Efficacy of Small Class Size in Improving Educational Outcomes

Nicholas Archambault

25 February 2021

Introduction

The Student-Teacher Achievement Ratio (STAR) project was a four-year study in which students were tracked to examine the effect of kindergarten class size on educational performance and personal development. The study under consideration involved 11,601 students between 1985 and 1989 and cost over $12 million to complete.

During those four years, kindergarten students were randomly assigned to small classes, medium-sized classes, or medium-sized classes with aid. Data collection on the participants continued through the end of their high school attendance. This project will analyze a portion of the collected data in order to investigate whether small kindergarten class sizes improved future educational performance.

Variables and their descriptions can be found in the following table.

Variable	Description
race	student's race (white = 1, black = 2, Asian = 3, Hispanic = 4, Native American = 5, other = 6)
classtype	kindergarten class type (small = 1, regular = 2, regular with aid = 3)
g4math	total scaled score for math section of fourth-grade standardized test
g4reading	total scaled score for reading section of fourth-grade standardized test
yearssmall	number of years in small classes
hsgrad	high school graduation (did graduate = 1, did not graduate = 0)

1. Variable Re-coding and Data Cleaning

We begin with some basic data cleaning to more easily manipulate the data set. Such cleaning includes removing numeric labels from the `race` and `classtype` variables and re-coding these variables as factors.

The factor version of the `classtype` variable will be renamed `kinder`. Within the `race` column, Asian and Native American students will be combined with students of 'other' race.

```
# Load the data
star <- read.csv("STAR.csv", as.is = TRUE, stringsAsFactors = FALSE)

# Create 'kinder' variable
star$kinder <- factor(star$classtype,
                      labels = c("small", "regular", "aid"))

# Modify 'race' variable
star$race <- factor(star$race, labels = c("white", "black", "others",
                                          "hispanic", "others","others"))
```

2. Comparing Test Performance by Class Size

With the data sufficiently re-coded, we can begin analyzing how educational performance varies with class size. After removing missing data, we can gain a basic familiarity with the performance gap between small and regular class sizes by examining the mean and standard deviation of standardized test scores taken after students' participation in the STAR program ended.

```
# Find means
math_mean_small <- mean(star$g4math[star$kinder == "small"], na.rm = T)
math_mean_regular <- mean(star$g4math[star$kinder == "regular"], na.rm = T)

reading_mean_small <- mean(star$g4reading[star$kinder == "small"], na.rm = T)
reading_mean_regular <- mean(star$g4reading[star$kinder == "regular"],
                             na.rm = T)

# Find standard deviations
math_sd_small <- sd(star$g4math[star$kinder == "small"], na.rm = T)
math_sd_regular <- sd(star$g4math[star$kinder == "regular"], na.rm = T)

reading_sd_small <- sd(star$g4reading[star$kinder == "small"], na.rm = T)
reading_sd_regular <- sd(star$g4reading[star$kinder == "regular"], na.rm = T)

# Identify treatment effect between class size groups
(math_treatment <- math_mean_small - math_mean_regular)
```

```
## [1] -0.3362425
```

```
(reading_treatment <- reading_mean_small - reading_mean_regular)
```

```
## [1] 3.501232
```

```
# Results
sum_stats <- cbind(c(math_mean_small, math_sd_small,
                     reading_mean_small, reading_sd_small),
                   c(math_mean_regular, math_sd_regular,
                     reading_mean_regular, reading_sd_regular))
colnames(sum_stats) <- c("small", "regular")
row.names(sum_stats) <- c("math (mean)", "math (sd)", "reading (mean)",
                          "reading (sd)")

knitr::kable(sum_stats, format = "pipe", digits = 3)
```

	small	regular
math (mean)	709.185	709.521
math (sd)	43.573	41.021
reading (mean)	723.391	719.890
reading (sd)	51.545	53.168

Students in small classes scored 3.5 points higher on the fourth grade standardized reading test, but 0.336 points lower on the math test. While assigning students to smaller classes appears to have modestly improved reading performance, it did not have the intended effect for math.

With that said, the treatment effects are only small fractions of the standard deviations of scores for each subject, so these results may not be substantive. The difference of mean performance in math is just 0.77% of a standard deviation. For reading, the difference represents 6.7% of a standard deviation.

3. Quantile Breakdown

After identifying somewhat murky results when directly comparing performance between small and regular class sizes, we can dig deeper by assessing quantile treatment effects. These will be calculated as the treatment effects between the 33rd and 66th percentiles of small and regular-sized classroom score distributions.

```
# Find 33rd and 66th quantiles for small and regular classes, both subjects
reading_terciles_small <- quantile(star$g4reading[star$kinder == "small"],
                                   na.rm = TRUE, probs = c(1/3, 2/3))
reading_terciles_regular <- quantile(star$g4reading[star$kinder == "regular"],
                                     na.rm = TRUE, probs = c(1/3, 2/3))

math_terciles_small <- quantile(star$g4math[star$kinder == "small"],
                                na.rm = TRUE, probs = c(1/3, 2/3))
math_terciles_regular <- quantile(star$g4math[star$kinder == "regular"],
                                  na.rm = TRUE, probs = c(1/3, 2/3))

# Identify quantile effects
(reading_effect <- reading_terciles_small - reading_terciles_regular)
```

```
## 33.33333% 66.66667%
##         0         1
```

```
(math_effect <- math_terciles_small - math_terciles_regular)
```

```
## 33.33333% 66.66667%
##        -1         1
```

Analysis of the terciles for each academic subject compared across class size shows that treatment effects are minimal. The performance differences at the 33rd and 66th percentiles are either zero or one point for both math and reading, minuscule fractions of the standard deviations for two subjects' score distributions. This confirms that small kindergarten class sizes did little to nothing to bolster performance on fourth grade examinations.

4. Analyzing Performance By Year

While many students remained in small classes for all four years of the STAR study, others were assigned to small classes for less than four years before being transferred to regular classes or regular classes with aid. By creating proportion tables and comparing the standardized test performance of students with different years of exposure to small class sizes, we can better understand whether longer participation in the program improved educational outcomes.

```
# Proportion table of years in small classes by treatment group
round(prop.table(table(star$yearssmall, star$kinder)), 3)
```

```
##
##     small regular   aid
##   0 0.000   0.310 0.316
##   1 0.091   0.015 0.015
##   2 0.043   0.009 0.009
##   3 0.031   0.013 0.012
##   4 0.135   0.000 0.000
```

```
# Proportion table marginalized by column
round(prop.table(table(star$yearssmall, star$kinder), margin = 2), 3)
```

```
##
```

```
##     small regular   aid
##   0 0.000   0.894 0.895
##   1 0.303   0.043 0.043
##   2 0.143   0.026 0.027
##   3 0.103   0.036 0.035
##   4 0.451   0.000 0.000
```

```
prop_stats <- cbind(tapply(star$g4reading, star$yearssmall, mean, na.rm = TRUE),
                    tapply(star$g4reading, star$yearssmall, median,
                           na.rm = TRUE),
                    tapply(star$g4math, star$yearssmall, mean, na.rm = TRUE),
                    tapply(star$g4math, star$yearssmall, median, na.rm = TRUE))
colnames(prop_stats) <- c("reading (mean)", "reading (median)", "math (mean)",
                          "math (median)")

knitr::kable(prop_stats, "pipe", digits = 1)
```

	reading (mean)	reading (median)	math (mean)	math (median)
0	719.9	722.0	708.0	710
1	723.1	724.5	707.6	709
2	717.9	720.0	711.9	714
3	719.9	721.0	709.6	712
4	724.7	726.0	710.1	711

The first proportion table shows the percentages of all students found within groups defined by class size and number of years spent in small classes. The second table shows the proportion of each treatment group, or class size, with various numbers of years in small classes. Only a few children assigned to 'regular' or 'aid' classrooms were exposed to small class sizes in kindergarten.

The table above shows the mean and median test scores for math and reading among students who spent between zero and four years in small classes. There is little difference in mean and median score based on the number of years of small-class participation, further indicating that the STAR program was largely ineffective.

5. Racial Achievement Gaps

Having compared achievement outcomes between treatment and control groups and found the efficacy of the STAR program limited, we turn now to analyzing whether the program produced performance disparities between racial groups.

```
# Create function
racial_gap <- function(star, subject, treatment){
  # Subset into white and minority student groups
  white <- subset(star, (race == "white") & (kinder == treatment))
  nonwhite <- subset(star, (race == "black" | race == "hispanic") &
                       kinder == treatment)

  c("Minority" = mean(nonwhite[, subject], na.rm = TRUE),
    "White" = mean(white[, subject], na.rm = TRUE),
    "Difference" = mean(white[, subject], na.rm = TRUE) -
      mean(nonwhite[, subject], na.rm = TRUE))
}

# Racial gap results
```

```
racial_results <- rbind(racial_gap(star, "g4reading", "small"),
                        racial_gap(star, "g4reading", "regular"),
                        racial_gap(star, "g4math", "small"),
                        racial_gap(star, "g4math", "regular"))
row.names(racial_results) <- c("reading, small", "reading, regular",
                               "math, small", "math, regular")
knitr::kable(racial_results, "pipe", digits = 1)
```

	Minority	White	Difference
reading, small	699.3	727.8	28.6
reading, regular	689.4	725.1	35.8
math, small	698.2	711.2	13.0
math, regular	698.5	711.4	12.9

In both small and regular classes, white students performed about 13 points better in math than nonwhite (Black and Hispanic) students. For reading, white students performed 29 points better in small classes and 36 points better in regular classes. The race-based breakdown shows the most dramatic disparities we have observed so far between any groups. Minority students benefited more from small classes than white students, particularly in reading, where the racial performance gap is seven points smaller within small classes than regular.

6. Graduation Rates

Finally, we investigate how the effects of kindergarten class size propagate forward to affect high school graduation rates.

```
# Graduation rate by class size
tapply(star$hsgrad, star$kinder, mean, na.rm = TRUE)
```

```
##     small   regular       aid
## 0.8359202 0.8251619 0.8392857
```

```
# Graduation rate by number of years in small classes
tapply(star$hsgrad, star$yearssmall, mean, na.rm = TRUE)
```

```
##         0         1         2         3         4
## 0.8286020 0.7910448 0.8131868 0.8324607 0.8775510
```

```
# Means for racial groups within treatment group
racial_gap(star, "hsgrad", "small")
```

```
##   Minority      White Difference
##  0.7446809  0.8674699  0.1227890
```

```
# Means for racial groups within control group
racial_gap(star, "hsgrad", "regular")
```

```
##   Minority      White Difference
##  0.7395833  0.8569620  0.1173787
```

We observe minimal difference in graduation rates between students assigned to small, regular, and regular-with-aid kindergarten classes. The average graduation rate for each group hovers around 83%. Examining graduation rates broken down by number of years spent in small classes, however, reveals a relationship between better graduation outcomes and more years spent in small classes. This trend is not perfect — children with zero years spent in small classes attain higher graduation rates than those with one or two years

— but it supports the notion that greater participation in the program led to better graduation outcomes, even if it had little impact on examination performance

There is a substantial racial gap in the graduation rates of both small-class and regular-class students, and there is minimal difference — either for white or nonwhite students — between the graduation rates of students in small and regular kindergarten classes. Despite promising findings with respect to the future positive impact of increased participation in STAR, it does not appear that the program reduced race-based disparities in graduation outcomes or provides clear support for the merits of smaller early classes.

Conclusion

In this project, we have analyzed the efficacy of the STAR program, which leveraged smaller early-year class sizes in an attempt to produce better educational performance and personal development outcomes. The results indicate that the program was largely ineffective, as it failed to generate meaningful differences in exam performances or high school graduation rates between students in small classes and those in regular-sized classes. There is little evidence that students with more years of participation in the program achieved better standardized test scores, but there appears to be a modest relationship between more years of participation and higher high school graduation rates.

Examining the program from a race-based perspective, we find substantive differences in test performance between classroom sizes among white and minority (Black and Hispanic) students. This indicates that minority students benefited more from STAR's small-class focus, namely with regard to reading performance. Though there exists a stark disparity in graduation rates between white and nonwhite students, differences within racial groups based on participation in STAR's small classes are negligible.

This exercise was based in part on the 1995 work of Frederick Mosteller: "The Tennessee study of class size in the early school grades," which appeared in *The Future of Children*, vol. 5, no. 2, pp. 113-127.