaceCombined OUT= assess.MathPlaceCombined_SortScore;
BY PlaceScore Course0;
RUN;

PROC FREQ DATA = assess.MathPlaceCombined;
Table PlaceType*PlaceScore*Course0/nopercent nocol;
RUN;
```