Multi-level Model
Bias in PERSUADE

```r
library(sjPlot)
library(lme4)
library(car) # level names in contrasts
library(stargazer) # LaTeX tables
library(emmeans)
library(performance) # ICC
library(tidyverse)
library(forcats)

dfp <- read.csv("../data/persuade-predictions-long.csv",
                na.strings = c("", " ", "NA")) %>%
  select(-c(full_text, X, assignment, source_text)) %>%
  mutate_if(is.character, as.factor) %>%

  # drop American Indian/Alaskan Native group because there are < 200 samples
  filter(race_ethnicity != "American Indian/Alaskan Native") %>%

  # remove unused levels (American Indian/Alaskan Native)
  droplevels() %>%

  # simplify level and variable names
  mutate(
    Transformer=fct_recode(model,
                   Black="baa_pred",
                   Control="control_pred",
                   Hispanic="hl_pred",
                   White="w_pred"),
    race=fct_recode(race_ethnicity,
                     asian="Asian/Pacific Islander",
                     black="Black/African American",
```

```r
                    hisp="Hispanic/Latino",
                    other="Two or more races/Other",
                    white="White")
  ) %>%

# calculate root mean squared error
mutate(rmse=sqrt(squared_error)) %>%

# calculate percent error
mutate(percent_error = (pred_error/holistic_score_adjudicated)*100) %>%

# collapse NaNs into negative level of binary factors
mutate(
  is_ell=fct_collapse(addNA(ell), Yes = "Yes", No = c("No", NA)),
  is_disadvantaged=fct_collapse(addNA(economically_disadvantaged),
                                Yes = "Economically disadvantaged",
                                No = c("Not economically disadvantaged",
                                       NA)
                                ),
  has_disability=fct_collapse(addNA(student_disability_status),
                              Yes = "Identified as having disability",
                              No = c("Not identified as having disability",
                                     NA)
                              )
  ) %>%

# set the first level, which will be used as "reference"
mutate(
  is_disadvantaged=fct_relevel(is_disadvantaged, "No"),
  has_disability=fct_relevel(has_disability, "No"),
  Transformer=fct_relevel(Transformer, "Control")
) %>%

mutate(
  Direction = case_when(
    pred_error >= 0 ~ "Positive",
    pred_error < 0 ~ "Negative"
    ),
  Direction = factor(Direction, levels=c("Positive","Negative")
                  )
)
```

```
# configure contrasts for race and source
options(decorate.contr.Sum = c("", ""))
contrasts(dfp$race) = contr.Sum(levels(dfp$race))
contrasts(dfp$source) = contr.Sum(levels(dfp$source))

summary(dfp)
```

```
     essay_id      holistic_score_1 holistic_score_2
2021000047:    4   Min.   :1.000    Min.   :1.000
2021000136:    4   1st Qu.:2.000    1st Qu.:2.000
2021000241:    4   Median :3.000    Median :3.000
2021000462:    4   Mean   :3.364    Mean   :3.388
2021000730:    4   3rd Qu.:4.000    3rd Qu.:4.000
2021000764:    4   Max.   :6.000    Max.   :6.000
(Other)   :19156
holistic_score_adjudicated              source
Min.   :1.000              Florida        :2436
1st Qu.:3.000              Georgia Virtual: 876
Median :3.000              Indiana        :6384
Mean   :3.405             NCES            :3120
3rd Qu.:4.000              Virginia       :6364
Max.   :6.000


                          prompt_name                  task      gender
Distance learning                 :1872   Independent   :10360   F:9612
Summer projects                   :1640   Text dependent: 8820   M:9568
Facial action coding system       :1576
Mandatory extracurricular activities:1532
Exploring Venus                   :1448
Driverless cars                   :1384
(Other)                           :9728
    grade         ell                    race_ethnicity
Min.   : 6.000   No  :16408   Asian/Pacific Islander :3836
1st Qu.: 8.000   Yes : 1812   Black/African American :3836
Median :10.000   NA's:  960   Hispanic/Latino        :3836
Mean   : 9.244               Two or more races/Other:3836
3rd Qu.:10.000               White                  :3836
Max.   :12.000


                economically_disadvantaged
```

```
Economically disadvantaged     :7660
Not economically disadvantaged:8120
NA's                          :3400




                    student_disability_status          model
Identified as having disability    : 2300     baa_pred    :4795
Not identified as having disability:16004     control_pred:4795
NA's                               :  876     hl_pred     :4795
                                              w_pred      :4795




  prediction       pred_error        squared_error       Transformer
Min.   :0.795   Min.   :-2.2323   Min.   : 0.00000   Control :4795
1st Qu.:2.709   1st Qu.:-0.2752   1st Qu.: 0.03495   Black   :4795
Median :3.453   Median : 0.1179   Median : 0.14918   Hispanic:4795
Mean   :3.527   Mean   : 0.1212   Mean   : 0.36447   White   :4795
3rd Qu.:4.285   3rd Qu.: 0.4887   3rd Qu.: 0.42920
Max.   :5.884   Max.   : 3.4119   Max.   :11.64123

   race          rmse           percent_error      is_ell      is_disadvantaged
asian:3836   Min.   :0.00003   Min.   :-49.861   No :17368   No :11520
black:3836   1st Qu.:0.18696   1st Qu.: -6.937   Yes: 1812   Yes: 7660
hisp :3836   Median :0.38623   Median :  3.677
other:3836   Mean   :0.47062   Mean   :  9.335
white:3836   3rd Qu.:0.65514   3rd Qu.: 16.232
             Max.   :3.41193   Max.   :341.192

has_disability    Direction
No :16880      Positive:11184
Yes: 2300      Negative: 7996
```

Summary statistics

```
dfp %>%
  filter(Transformer == "Control") %>%
  summarize(
    Score = mean(holistic_score_adjudicated),
    N = n(),
    ELLs = sum(as.numeric(is_ell)),
    Disability = sum(as.numeric(has_disability)),
    Disadvantaged = sum(as.numeric(is_disadvantaged)),
    .by=c(race_ethnicity)
    ) %>%
  arrange(race_ethnicity)
```

```
          race_ethnicity    Score   N ELLs Disability Disadvantaged
1  Asian/Pacific Islander 3.910323 959 1117       1018          1229
2  Black/African American 3.103233 959  997       1086          1464
3         Hispanic/Latino 3.104275 959 1193       1068          1524
4 Two or more races/Other 3.453597 959  970       1104          1344
5                   White 3.455683 959  971       1094          1149
```

Model from Research Question 1

Just adding in transformer model as a predictor and a random intercept for essay ID. This model has poor fit, with many useless parameters. Prompt_name is no longer useful as a grouping variable, so we can drop it.

```
mod.rq1 <- lmer(pred_error
                ~ trf
                + gender
                + race*is_ell
                + is_disadvantaged*has_disability
                + is_disadvantaged*is_ell
                + has_disability*is_ell
                + (1|prompt_name)
                + (1|essay_id),
                data=dfp
                )
tab_model(mod.rq1, show.aic=TRUE, show.reflvl=TRUE, prefix.labels="varname", p.adjust = "B
```

Save Table

```r
coefs <- summary(fit)$coefficients
cat(rownames(coefs), sep="\", \"")

labels <- c("(Intercept)",
            "Transformer[Black]",
            "Transformer[Hispanic]",
            "Transformer[White]",
            "Black/African American",
            "Hispanic/Latino",
            "Two or more races/Other",
            "White",
            "English language learner",
            "Identified as having disability",
            "Economically disadvantaged",
            "Transformer[Black]:Black/African American",
            "Transformer[Hispanic]:Black/African American",
            "Transformer[White]:Black/African American",
            "Transformer[Black]:Hispanic/Latino",
            "Transformer[Hispanic]:Hispanic/Latino",
            "Transformer[White]:Hispanic/Latino",
            "Transformer[Black]:Two or more races/Other",
            "Transformer[Hispanic]:Two or more races/Other",
            "Transformer[White]:Two or more races/Other",
            "Transformer[Black]:White",
            "Transformer[Hispanic]:White",
            "Transformer[White]:White",
            "Transformer[Black]:English language learner",
            "Transformer[Hispanic]:English language learner",
            "Transformer[White]:English language learner",
            "Transformer[Black]:Identified as having disability",
            "Transformer[Hispanic]:Identified as having disability",
            "Transformer[White]:Identified as having disability",
            "Transformer[Black]:Economically disadvantaged",
            "Transformer[Hispanic]:Economically disadvantaged",
            "Transformer[White]:Economically disadvantaged")

tab_model(fit,
          title = "Prediction Error Regressed on Demographic Variables",
          dv.labels = "Prediction Error",
```

```
        show.aic=TRUE,
        show.reflvl=TRUE,
        prefix.labels="varname",
        p.adjust = "BH",
        pred.labels=labels,
        file = "../results/RQ2.html")
```

Tear-down

We would be left with very little in the model.

```
fit.min <- lmer(pred_error
          ~ trf
        # + race # increases AIC
        # + source # increases AIC
        # + is_disadvantaged # increases AIC
        + is_ell
        + has_disability
        # + gender # increases AIC
        # + race*is_disadvantaged # increases AIC
        # + race*is_ell # increases AIC
        # + race*has_disability # increases AIC
        # + race*source # increases AIC
        # + race*gender # increases AIC
        # + race*grade # increases AIC
        # + is_disadvantaged*has_disability # increases AIC
        # + is_disadvantaged*is_ell # increases AIC
        # + has_disability*is_ell # increases AIC
        # + gender*has_disability # increases AIC
        # + gender*is_disadvantaged # increases AIC
        # + gender*is_ell # small decrease (-1) to AIC
        + (1 | essay_id),
        data=dfp
        )

tab_model(fit.min, show.aic=TRUE, show.reflvl=TRUE, prefix.labels="varname", p.adjust = "B
```

Build-up: Interactions

While race did not help model fit on its own, its interaction with transformer model substantially improves model fit. The same pattern holds for interactions between transformer model and other demographic variables.

```r
fit <- lmer(pred_error
            ~ Transformer
            + Transformer*race_ethnicity
            + Transformer*is_ell
            + Transformer*has_disability
            + Transformer*is_disadvantaged
            + (1 | essay_id),
            REML = TRUE,
            data=dfp
            )

# tab_model(fit, show.aic=TRUE, show.reflvl=TRUE, prefix.labels="varname", p.adjust = "BH"
```

Deviance Tests

These require non-restricted ML estimations.

```r
fit.ml <- lmer(pred_error
            ~ Transformer
            + Transformer*race
            # + Transformer*is_ell
            + Transformer*has_disability
            + Transformer*is_disadvantaged
            + (1 | essay_id),
            REML = FALSE,
            data=dfp
            )

fit.ell <- lmer(pred_error
            ~ Transformer
            + Transformer*race
            + Transformer*is_ell
            + Transformer*has_disability
            + Transformer*is_disadvantaged
```

```
              + (1 | essay_id),
              REML = FALSE,
              data=dfp
              )

1-pchisq(deviance(fit.ml) - deviance(fit.ell), df=2)
```

[1] 0

RMSE

RMSE is more interpretable. Possibly a separate research question?

```
fit.rmse <- lmer(rmse
              ~ Transformer
              + Transformer*race_ethnicity
              + Transformer*is_ell
              # + Transformer*has_disability
              + Transformer*is_disadvantaged
              + (1 | essay_id),
              REML = TRUE,
              data=dfp
              )

tab_model(fit.rmse, show.aic=TRUE, show.reflvl=TRUE, prefix.labels="varname", p.adjust = "
```

Interaction Plots

```
# c("Asian/Pacific Islander",
#    "Black/African American",
#              "Hispanic/Latino",
#              "Two or more races/Other",
#              "White")

emmip(fit, race_ethnicity ~ Transformer,
      mode="asymp",
      type = "scale",
      style = "factor",
```
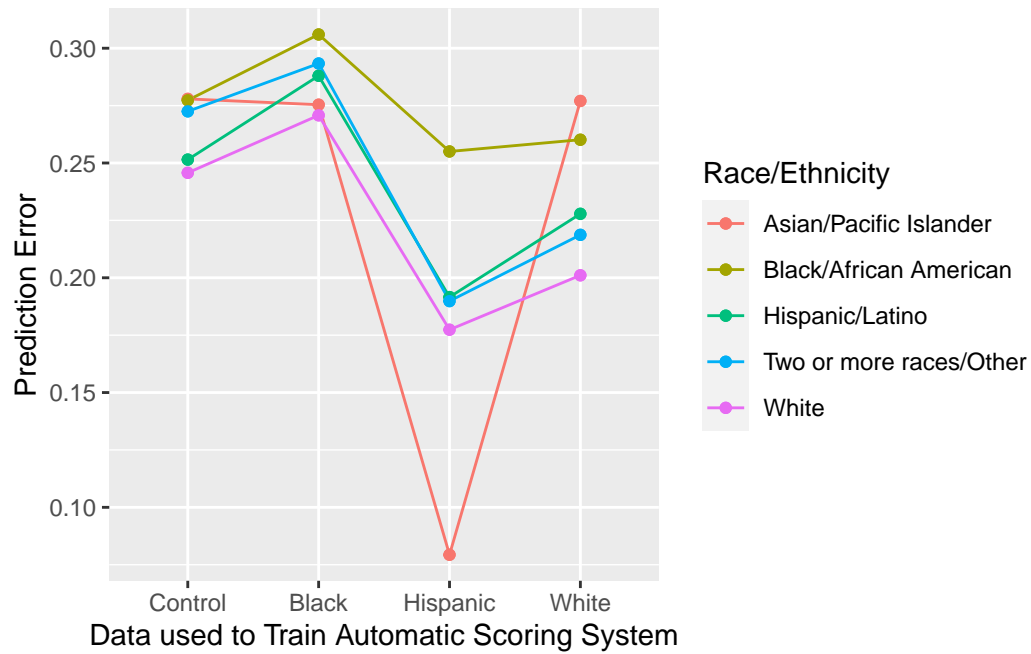
```
        tlab = "Race/Ethnicity",
        ylab = "Prediction Error",
        xlab = "Data used to Train Automatic Scoring System"
)
```



```
# ggsave("../results/transformer_race.png")
```