The Effect of Usable Mathematic Knowledge of Teachers on Student Learning

Rao Xiong

Spring 2022

Introduction

This analysis is part of the research study entitled “The Classroom Video Analysis -Common Core Mathematics (CVA-M) Instrument: A Development and Validation Study to Measure Usable, Content-focused Mathematics Knowledge for Teaching” by LAUSD schools, leads by Dr. Nicole Kersting, University of Arizona. We believe teachers’ usable mathematics knowledge in teaching is different from the related mathematics knowledge they have due the teaching experience and environment. The CVA-M instrument tried to measure this usable knowledge teachers have with classroom video analysis following by questions related to different magnitude of teaching. We wish that this measurement can positively predict students’ math performance.

Hypothesis: the higher level of Usable Mathematics Knowledge of teachers is, the higher Math scores of students.

This analysis focus on the topic of fraction in 4th and 5th grade. The independent variable, usable mathematics knowledge of teachers (CVAM score), is the sum score of teachers’ responses of 15 questions in CVA-M Instrument, coded by the research team with qualified internal reliability. While each question is coded with 0, 1, and 2, the sum score ranges from 0 to 30. The Math scores of students, are measured using Math scores from Smarter Balanced Assessment Consortium (SBAC) test. It’s a standardized test, range between 2000 to 3000, with higher score representing better student performance. The dependent variable is the math score in 2019, and the math score in 2018 is the covariate variable.

The data original has 5202 students in 202 classrooms. In order to better visualizing the data and the model, I random sampled a subset of 10 classrooms with 275 students.

To address Hypothesis, I will run a Hierarchical Linear Modeling (HLM) with math scores in 2019 included as the dependent variable, and teachers’ CVAM scores included as our class level (level 2) independent variable. We will enter math scores in 2018 as student level (level 1) covariates in the model. A p-value of less than .05 will indicate a significant effect of the CVAM score on our dependent variable, math scores in 2019.

```
fproject<-read_csv("data/final_project_data.csv")
```

```
## Rows: 5202 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (5): ID, Course_Name, X2018_Overall_Scale, X2019_Overall_Scale, CVA
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
set.seed(2022)
fproject_subsample_ID<-sample( unique(fproject$ID),10)

fproject_sub<-fproject%>%
  filter(ID %in% fproject_subsample_ID)%>%
  rename(Math_2019 = X2019_Overall_Scale,
         Math_2018 = X2018_Overall_Scale,
         CVAM = CVA,
         Class = ID)%>%
  mutate(Class = factor(Class))
```

Checking Assumptions

Class Level Variance

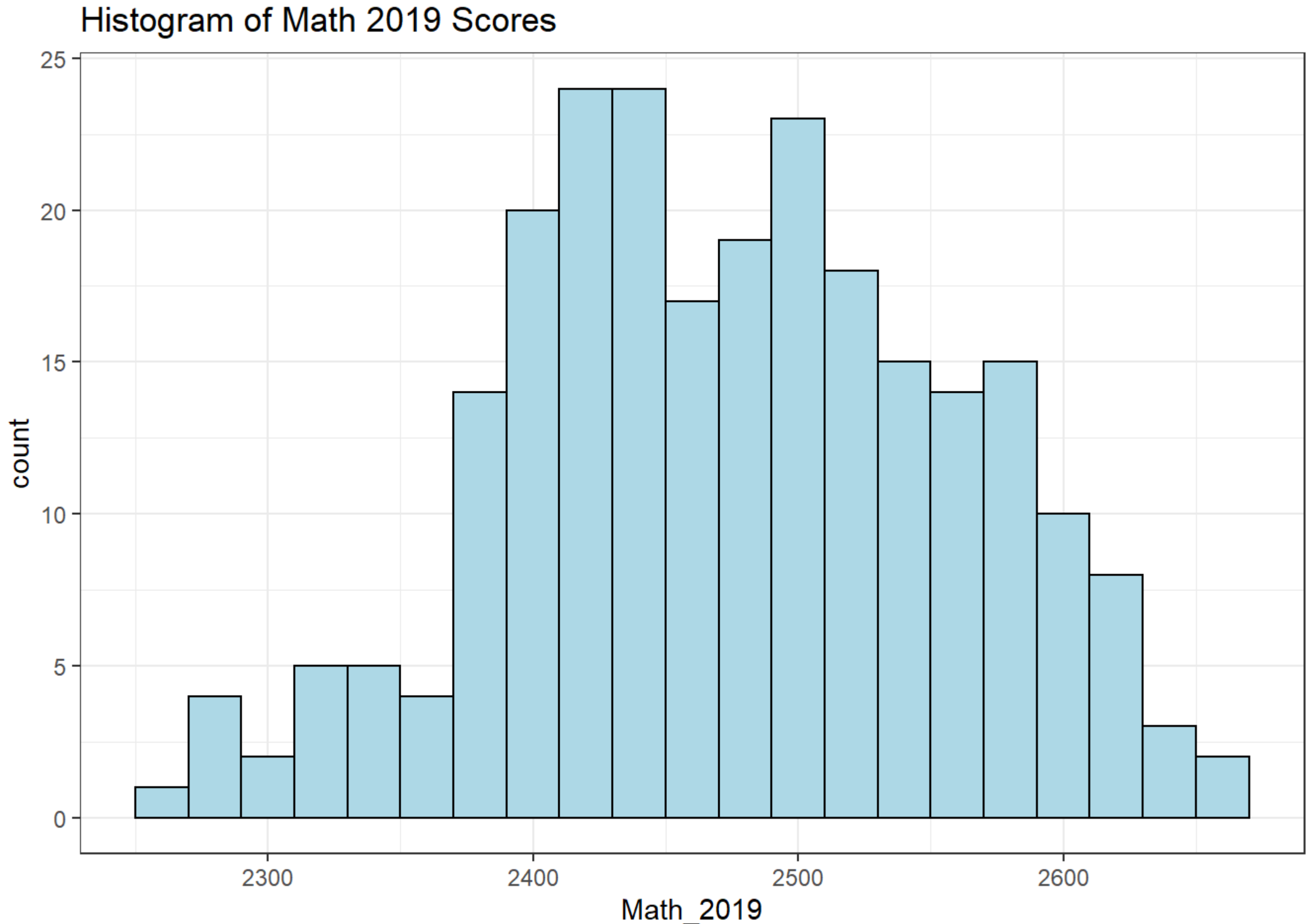
```
ucm <- lme(fixed = Math_2019 ~ 1,
        random = ~ 1 | Class,
        data = fproject_sub)

VarCorr(ucm)
```

```
## Class = pdLogChol(1)
##             Variance StdDev
## (Intercept) 1833.269 42.81669
## Residual    5137.751 71.67811
```

Intraclass Correlation Coefficient: 1833.269 /(1833.269 + 5137.751) = 0.2630
Since 26.30% of variance in Math_2019 scores are explained by the difference between classed, which exceeds the cut-off value of 10%, we should use Hierarchical Linear Modeling.

Normality & Outliers

```
fproject_sub%>%
  ggplot(aes(x = Math_2019))+
  geom_histogram(binwidth = 20,color = "black",fill = "lightblue")+
  labs(title = "Histogram of Math 2019 Scores")+
  theme_bw()
```

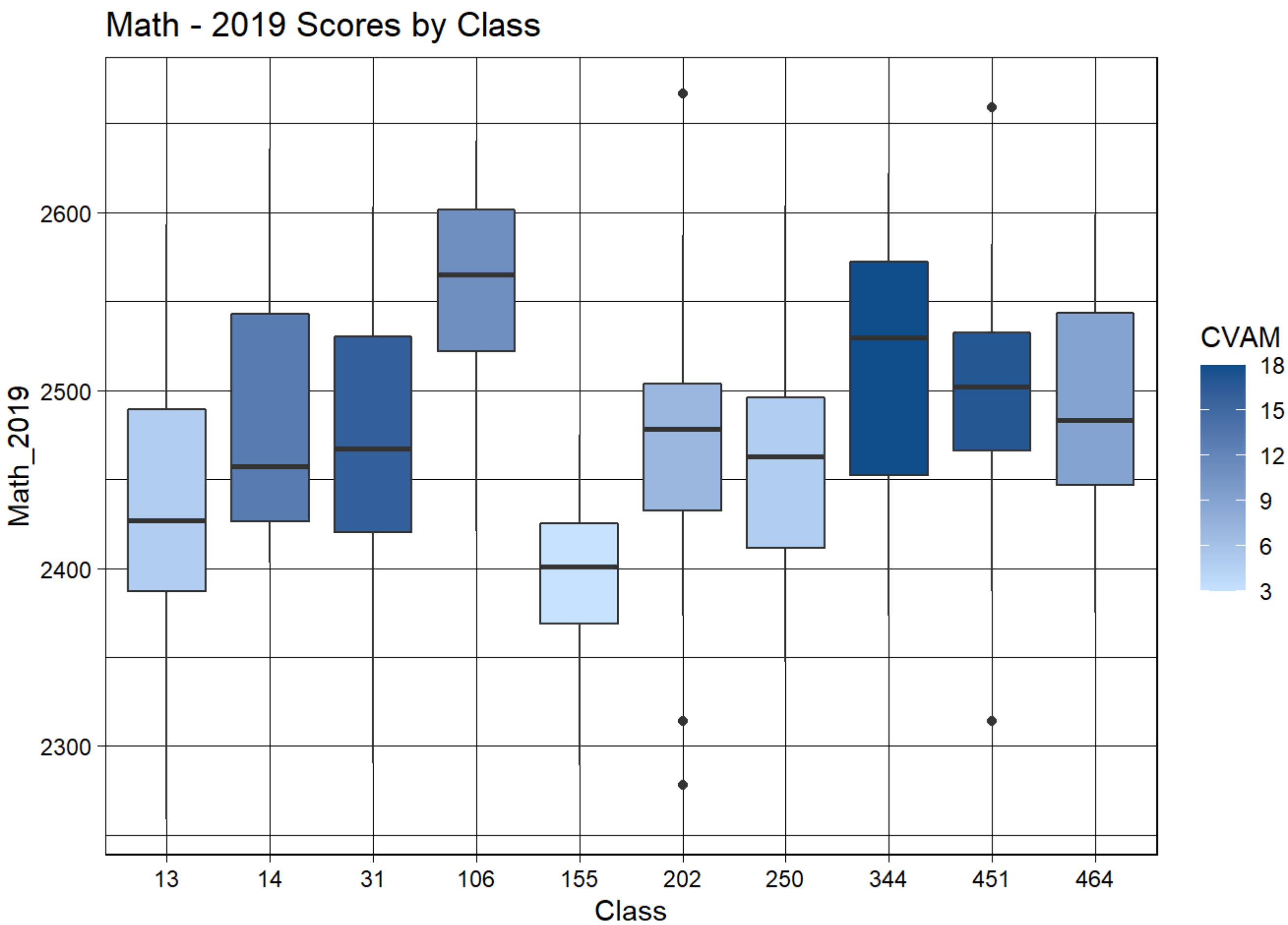


The histogram of Math 2019 Scores at overall student level show that the dependent variable is normally distributed.

```
fproject_sub%>%count(Class)
```

```
## # A tibble: 10 x 2
##    Class     n
##    <fct> <int>
##  1 13       27
##  2 14       27
##  3 31       26
##  4 106      27
##  5 155      25
##  6 202      26
##  7 250      28
##  8 344      28
##  9 451      19
## 10 464      14
```

```
fproject_sub%>%
  ggplot(aes(x = Class, y = Math_2019,fill = CVAM))+
  geom_boxplot()+
  scale_fill_continuous(low = "slategray1",high = "dodgerblue4")+
  labs(title = "Math - 2019 Scores by Class ")+
  theme_linedraw()
```

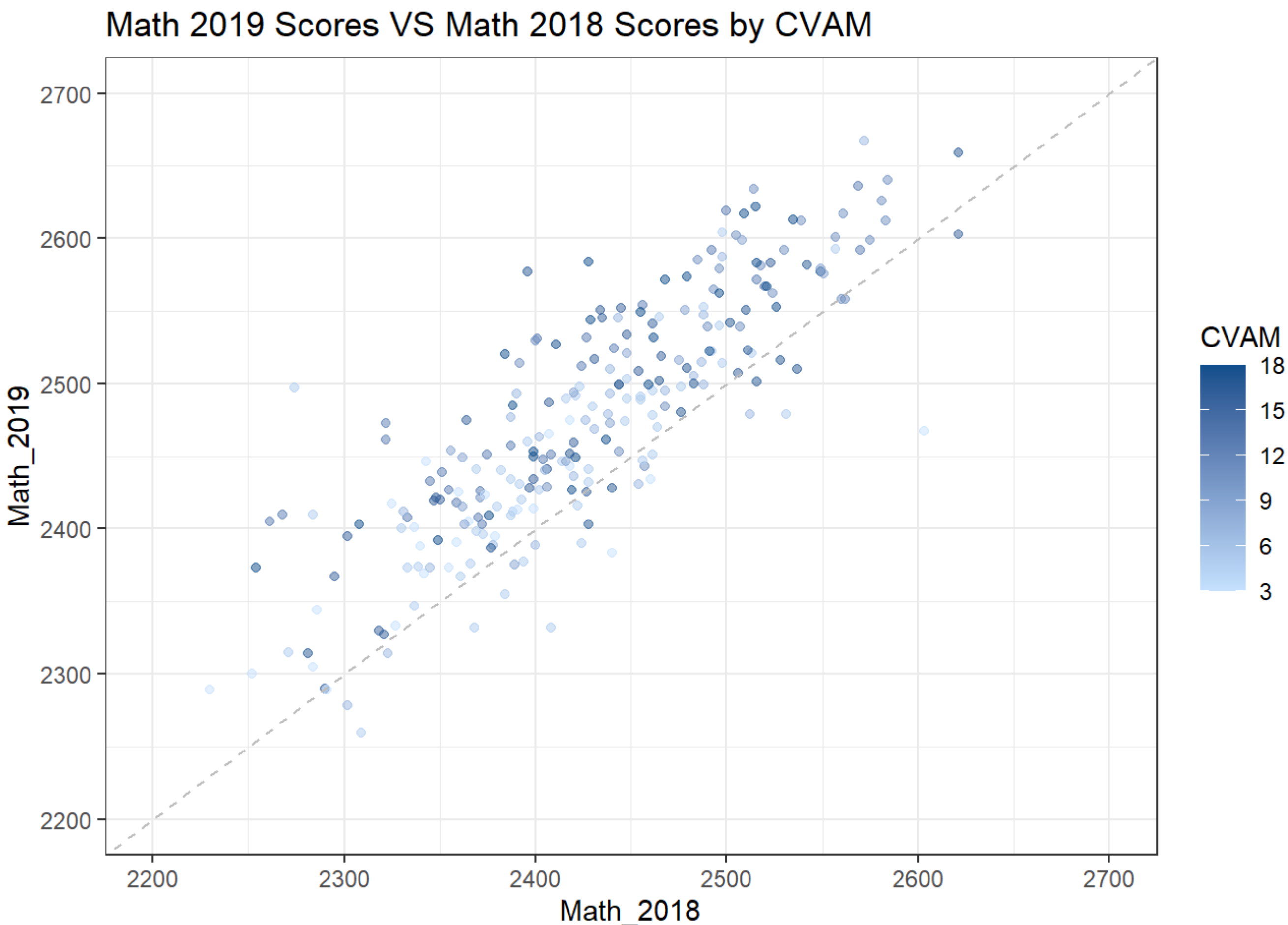


The boxplot of the Math 2019 scores for each classes shows that most of the classes have normally distributed Math 2019 scores. Although Class 202 and 451 have several outliers, I will still keep them in the data since the class sizes are relatively small.

Exploring Data

Student Level

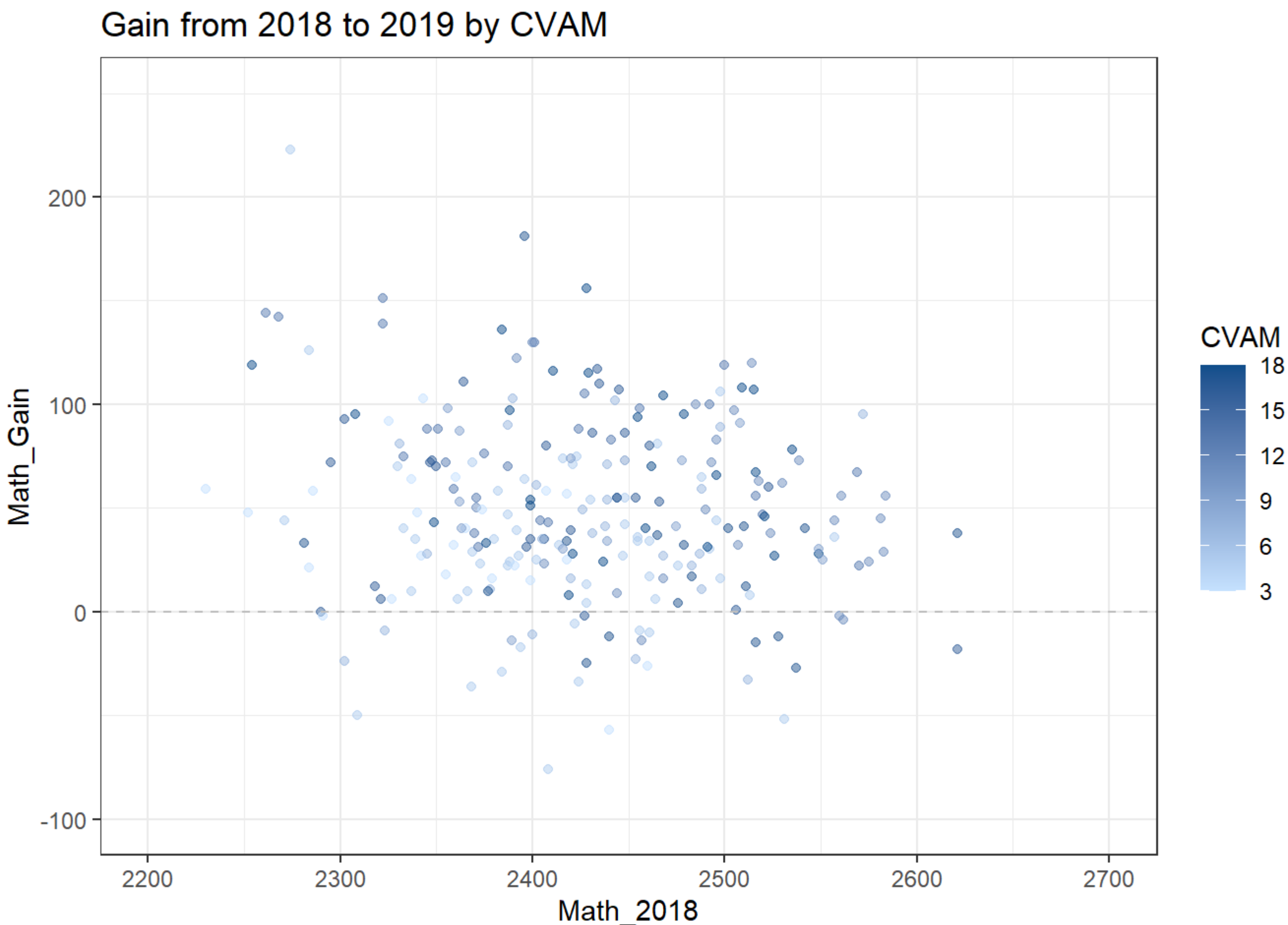
```
fproject_sub%>%
  ggplot(aes(x = Math_2018,y = Math_2019,color = CVAM))+
  geom_point(alpha=0.5)+
  scale_x_continuous(limits = c(2200,2700))+
  scale_y_continuous(limits = c(2200,2700))+
  scale_color_continuous(low = "slategray1",high = "dodgerblue4")+
  geom_abline(intercept = 0, slope = 1, color = "grey",linetype = "dashed")+
  labs(title= "Math 2019 Scores VS Math 2018 Scores by CVAM")+
  theme_bw()
```



This plot shows that the more of the points with bigger distance above the dotted gray line, which has math improvement of 0, are in darker color, and more of the points close or below the dotted line of 0 improvement, are in lighter color.

```
fproject_sub%>%
  mutate(Math_Gain = Math_2019-Math_2018)%>%
  ggplot(aes(x = Math_2018,y = Math_Gain,color = CVAM))+
  geom_point(alpha=0.5)+
  scale_x_continuous(limits = c(2200,2700))+
  scale_y_continuous(limits = c(-100,250))+
  scale_color_continuous(low = "slategray1",high = "dodgerblue4")+
  geom_hline(yintercept = 0, color = "grey",linetype = "dashed")+
  labs(title= "Gain from 2018 to 2019 by CVAM")+
  theme_bw()
```

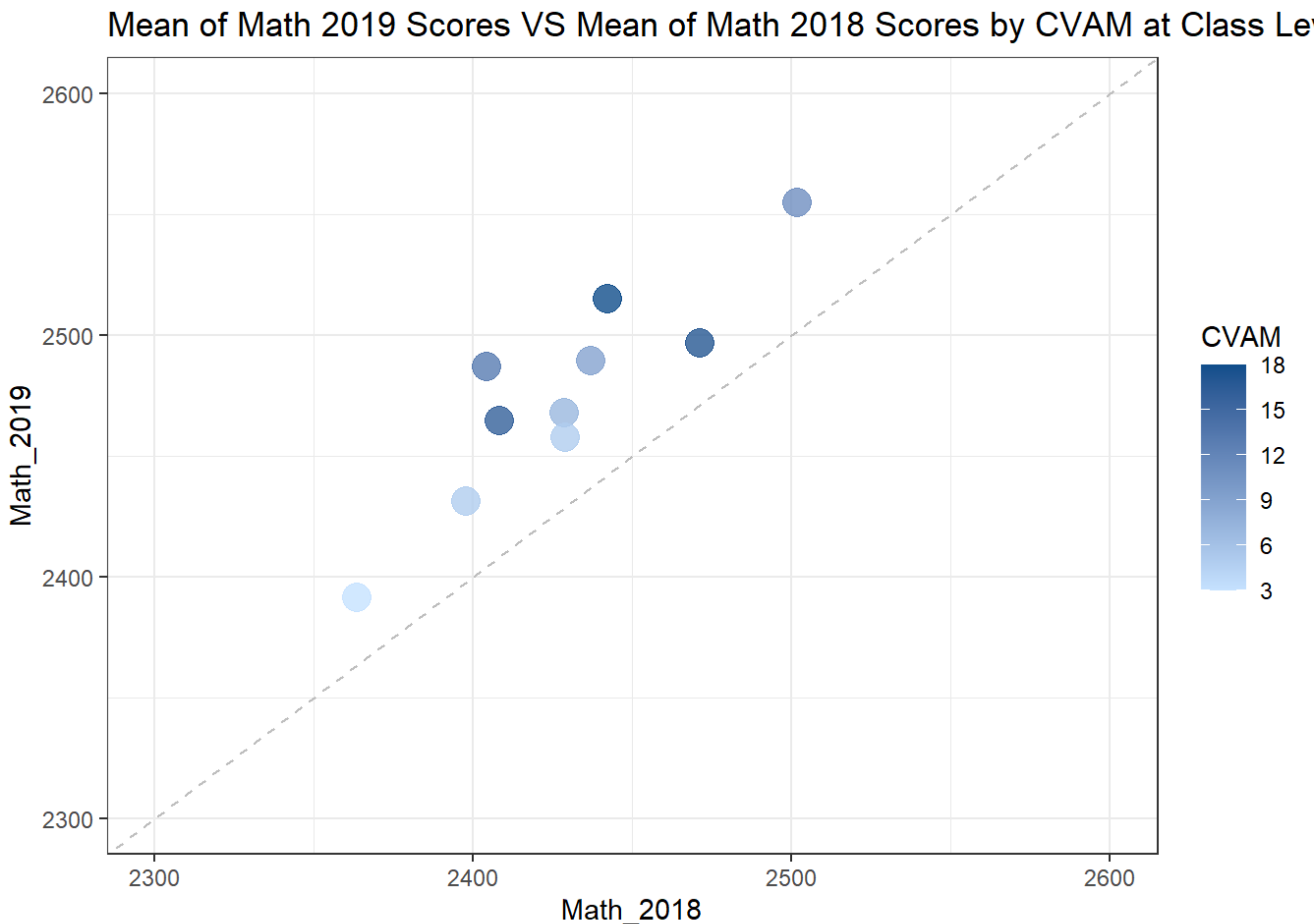
```
## Warning: Removed 1 rows containing missing values (geom_point).
```



In this rotated plot with horizontal axis presenting the gain of math scores from 2018 to 2019, it is much clearer that points with greater gain have higher CVAM score than the points with lower or negative gain. Therefore, at student level, the higher the teacher's CVAM score is, the greater the improvement student has in math.

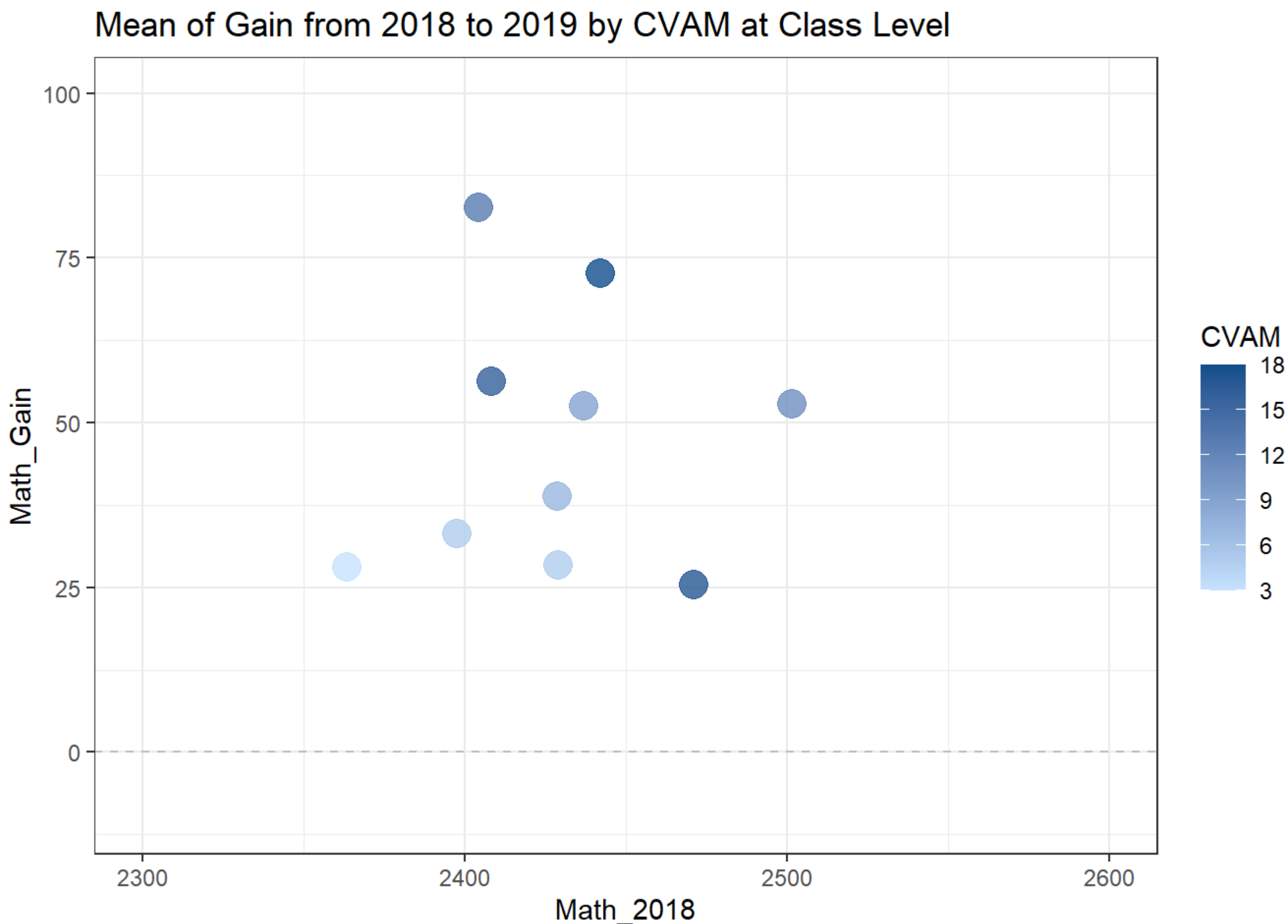
Class Level

```
fproject_sub%>%
  group_by(Class)%>%
  summarize(Math_2018 = mean(Math_2018),
         Math_2019 = mean(Math_2019),
         CVAM = mean(CVAM))%>%
  ggplot(aes(x = Math_2018,y = Math_2019,color = CVAM))+
  geom_point(alpha=0.8,size = 5)+
  scale_x_continuous(limits = c(2300,2600))+
  scale_y_continuous(limits = c(2300,2600))+
  scale_color_continuous(low = "slategray1",high = "dodgerblue4")+
  geom_abline(intercept = 0, slope = 1, color = "grey",linetype = "dashed")+
  labs(title= "Mean of Math 2019 Scores VS Mean of Math 2018 Scores by CVAM at Class Level")+
  theme_bw()
```



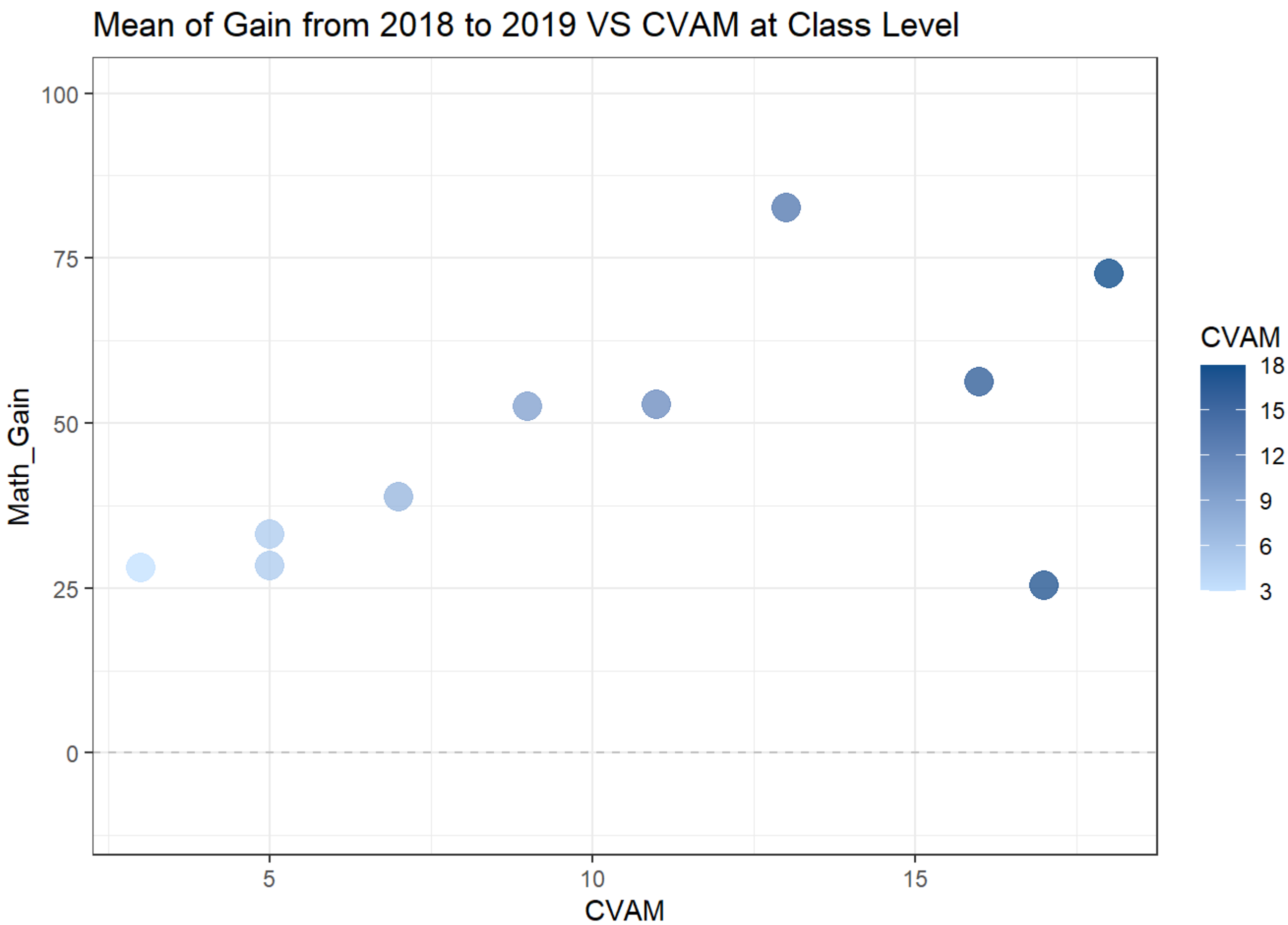
This plot shows the mean of math scores for each class in 2018 and 2019. The dotted gray line represents math improvement of 0.

```
fproject_sub%>%
  mutate(Math_Gain = Math_2019-Math_2018)%>%
  group_by(Class)%>%
  summarize(Math_2018 = mean(Math_2018),
         Math_Gain = mean(Math_Gain),
         CVAM = mean(CVAM))%>%
  ggplot(aes(x = Math_2018,y = Math_Gain,color = CVAM))+
  geom_point(alpha=0.8,size = 5)+
  scale_x_continuous(limits = c(2300,2600))+
  scale_y_continuous(limits = c(-10,100))+
  scale_color_continuous(low = "slategray1",high = "dodgerblue4")+
  geom_hline(yintercept = 0, color = "grey",linetype = "dashed")+
  labs(title= "Mean of Gain from 2018 to 2019 by CVAM at Class Level")+
  theme_bw()
```



This rotated plot presents the mean of gain of math scores from 2018 to 2019 of each classes. It shows that classes with greater mean of gain have higher CVAM score on average.

```
fproject_sub%>%
  mutate(Math_Gain = Math_2019-Math_2018)%>%
  group_by(Class)%>%
  summarize(Math_2018 = mean(Math_2018),
         Math_Gain = mean(Math_Gain),
         CVAM = mean(CVAM))%>%
  ggplot(aes(x = CVAM,y = Math_Gain,color = CVAM))+
  geom_point(alpha=0.8,size = 5)+
  scale_x_continuous()+
  scale_y_continuous(limits = c(-10,100))+
  scale_color_continuous(low = "slategray1",high = "dodgerblue4")+
  geom_hline(yintercept = 0, color = "grey",linetype = "dashed")+
  labs(title= "Mean of Gain from 2018 to 2019 VS CVAM at Class Level")+
  theme_bw()
```



This plot shows the relationship between mean of gain in math for each class and teachers' CVAM scores directly. Besides the point at lower right, which shows the class with mean gain of about 25 and teacher's CVAM score of about 17, rest of the plot shows a positive relationship between the improvement in math score and teachers' CVAM score.

HLM Model

```
mod <- lme(fixed = Math_2019 ~ 1 + Math_2018+CVAM,
        random = ~ 1| Class,
        data = fproject_sub,
        control = lmeControl(opt="optim"))

summary(mod)
```

```
## Linear mixed-effects model fit by REML
##   Data: fproject_sub
##        AIC      BIC    logLik
##   2550.042 2567.528 -1270.021
##
## Random effects:
##  Formula: ~1 | Class
##         (Intercept) Residual
## StdDev:    14.41384 40.65475
##
## Fixed effects:  Math_2019 ~ 1 + Math_2018 + CVAM
##                Value Std.Error  DF   t-value p-value
## (Intercept) 387.9088  88.89409 236  4.363719  0.0000
## Math_2018     0.8486   0.03681 236 23.056031  0.0000
## CVAM          2.5950   1.02114   8  2.541310  0.0346
##  Correlation:
##           (Intr) M_2018
## Math_2018 -0.991
## CVAM       0.000 -0.118
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -3.201753136 -0.641047878 -0.008177402  0.636046029  4.181363138
##
## Number of Observations: 247
## Number of Groups: 10
```

The result of the HLM shows that the coefficient of CVAM is 2.595, which means that for every 1 point increase in teachers' CVAM score, the predicted mean of math 2019 scores at class level increase 2.59 points, controlling for math 2018 scores. However, comparing to the range of about 100 points among the mean of 2019 score at class level, 2.59 point increase is very small. Therefore, although the p-value ($0.0346<0.5$) shows that the effect of CVAM is statistically significant, the real effect size is tiny.