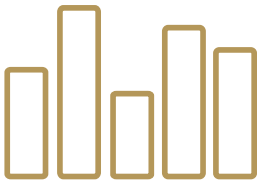


STRATEGIC DATA PROJECT

ANALYZE: HUMAN CAPITAL ANALYSIS GUIDE

SDP TOOLKIT

FOR EFFECTIVE DATA USE IN EDUCATION AGENCIES

www.gse.harvard.edu/sdp/toolkit

Toolkit Documents

An Introduction to the SDP Toolkit for Effective Data Use

Identify: Data Specification Guide

Clean: Data Building Guide for Human Capital BETA

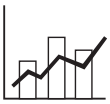
Connect: Data Linking Guide for Human Capital BETA

Analyze: Human Capital Analysis Guide

Adopt: Coding Style Guide

SDP Stata Glossary

VERSION: 1.0
Last Modified: April 4th, 2014



4. Analyze: Human Capital Analysis Guide

Conduct analyses that help answer key questions in your agency.

Analyze: Human Capital Analysis Guide is a set of step-by-step instructions to help any analyst in an education agency generate data visualizations about teacher recruitment, placement, development, evaluation, and retention. Through **Analyze**, your previous work identifying, cleaning, and connecting data will generate actual analyses to inform decision-making in your agency!

HUMAN CAPITAL ANALYSIS GUIDE

By now, you should have identified, cleaned, and connected your data into two analysis files named `Student_Teacher_Year_Analysis` and `Teacher_Year_Analysis`. You will now use these final analysis files to generate a number of analyses along the teacher pathway of being recruited, placed, developed, evaluated, and retained.

Analyze Structure

With each analysis, you will find:

- a picture of the analysis, based on the synthetic data;
- **Purpose:** an explanation of each analysis' value, and its ability to support understanding of teacher career trajectories and effectiveness patterns in your agency;
- **Required analysis file variables:** the variables from the analysis file you will need;
- **Analysis-specific sample restrictions:** a list of restrictions that you will apply to define the sample for the analysis;
- **Ask yourself:** a set of questions to help interpret results and invite deeper inquiry;
- **Potential further analyses:** further analyses you may conduct to understand underlying causes or interventions needed; and
- **Analytic technique:** how to produce the analysis step-by-step using your analysis file and code in Stata.

Analytic Samples

There are two main analytic samples: a **teacher sample** corresponding to the `Teacher_Year_Analysis` file and a **student sample** corresponding to the `Student_Teacher_Year_Analysis` file created in **Connect**. These samples are further restricted using the **analysis-specific sample restrictions** specific to the individual analyses using indicator variables, such as `t_is_teacher`, `t_novice`, `t_experience`, and others you created in **Connect**. In the next section, we describe the **teacher sample** and **student sample** in greater detail, which you should read carefully if you are using your agency's own data to run these analyses.

Teacher sample: `Teacher_Year_Analysis`

The file will be unique by teacher ID (`tid`) and `school_year`, and includes many additional variables. For the full list of variables, refer to **Connect**.

tid	school_year	school_name	t_male	t_experience	other variables...
985	2004	Jackson Elementary	1	1	...
985	2005	Jackson Elementary	1	2	...
985	2006	Jackson Elementary	1	3	...

Ideally, teacher sample should include agency staff who:

1. received an agency salary in a given school year,
2. had a job code of "Teacher" in the same school year, and
3. were tied to students as the teacher-of-record in course or roster records in the same school year.

Generally, the teacher sample is defined by the indicator variable, `t_is_teacher`, which was generated in **Staff Task 3**: Staff Degrees and Codes in **Clean.** In this toolkit we primarily use the second criterion to identify teachers as our sample data does not include agency salary. Links to students in a class for a teacher are tackled separately in **Student Task 4**: Student Class Enrollment. In the event that teacher-student link data are poor or unavailable in your agency, you may define the teacher sample based on just the first two criteria, particularly for analyses that do not rely on estimates of teacher effectiveness.

There are a few **exceptions** to the three decision rules defined above:

- In some agencies, teaching positions are filled by long-term substitutes or other staff with job codes other than "Teacher." If this is the case, in your recruitment analyses you should consider separately reporting the number of new hires with "Teacher" job codes and the number of new hires with nonteaching job codes when many staff with nonteaching job codes are linked to students in course or roster records.
- When using a measure of teacher effectiveness, include all staff who serve as the primary teacher-of-record to students in tested grades and subjects. In your evaluation analyses, consider expanding the teacher sample to include all staff tied to students in tested grades and subjects, irrespective of their job codes. You may also consider excluding charter school teachers from your analysis if charter school data coverage is poor.
- One more caveat applies to retention analyses. First, teachers in the most recent school year for which data are available cannot be included, since we cannot identify the retention status of teachers when we have no way of observing their employment status in the following school year (because these data do not exist yet).
- Similarly, in the recruitment analyses we do not include teachers in the first year. Those data are not available, because new hire status is determined by whether or not a teacher was employed in the previous year.

Student sample: `Student_Teacher_Year_Analysis`

The file will be unique by student ID (`sid`), `school_year`, `tid_math` or `tid_ela`, and includes many additional variables. For the full list of variables, refer to **Connect**. To link students to teachers, this file is built upon a class-level file that lists the students taking a course with course and teacher information.

sid	school_year	tid_math	tid_ela	school_name	other variables...
1	2007	2657	.	Jackson Elementary	...
1	2007	.	2657	Jackson Elementary	...
2	2006	1354	.	Monroe Elementary	...
2	2006	.	5979	Monroe Elementary	...

In some analyses, you will be interested only in teachers and their employment patterns. In other analyses, however, you will need to rely on information about students, too. Our standard student sample includes all students in core courses (the result of Step 2 Restrict and Step 3 Generate in **Connect**) tied to the teachers in the teacher sample.

As a final note, it should be noted that agency source data for teachers and students often reflect the fact that students and teachers may transfer schools during the school year, and teachers may teach in more than one school. In these cases, use decision rules to simplify the structure of the data. For example, a good decision rule for assigning students to a single school is to include only students present in a school for more than half the school year.

A Note About Sample Data

Since the analyses in this section of the toolkit are generated using sample data, in some cases the findings are not typical of results generated using education agency data. If you want to compare findings from your own agency with results from other agencies, you can read some of SDP's human capital diagnostic reports at http://**www.gse.harvard.edu/sdp/diagnostics/published-findings.php** to see if analyses for your agency show similar trends. You can also look at SDP's Strategic Performance Indicators, which use data from a number of partner agencies to provide deeper insight into the human capital performance of education systems. Read about the SPIs at **www.gse.harvard.edu/sdp/spi.**

Running Your Analyses

At the beginning of your do file for each section of the human capital pathway, you will have a series of global switches indicating which analyses you would like to run. For example, here are the globals for the retention do file of the human capital pathway:

```
global retention_pie                    1
global retention_year                   1
global retention_school_poverty         0
global retention_teacher_effectiveness  1
global retention_trajectory_novices     0
```

Each analysis is enclosed by a check for this global switch so that the code inside will run if the switch has a value of "1". In this way, you can control which analyses you want to run.

```
if $retention_pie==1 {
```

```
...
}
```

Summary

After completing **Analyze**, you will have:

- used your final analysis files from **Connect** to generate and display many different analyses on teacher effectiveness,
- obtained new and confirmatory information about teachers in your agency, and
- learned essential methodologies to embark on your own "deeper dives" into the data.

Share these analyses with colleagues, peers, and senior leadership in your agency.

- Ask yourself how these analyses might further inquiry and inform policy.
- How might you adapt these analyses to track performance over time?
- What relationships were particularly informative?
- How might you extend certain analyses to be even more informative?
- Who should have this information?

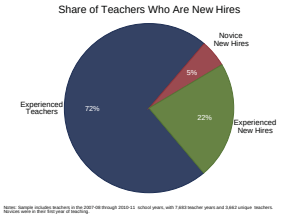
As a final note, the analyses presented here do not capture all of our research team's efforts to understand human capital in education agencies. We believe the analyses presented are some of the most widely applicable to drive discussions about change. Moreover, we believe these analyses serve as a model to seek answers about human capital trends in your agency.

We would love to hear how these model analyses inspired different analyses and "deeper dives" in your agency. As always, if you require additional support, feel free to email us at **sdp@gse.harvard.edu**.

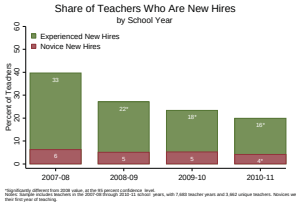
Map of Analyses

A. Recruitment

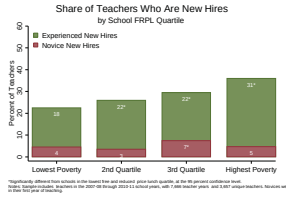
An examination of hiring within the agency, including the distribution of new hires across schools, their demographic characteristics, and then effectiveness of recent hires by certification pathway.



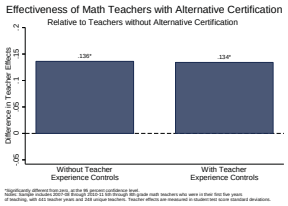
1. Share of Teachers Who Are New Hires (p. 10)



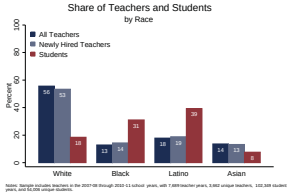
2. Share of Teachers Who Are New Hires by School Year (p. 13)



3. Share of Teachers Who Are New Hires by School Poverty Level (p. 17)



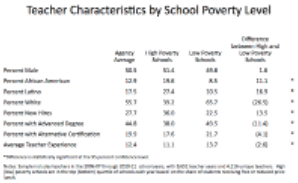
4. Teacher Effectiveness for Alternatively Certified Teachers (p. 21)



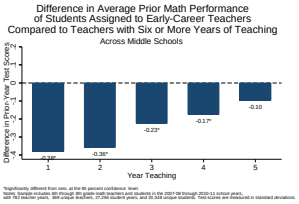
5. Share of Teachers and Students by Race (p. 27)

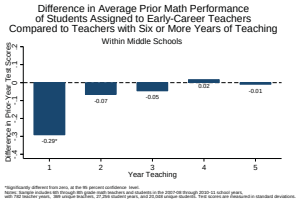
B. Placement

An examination of the patterns in student assignment to teachers across and within schools to identify places where efforts to reform placement policies could positively impact students and teachers.



1. Teacher Characteristics by School Poverty (p. 32)



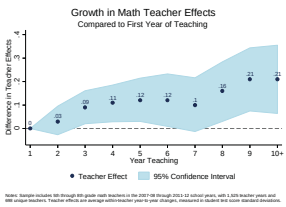


2. Student Prior Achievement by Teacher Experience (p. 36)

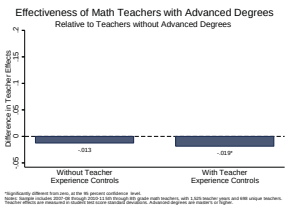
Strategic Performance Indicator

C. Development

An examination of the ways teachers develop during their careers and an exploration of whether agency incentives are aligned with gains in teacher effectiveness.



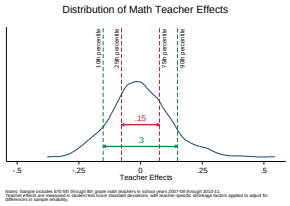
1.Growing in Teacher Effects for Early Career Teachers (p. 44)



2. Differences in Teacher Effects for Teachers with and without Advanced Degrees (p. 50)

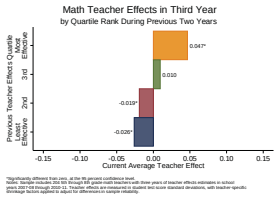
D. Evaluation

A good measure of teacher effectiveness will be spread out enough to distinguish exemplary teachers from developing ones in addition to being well correlated over time. The Evaluation section of the diagnostic examines the extent to which a teacher effectiveness measure meets criteria.



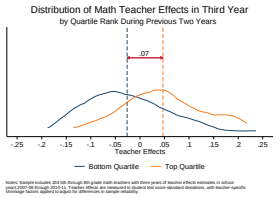
1. Distribution of Teacher Effects

(p. 57)



2. Average Teacher Effects in Third Year by Quartile Rank During Previous Two Years

(p. 61)

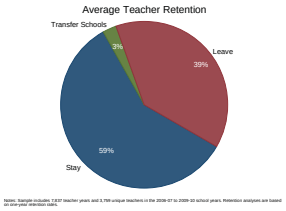


3. Distribution of Teacher Effects in Third Year by Quartile Rank During Previous Two Years

(p. 66)

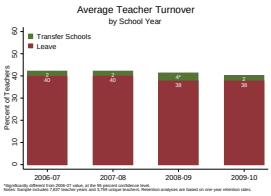
E. Retention

An examination of how retention and turnover patterns vary by school characteristics, and among teachers with different teacher effectiveness estimates.



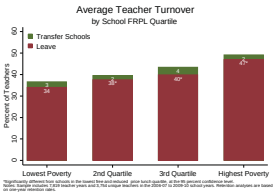
1. Average Annual Teacher Retention

(p. 71)



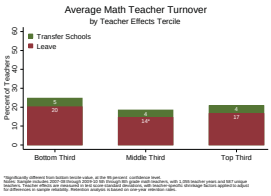
2. Teacher Turnover by School Year

(p. 74)



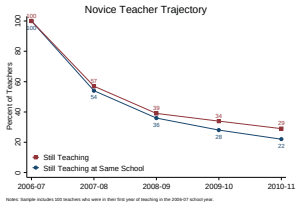
3. Teacher Turnover by School Poverty Quartile

(p. 78)



4. Teacher Turnover by Teacher Effectiveness Tercile

(p. 82)

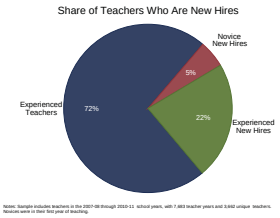


5. Novice Teacher Retention Trajectory

(p. 86)

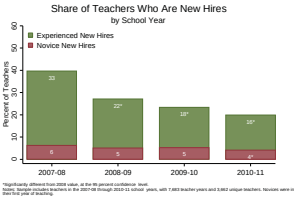
A. Recruitment

The recruitment process is the first opportunity education agencies have to secure highly effective teachers for their students. This section of the human capital pathway documents the kinds of teachers the agency hires (e.g., novies and experienced new hires), their demographic characteristics, and their distribution within the agency. Analyzing recruitment may be helpful in improving teacher recruitment and hiring by, for instance, identifying high-turnover schools, examining the demographic matches and mismatches between students and teachers, and determining which certification pathways produce the most effective teachers. Analyzing recruitment trends can provide direction for a human resource strategy that aims to attract and place highly effective teachers in all classrooms.



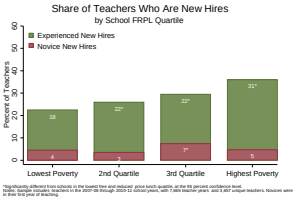
1. SHARE OF TEACHERS WHO ARE NEW HIRES

Describes the overall shapes of novice and experienced new hires.



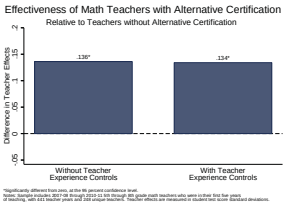
2. SHARE OF TEACHERS WHO ARE NEW HIRES BY SCHOOL YEAR

Describes the share of novice and experienced new hires by year.



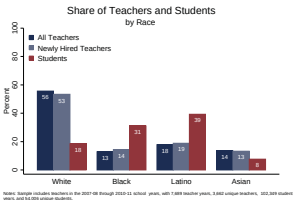
3. SHARE OF TEACHERS WHO ARE NEW HIRES BY POVERTY LEVEL

Examines the extent to which new hires are distributed unevenly across the agency according to school poverty characteristics.



4. TEACHER EFFECTIVENESS FOR ALTERNATIVELY CERTIFIED TEACHERS

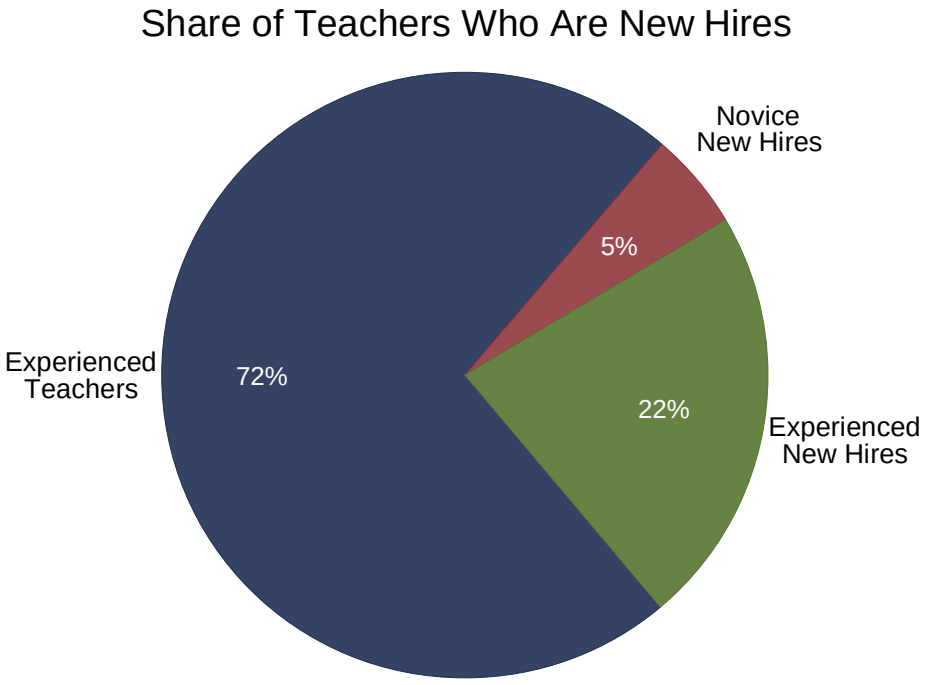
Estimates the difference in teacher effectiveness between teachers with transitional and alternative certifications.



5. SHARE OF TEACHERS AND STUDENTS BY RACE

Compares the shares of all teachers, newly hired teachers, and students by race.

1. SHARE OF TEACHERS WHO ARE NEW HIRES



Notes: Sample includes teachers in the 2007-08 through 2010-11 school years, with 7,683 teacher years and 3,662 unique teachers. Novices were in their first year of teaching.

Purpose:
Describe the share of teachers in the agency who are new hires.

Required analysis file variables:

```
tid
school_year
t_is_teacher
t_newhire
t_novice
```

Analysis-specific sample restrictions:

- Keep only employees whose job code is "teacher."
- Keep only years for which new hire information is available.

Ask yourself:

- How is your workforce balanced between novice and veteran teachers? Is the ratio what you expected?
- What are the major sources of novice new hires in your agency? Experienced new hires?
- How does your recruiting strategy affect the composition of yur teacher workforce?

Potential further analyses:
You can use a pie chart like this one to examine the overall distribution of various characteristics of your teacher workforce. For example, you can use a pie chart to examine categorical variables such as teacher gender, race, or tenure status, or group continuous variables such as in-district experience, total teaching experience, or teacher age into three to seven categories and then display the share of teachers in each category.

1. SHARE OF TEACHERS WHO ARE NEW HIRES

// Step 1: Load the Teacher_Year_Analysis data file.

```
use "${analysis}\Teacher_Year_Analysis.dta", clear
isid tid school_year
```

// Step 2: Restrict the analysis sample. Keep only employees who are teachers. Drop the first year of data, since new hires are not defined for that year. Drop records with missing values for variables important to the analysis.

```
keep if school_year > 2007
keep if t_is_teacher==1
keep if !missing(t_newhire)
keep if !missing(t_novice)
assert !missing(t_experience)
```

// Step 3: Review the values of variables to be used in the analysis.

```
tab t_newhire, mi
tab t_novice, mi
tab t_novice t_newhire, mi col
```

// Step 4: Define a new variable which includes both novice and experienced new hires.

```
gen pie_hire = .
replace pie_hire = 1 if t_newhire == 0
replace pie_hire = 2 if t_newhire == 1 & t_novice == 1
replace pie_hire = 3 if t_newhire == 1 & t_novice == 0
tab pie_hire, mi
```

// Step 5: Calculate and store sample sizes for the chart footnote.

```
summ tid
local teacher_years = string(r(N), "%9.0fc")
preserve
       bys tid: keep if _n == 1
       summ tid
       local unique_teachers = string(r(N), "%9.0fc")
restore
```

// Step 6: Create a pie chart. Footnote text is flush left to allow wrapping lines without inserting tabs in footnote.

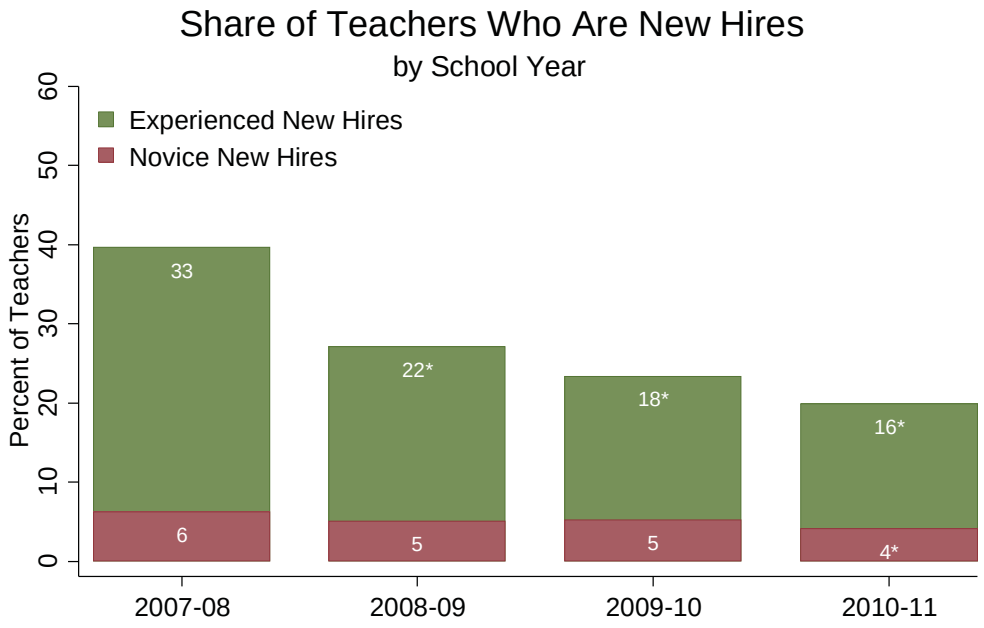
```
#delimit ;
graph pie, over (pie_hire) angle(-50)
       pie(1, color(dknavy))
       pie(2, color(maroon))
       pie(3, color(forest_green))
       plabel(_all percent, format(%3.0f) color(white) size(*1.2))
       plabel(1 "Experienced" "Teachers", gap(30) color(black) size(medsmall))
       plabel(2 "Novice" "New Hires", gap(30) color(black) size(medsmall))
       plabel(3 "Experienced" "New Hires", gap(30) color(black) size(medsmall))
```

```
       legend(off)
       graphregion(color(white) fcolor(white) lcolor(white))
       plotregion(color(white) fcolor(white) lcolor(white) margin(1 1 3 3))
       title("Share of Teachers Who Are New Hires", span)
       note(" " "Notes: Sample includes teachers in the 2007-08 through 2010-11
school years, with `teacher_years' teacher years and `unique_teachers' unique
teachers." "Novices were in their first year of teaching.", size(vsmall) span);
#delimit cr
```

// Step 7: Save the chart in Stata Graph and EMF formats.

```
graph export "${graphs}/A1_Share_of_Teachers_New_Hires.emf", replace
graph save "${graphs}/A1_Share_of_Teachers_New_Hires.gph", replace

}
```

2. SHARE OF TEACHERS WHO ARE NEW HIRES BY SCHOOL YEAR



*Significantly different from 2008 value, at the 95 percent confidence level.
Notes: Sample includes teachers in the 2007-08 through 2010-11 school years, with 7,683 teacher years and 3,662 unique teachers. Novices were in their first year of teaching.

Purpose:
Examine trends in hiring over time.

Required analysis file variables:

`tid`
`school_year`
`t_is_teacher`
`t_novice`
`t_veteran_newhire`

Analysis-specific sample restrictions:

- Keep only employees whose job code is "teacher."
- Keep only years for which new hire information is available.

Ask yourself:

- How have hiring trends changed over time?
- What factors might account for the trends that I see?

Potential further analyses:
At the state level, you may wish to examine hiring trends by year for specific school types or geographic areas. At the district level, you can make a graph of this type in order to examine overall hiring by school or for specific groups of schools, instead of by year.

2. SHARE OF TEACHERS WHO ARE NEW HIRES BY SCHOOL YEAR

// Step 1: Load the Teacher_Year_Analysis data file.

```
use "${analysis}\Teacher_Year_Analysis.dta", clear
isid tid school_year
```

// Step 2: Restrict the analysis sample.

```
keep if school_year > 2007
keep if t_is_teacher==1
keep if !missing(t_newhire)
keep if !missing(t_novice)
assert !missing(t_experience, t_veteran_newhire)
```

// Step 3: Review variables to be used in the analysis.

```
tab school_year t_novice, mi row
tab school_year t_veteran_newhire, mi row
tab t_novice t_veteran_newhire
```

// Step 4: Calculate sample size.

```
summ tid
local teacher_years = string(r(N), "%9.0fc")
preserve
       bys tid: keep if _n == 1
       summ tid
       local unique_teachers = string(r(N), "%9.0fc")
restore
```

// Step 5: Calculate significance indicator variables by year.

```
foreach var in t_novice t_veteran_newhire {
       gen sig_`var' = .
       xi: logit `var' i.school_year, robust
       forvalues year = 2009/2011 {
              replace sig_`var' = abs(_b[_Ischool_ye_`year'] / _se[_Ischool_ye_`year']) ///
                     if school_year == `year'
              replace sig_`var' = 0 if sig_`var' <= 1.96 & school_year == `year'
              replace sig_`var' = 1 if sig_`var' > 1.96 & school_year == `year'
       }
              replace sig_`var' = 0 if school_year == 2008
       }
```

2. SHARE OF TEACHERS WHO ARE NEW HIRES BY SCHOOL YEAR

// Step 6: Collapse the teacher-level data file to calculate percent of new hires by year.

```
collapse (mean) t_novice t_veteran_newhire sig_*, by(school_year)
foreach var in t_novice t_veteran_newhire {
	replace `var' = 100 * `var'
}
```

// Step 7: Concatenate values and significance asterisks to make value labels.

```
foreach var of varlist t_novice t_veteran_newhire {
	tostring(sig_`var'), replace
	replace sig_`var' = "*" if sig_`var' == "1"
	replace sig_`var' = "" if sig_`var' == "0"
	gen `var'_str = string(`var', "%9.0f")
	egen `var'_label = concat(`var'_str sig_`var')
}
```

// Step 8: Get the total new hire percent for each year for graphing.

```
gen t_total = t_novice + t_veteran_newhire
```

// Step 9: Create a stacked bar graph using overlaid bars. Use scatter plots with invisible symbols for the value labels.

```
#delimit ;
twoway (bar t_total school_year,
		fcolor(forest_green) lcolor(forest_green) lwidth(0) barwidth(0.75))
	(bar t_novice school_year,
		fcolor(maroon) lcolor(maroon) lwidth(0) barwidth(0.75))
	(scatter t_total school_year,
		mcolor(none) mlabel(t_veteran_newhire_label) mlabcolor(white) mlabpos(6)
		mlabsize(small))
	(scatter t_novice school_year,
		mcolor(none) mlabel(t_novice_label) mlabcolor(white) mlabpos(6)
		mlabsize(small)),
	title("Share of Teachers Who Are New Hires", span)
	subtitle("by School Year", span)
	ytitle("Percent of Teachers")
	ylabel(0(10)60, nogrid labsize(medsmall))
	xtitle("")
	xlabel(2008 "2007-08" 2009 "2008-09" 2010 "2009-10" 2011 "2010-11",

		labsize(medsmall))
	legend(order(1 "Experienced New Hires" 2 "Novice New Hires")
		ring(0) position(11) symxsize(2) symysize(2) rows(2) size(medsmall)
		region(lstyle(none) lcolor(none) color(none)))
	graphregion(color(white) fcolor(white) lcolor(white))
	plotregion(color(white) fcolor(white) lcolor(white) margin(2 0 2 0))
	note(" " "*Significantly different from 2008 value, at the 95 percent confidence
	level."	"Notes: Sample includes teachers in the 2007-08 through 2010-11 school
```

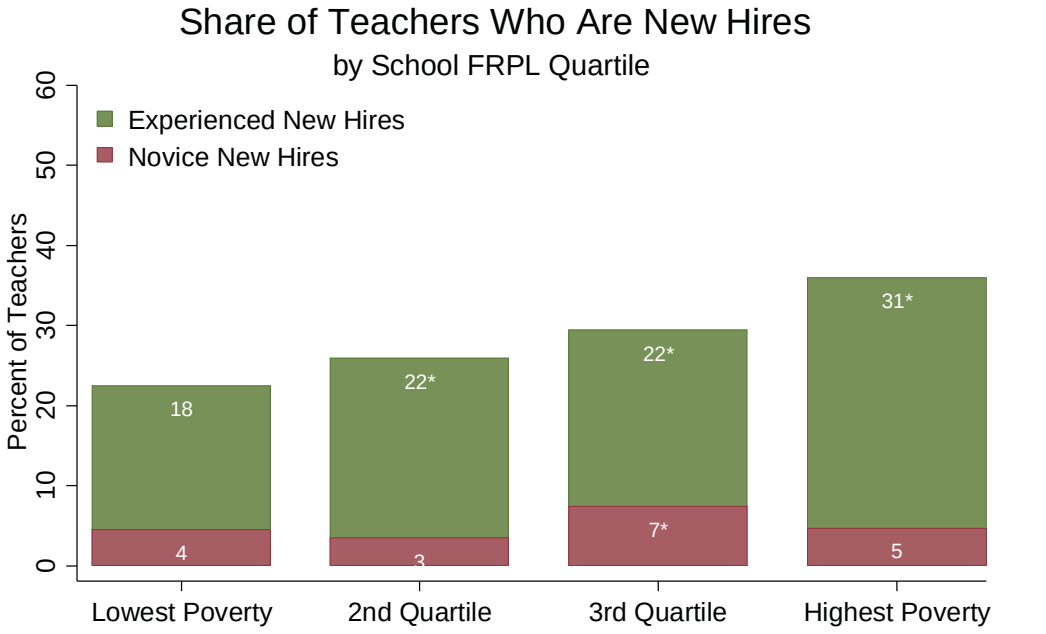
2. SHARE OF TEACHERS WHO ARE NEW HIRES BY SCHOOL YEAR

```
      years, with `teacher_years' teacher years and `unique_teachers' unique teachers.
      Novices were in" "their first year of teaching.", size(vsmall) span);

#delimit cr
```

// Step 10: Save the chart in Stata Graph and EMF formats.

```
graph export "${graphs}/A2_New Hires_by_School_Year.emf", replace
graph save "${graphs}/A2_New Hires_by_School_Year.gph", replace

}
```

3. SHARE OF TEACHERS WHO ARE NEW HIRES BY SCHOOL POVERTY LEVEL



*Significantly different from schools in the lowest free and reduced price lunch quartile, at the 95 percent confidence level.
Notes: Sample includes teachers in the 2007-08 through 2010-11 school years, with 7,666 teacher years and 3,657 unique teachers. Novices were in their first year of teaching.

Purpose:
Examine the extent to which new hires are distributed unevenly across the agency according to school characteristics.

Required analysis file variables:

```
tid
school_year
t_is_teacher
t_novice
t_veteran_newhire
school_poverty_quartile
```

Ask yourself:

- How do hiring patterns differ between high and low-poverty schools?
- Are the shares of novice and veteran hires distributed equitably and strategically across school poverty quartiles?

Analysis-specific sample restrictions:

- Keep only employees whose job code is "teacher."
- Keep only years for which new hire information is available.

Potential further analyses:

- You can use a version of this graph to look at how new hires are distributed across other quartiles of school characteristics. For example, you can examine new hiring by school average test score quartile, or school minority percent quartile.

3. SHARE OF TEACHERS WHO ARE NEW HIRES BY SCHOOL FRPL QUARTILE

// Step 1: Load the Teacher_Year_Analysis data file.

```
use "${analysis}\Teacher_Year_Analysis.dta", clear
isid tid school_year
```

// Step 2: Restrict the analysis sample.

```
keep if school_year > 2007
keep if t_is_teacher==1
keep if !missing(t_newhire)
keep if !missing(t_novice)
keep if !missing(school_poverty_quartile)
assert !missing(t_experience, t_veteran_newhire)
```

// Step 3: Review variables used in the analysis.

```
tab school_poverty_quartile, mi
tab school_poverty_quartile t_novice, mi row
tab school_poverty_quartile t_veteran_newhire, mi row
```

// Step 4: Calculate sample size.

```
summ tid
local teacher_years = string(r(N), "%9.0fc")
preserve
       bys tid: keep if _n == 1
       summ tid
       local unique_teachers = string(r(N), "%9.0fc")
restore
```

// Step 5: Calculate significance indicator variables by school poverty quartile.

```
foreach var of varlist t_novice t_veteran_newhire {
       gen sig_`var' = .
       xi: reg `var' i.school_poverty_quartile, robust
       forval quart = 2/4 {
              replace sig_`var' = abs(_b[_Ischool_po_`quart']/_se[_Ischool_po_`quart']) ///
                     if school_poverty_quartile == `quart'

              replace sig_`var' = 0 if sig_`var' <= 1.96 & school_poverty_quartile ==`quart'
              replace sig_`var' = 1 if sig_`var' > 1.96 & school_poverty_quartile == `quart'
       }
       replace sig_`var' = 0 if school_poverty_quartile == 1
}
```

3. SHARE OF TEACHERS WHO ARE NEW HIRES BY SCHOOL FRPL QUARTILE

// Step 6: Collapse to calculate shares of new hires in each quartile.

```
collapse (mean) t_novice t_veteran_newhire sig_*, by(school_poverty_quartile)
foreach var of varlist t_novice t_veteran_newhire {
	replace `var' = 100 * `var'
}
```

// Step 7: Concatenate values and significance asterisks to make value labels.

```
foreach var of varlist t_novice t_veteran_newhire {
	tostring(sig_`var'), replace
	replace sig_`var' = "*" if sig_`var' == "1"
	replace sig_`var' = "" if sig_`var' == "0"
	gen `var'_str = string(`var', "%9.0f")
	egen `var'_label = concat(`var'_str sig_`var')
}
```

// Step 8: Get the total new hire percent for each year for graphing.

```
gen t_total = t_novice + t_veteran_newhire
```

// Step 9: Create a bar graph using twoway bar and scatter for the labels.

```
#delimit ;
twoway (bar t_total school_poverty_quartile,
		fcolor(forest_green) lcolor(forest_green) lwidth(0) barwidth(0.75))
	(bar t_novice school_poverty_quartile,
		fcolor(maroon) lcolor(maroon) lwidth(0) barwidth(0.75))
	(scatter t_total school_poverty_quartile,
		mcolor(none) mlabel(t_veteran_newhire_label) mlabcolor(white) mlabpos(6)
		mlabsize(small))
	(scatter t_novice school_poverty_quartile,
		mcolor(none) mlabel(t_novice_label) mlabcolor(white) mlabpos(6)
		mlabsize(small)),
	title("Share of Teachers Who Are New Hires", span)
	subtitle("by School FRPL Quartile", span)
	ytitle("Percent of Teachers")
	ylabel(0(10)60, nogrid labsize(medsmall))
	xtitle("")
	xlabel(1 "Lowest Poverty" 2 "2nd Quartile" 3 "3rd Quartile" 4 "Highest Poverty",
		labsize(medsmall))
	legend(order(1 "Experienced New Hires" 2 "Novice New Hires")
		ring(0) position(11) symxsize(2) symysize(2) rows(2) size(medsmall)
		region(lstyle(none) lcolor(none) color(none)))
```

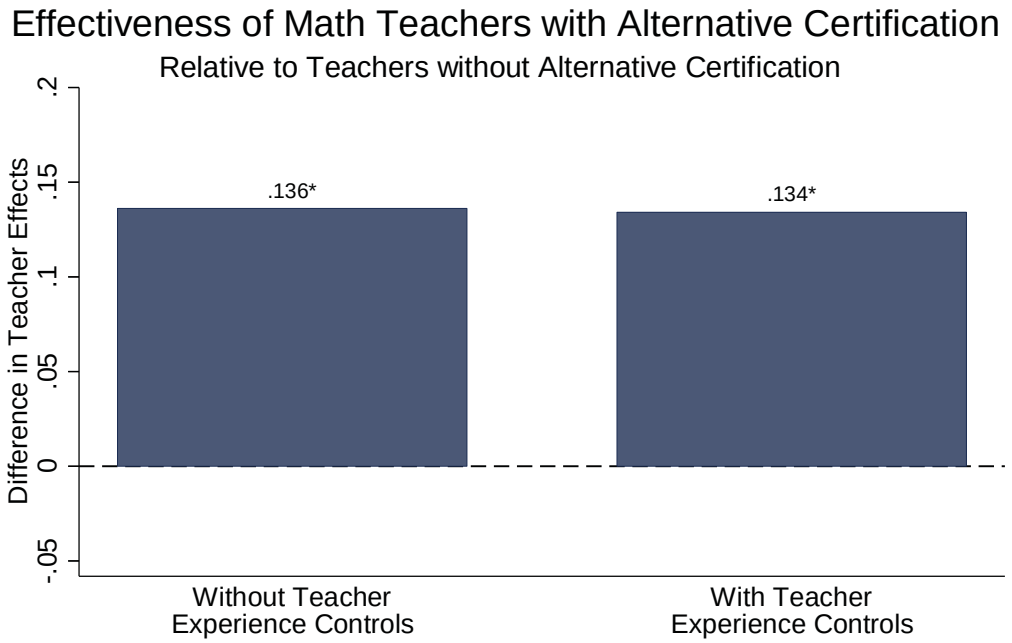
3. SHARE OF TEACHERS WHO ARE NEW HIRES BY SCHOOL FRPL QUARTILE

```
	graphregion(color(white) fcolor(white) lcolor(white))
	plotregion(color(white) fcolor(white) lcolor(white) margin(2 0 2 0))
	note(" " "*Significantly different from schools in the lowest free and reduced price
lunch quartile, at the 95 percent confidence level." "Notes: Sample includes teachers
in the 2007-08 through 2010-11 school years, with `teacher_years' teacher years
and `unique_teachers' unique teachers. Novices were" "in their first year of
teaching.", size(vsmall) span);
#delimit cr
```

// Step 10: Save the chart in Stata Graph and EMF formats. If marker labels needs to moved by hand using Stata Graph Editor, re-same .gph and .emf files after editing.

```
graph export "${graphs}/A3_New Hires_by_Poverty_Quartile.emf", replace
graph save "${graphs}/A3_New Hires_by_Poverty_Quartile.gph", replace

}
```

4. TEACHER EFFECTIVENESS FOR ALTERNATIVELY CERTIFIED TEACHERS



*Significantly different from zero, at the 95 percent confidence level.
Notes: Sample includes 2007-08 through 2010-11 5th through 8th grade math teachers who were in their first five years of teaching, with 441 teacher years and 248 unique teachers. Teacher effects are measured in student test score standard deviations.

Purpose:
Determine whether teacher effectiveness varies by certification status.

Required analysis file variables:

```
sid
school_year
tid_year
tid_math
cid_math
grade_level
t_is_teacher
t_experience
t_certification_pathway
std_scaled_score_math
std_scaled_score_math_tm1
std_scaled_score_ela_tm1
```

(See full list of variables in **Connect**): student characteristics vector cohort characteristics vector)

Analysis-specific sample restrictions:

- Restrict to teachers with five or fewer years of teaching experience.
- Restrict to teachers with certification pathway information.
- Restrict to grades and subjects included in value-added estimates.
- Restrict to years for which prior-year test scores are available.
- Employ all other value-added estimate sample restrictions (see the value-added model technical appendix for the full list of sample restrictions).

Ask yourself:

- Are certain certification pathways associated with larger teacher effects? One way to explore this is by comparing teacher effect estimates across specific programs. Oftentimes, program-specific data are not available to perform this analysis, but when the data are accessible, we undertake such analyses using the analytic guidelines above.
- Given the results, what recruitment policies might the agency implement or change?
- What questions does this raise about teacher preparation and professional development in your agency?

Potential further analyses:
If you are a state, you might examine these trends across/by agencies to understand how location is linked to recruitment of teachers with alternative vs. traditional certification.

4. TEACHER EFFECTIVENESS FOR ALTERNATIVELY CERTIFIED TEACHERS

// Step 1: Choose a subject for the analysis. Note: To change from math to ELA, switch the subjects in the next two lines. To generate ELA and math charts at the same time, enclose the analysis code within a loop.

```
local subject math
local alt_subject ela
```

// Step 2. Load the Student_Teacher_Year_Analysis data file.

```
use "${analysis}\Student_Teacher_Year_Analysis.dta", clear
isid sid school_year
```

// Step 3: Restrict the sample. Keep only teachers in their first five years of teaching. Keep grades and years for which prior-year test scores are available. Keep students with teachers with non-missing values for experience and certification pathway. Keep students with a single identified core course and current and prior-year test scores in the given subject.

```
keep if school_year >= 2008 & school_year <= 2011
keep if grade_level >= 5 & grade_level <= 8
keep if t_is_teacher == 1
keep if t_experience <= 5
keep if !missing(t_certification_pathway)
keep if !missing(cid_`subject')
keep if !missing(std_scaled_score_`subject', std_scaled_score_`subject'_tm1)
```

// Step 4: Review teacher variables.

```
tab school_year
unique tid_`subject'
unique tid_`subject' school_year
bysort tid_`subject' school_year: gen tag = (_n == 1)
tab t_experience t_certification_pathway if tag == 1, mi
drop tag
```

// Step 5: Create a dummy variable for alternative certification

```
gen alternative_certification = (t_certification_pathway > 1 & ///
       t_certification_pathway != .)
tab alternative_certification t_certification_pathway, mi
```

// Step 6: Create dummy variables for each year of teaching experience.

```
tab t_experience, gen(exp)
```

4. TEACHER EFFECTIVENESS FOR ALTERNATIVELY CERTIFIED TEACHERS

// Step 7: Create variable for grade-by-year fixed effects.

```
egen grade_by_year = group(grade_level school_year)
```

// Step 8: Create variables for previous year's score squared and cubed.

```
gen std_scaled_score_`subject'_tm1_sq = std_scaled_score_`subject'_tm1^2
gen std_scaled_score_`subject'_tm1_cu = std_scaled_score_`subject'_tm1^3
```

// Step 9: Create indicator for whether student is missing prior achievement for alternate subject. Make a replacement variable that imputes score to zero if missing.

```
gen miss_std_scaled_score_`alt_subject'_tm1 = ///
       missing(std_scaled_score_`alt_subject'_tm1)
gen _IMPstd_scaled_score_`alt_subject'_tm1 = std_scaled_score_`alt_subject'_tm1
replace _IMPstd_scaled_score_`alt_subject'_tm1 = 0 ///
       if miss_std_scaled_score_`alt_subject'_tm1 == 1
```

// Step 10: Identify prior achievement variables to use as controls.

```
#delimit ;
local prior_achievement
       "std_scaled_score_`subject'_tm1
       std_scaled_score_`subject'_tm1_sq
       std_scaled_score_`subject'_tm1_cu
       _IMPstd_scaled_score_`alt_subject'_tm1
       miss_std_scaled_score_`alt_subject'_tm1";
#delimit cr
```

// Step 11: Identify other student variables to use as controls.

```
#delimit;
local student_controls
       "s_male
       s_black
       s_asian
       s_latino
       s_naam
       s_mult
       s_racemiss
       s_reducedlunch
       s_freelunch
       s_lunch_miss
       s_retained
       s_retained_miss
       s_gifted
       s_gifted_miss
       s_iep
       s_iep_miss
```

4. TEACHER EFFECTIVENESS FOR ALTERNATIVELY CERTIFIED TEACHERS

```
       s_ell
       s_ell_miss
       s_absence_high
       s_absence_miss";
#delimit cr
```

// Step 12: Review all variables to be included in the teacher effectiveness models. Class and cohort (grade/school/year) variables should include means of all student variables, and means, standard deviations, and percent missing for prior-year test scores for both main and alternate subject. Class and cohort size should also be included as controls.

```
codebook std_scaled_score_`subject' alternative_certification exp2-exp5
codebook `prior_achievement'
codebook `student_controls'
codebook _CL*`subject'*
codebook _CO*
codebook grade_by_year cid_`subject'
```

// Step 13: Estimate differences in teacher effectiveness between alternatively and traditionally certified teachers, without teacher experience controls.

```
reg std_scaled_score_`subject' alternative_certification ///
       `student_controls' `prior_achievement' _CL*`subject'* _CO* ///
       i.grade_by_year, cluster(cid_`subject')
```

// Step 14: Store coefficient and standard error.

```
gen coef_noexp = _b[alternative_certification]
gen se_noexp = _se[alternative_certification]
```

// Step 15: Get teacher sample size for this model.

```
egen teachers_in_sample_noexp = nvals(tid_`subject') if e(sample)
summ teachers_in_sample_noexp
local teachers_in_sample_noexp = r(mean)
```

// Step 16: Estimate differences in teacher effectiveness between alternatively and traditionally certified teachers, with teacher experience controls.

```
reg std_scaled_score_`subject' alternative_certification exp2-exp5 ///
       `student_controls' `prior_achievement' _CL*`subject'* _CO* ///
       i.grade_by_year, cluster(cid_`subject')
```

// Step 17: Store coefficient and standard error.

```
gen coef_wexp = _b[alternative_certification]
gen se_wexp = _se[alternative_certification]
```

4. TEACHER EFFECTIVENESS FOR ALTERNATIVELY CERTIFIED TEACHERS

// Step 18: Get teacher sample size for this model and compare sample size for the two models.

```
egen teachers_in_sample_wexp = nvals(tid_`subject') if e(sample)
summ teachers_in_sample_wexp
local teachers_in_sample_wexp = r(mean)
assert `teachers_in_sample_wexp' == `teachers_in_sample_noexp'
```

// Step 19: Store teacher sample size for footnote.

```
egen teacher_years = nvals(tid_`subject' school_year) if e(sample)
summ teacher_years
local teacher_years = string(r(mean), "%9.0fc")
egen unique_teachers = nvals(tid_`subject') if e(sample)
summ unique_teachers
local unique_teachers = string(r(mean), "%9.0fc")
```

// Step 20: Collapse dataset for graphing.

```
collapse(max) coef* se*
```

// Step 21: Get signficance.

```
foreach spec in noexp wexp {
       gen sig_`spec' = abs(coef_`spec') - 1.96 * se_`spec' > 0
}
```

// Step 22: Reshape for graphing.

```
gen results = 1
reshape long coef_ se_ sig_, i(results) j(spec) string
rename coef_ coef
rename se_ se
rename sig_ sig
replace spec = "1" if spec == "noexp"
replace spec = "2" if spec == "wexp"
destring spec, replace
```

// Step 23: Make value labels with significance indicator.

```
tostring sig, replace
replace sig = "*" if sig == "1"
replace sig = ""  if sig == "0"
replace coef = round(coef,.001)
egen coef_label = concat(coef sig)
```

// Step 24: Define subject titles for graph.

4. TEACHER EFFECTIVENESS FOR ALTERNATIVELY CERTIFIED TEACHERS

```
if "`subject'" == "math" {
       local subject_foot "math"
       local subject_title "Math"
}

if "`subject'" == "ela" {
       local subject_foot "English/Language Arts"
       local subject_title "ELA"
}
```

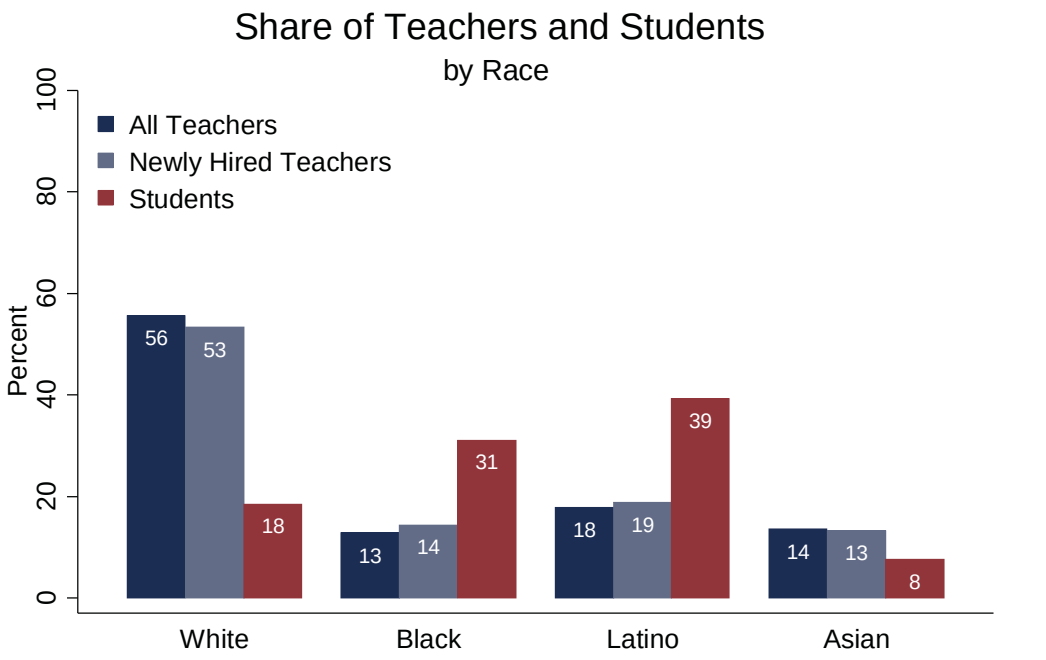
// Step 25: Create a bar graph of the estimation results.

```
#delimit ;
graph twoway (bar coef spec,
              fcolor(dknavy) lcolor(dknavy) lwidth(0) barwidth(0.7))
       (scatter coef spec,
              mcolor(none) mlabel(coef_label) mlabcolor(black) mlabpos(12)
              mlabsize(small)),
       yline(0, style(extended) lpattern(dash) lwidth(medthin) lcolor(black))
       title("Effectiveness of `subject_title' Teachers with Alternative Certification", span)
       subtitle("Relative to Teachers without Alternative Certification", span)
       ytitle("Difference in Teacher Effects", size(medsmall))
       yscale(range(-.05 .2))
       ytick(-.05(.05).2)
       ylabel(-.05(.05).2, nogrid)
       xlabel("", notick)
       xtitle("")
       xlabel(1 `""Without Teacher" "Experience Controls""'
              2 `""With Teacher" "Experience Controls""', labsize(medsmall))
       legend(off)
       graphregion(color(white) fcolor(white) lcolor(white))
       plotregion(color(white) fcolor(white) lcolor(white) margin(5 5 2 0))
       note(" " "*Significantly different from zero, at the 95 percent confidence level."
"Notes: Sample includes 2007-08 through 2010-11 `subject_foot' 5th through 8th grade teachers who
were in their first five years of teaching, with `teacher_years' teacher years and `unique_teachers'
unique teachers. Teacher effects are measured in student test score standard deviations.",
size(vsmall) span);
#delimit cr
```

// Step 26: Save chart.

```
graph export "${graphs}/A4_Teacher_Effects_Cert_Pathway_`subject_title'.emf", replace
graph save "${graphs}/A4_Teacher_Effects_Cert_Pathway_`subject_title'.gph", replace
```

5. SHARE OF TEACHERS AND STUDENTS BY RACE



Notes: Sample includes teachers in the 2007-08 through 2010-11 school years, with 7,689 teacher years, 3,662 unique teachers, 102,349 student years, and 54,006 unique students.

Purpose:
Compares the shares of all teachers, newly hired teachers, and students by race.

Required analysis file variables:

```
tis
school_year
t_is_teacher
t_newhire
t_white
t_black
t_latino
t_asian
sid
s_race_ethnicity
```

Ask yourself:

- Is the racial composition of your teacher workforce similar to the racial composition of your student body? Is there a difference in racial composition between all teachers and newly hired teachers?
- If there is a difference between teachers and students, what impact might this have on student learning?

Analysis-specific sample restrictions:

- For the teacher sample, keep only employees whose job code is "teacher."
- For the student and teacher samples, keep only records for which race information is not missing.
- For the student and teacher samples, keep only years for which teacher new hire information is available.

Potential further analyses:

- You may wish to replicate this analysis for specific schools or groups of schools.

5. SHARE OF TEACHERS AND STUDENTS BY RACE

// Step 1: Set up matrix to hold teacher, new teacher, and student results.

```
matrix race = J(4, 4, .)
matrix colnames race = race teacher new_teacher student
```

// Step 2: Load the Teacher_Year_Analysis data file.

```
use "${analysis}\Teacher_Year_Analysis.dta", clear
isid tid school_year
```

// Step 3: Restrict the teacher sample.

```
keep if school_year > 2007
keep if t_is_teacher == 1
keep if !missing(t_race_ethnicity)
keep if !missing(t_newhire)
```

// Step 4: Review teacher variables.

```
tab school_year t_race_ethnicity, mi
tab t_newhire t_white, mi row
tab t_newhire t_black, mi row
tab t_newhire t_latino, mi row
tab t_newhire t_asian, mi row
```

// Step 5: Get teacher sample sizes.

```
summ tid
local teacher_years = string(r(N), "%6.0fc")
preserve
       bys tid: keep if _n == 1
       summ tid
       local unique_teachers = string(r(N), "%6.0fc")
restore
```

// Step 6: Store percentages by race for all teachers and newly hired teachers.

```
local i = 1
foreach race of varlist t_white t_black t_latino t_asian {
       matrix race[`i', 1] = `i'
       summ `race'
       matrix race[`i', 2] = 100 * r(mean)
       summ `race' if t_newhire == 1
       matrix race[`i', 3] = 100 * r(mean)
       local i = `i' + 1
}
```

// Step 7: Load the Connect_Step1 data file to get student data.

```
use "${analysis}\Connect_Step1.dta", clear
isid sid cid
```

5. SHARE OF TEACHERS AND STUDENTS BY RACE

// Step 8: Make the file unique by sid and school_year.

```
keep sid school_year s_race_ethnicity
duplicates drop
isid sid school_year
```

// Step 9: Restrict the student sample.

```
keep if school_year > 2007
keep if !missing(s_race_ethnicity)
```

// Step 10: Review student variables.

```
tab school_year s_race_ethnicity, mi
```

// Step 11: Create dummy variables for major student race/ethnicity categories.

```
gen s_black = (s_race_ethnicity == 1)
gen s_asian = (s_race_ethnicity == 2)
gen s_latino = (s_race_ethnicity == 3)
gen s_white = (s_race_ethnicity == 5)
```

// Step 12: Get student sample sizes.

```
summ sid
local student_years = string(r(N), "%9.0fc")
preserve
       bys sid: keep if _n == 1
       summ sid
       local unique_students = string(r(N), "%9.0fc")
restore
```

// Step 13: Store percentages by race for students.

```
local i = 1
foreach race of varlist s_white s_black s_latino s_asian{
       summ `race'
       matrix race[`i', 4] = 100 * r(mean)
       local i = `i' + 1
}
```

// Step 14: Replace the dataset with the matrix of results.

```
clear
svmat race, names(col)
```

// Step 15: Graph the results.

```
#delimit ;
graph bar teacher new_teacher student,
       bar(1, fcolor(dknavy) lcolor(dknavy))
```

5. SHARE OF TEACHERS AND STUDENTS BY RACE

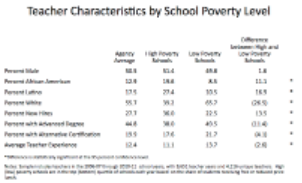
```
       bar(2, fcolor(dknavy*.7) lcolor(dknavy*.7))
       bar(3, fcolor(maroon) lcolor(maroon))
       blabel(bar, position(inside) color(white) format(%10.0f))
       over(race, relabel(1 "White" 2 "African American" 3 "Hispanic" 4 "Asian")
              label(labsize(medsmall)))
       title("Share of Teachers and Students", span)
       subtitle("by Race", span)
       ytitle("Percent", size(medsmall))
       ylabel(0(20)100, labsize(medsmall) nogrid)
       legend(order(1 "All Teachers" 2 "Newly Hired Teachers" 3 "Students")
              ring(0) position(11) symxsize(2) symysize(2) rows(3)
              size(medsmall) region(lstyle(none) lcolor(none) color(none)))
       graphregion(color(white) fcolor(white) lcolor(white))
       plotregion(color(white) fcolor(white) lcolor(white) margin(5 5 2 0))
       note(" " "Notes: Sample includes teachers in the 2007-08 through 2010-11 school
years, with `teacher_years' teacher years, `unique_teachers' unique teachers,
`student_years' student" "years, and `unique_students' unique students.", size(vsmall) span);
#delimit cr
```

// Step 16: Save the chart.

```
graph export "${graphs}/A5_Share_Teachers_Students_by_Race.emf", replace
graph save "${graphs}/A5_Share_Teachers_Students_by_Race.gph", replace
```

B. Placement

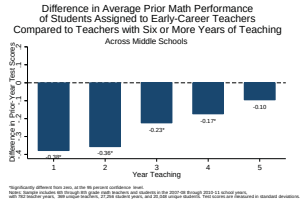
Students are not randomly assigned to teachers' classrooms. Sometimes nonrandom assignment benefits students. For example, some teachers have a talent for working with hard-to-reach students and might get assigned to more low-performing students than other teachers. Unfortunately, some placement decisions have little to do with students' needs. When the most senior teachers concentrate in districts, schools, and classrooms with the most advantaged students, while novice teachers teach lower-performing students in hard-to-staff schools, achievement gaps widen. The SDP Placement analyses reveal patterns in student assignment to teachers across and within schools to identify places where efforts to reform placement policies could positively impact students and teachers.

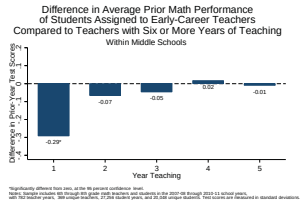

Teacher Characteristics by School Poverty Level

	Agency Average	High Poverty Schools	Low Poverty Schools	Difference between High and Low Poverty Schools	
Percent Male	48.6	50.4	49.8	1.6	
Percent African American	52.8	78.8	8.8	11.1	*
Percent Latino	13.5	29.4	10.5	18.9	*
Percent White	55.7	30.3	65.7	(26.5)	*
Percent New Hires	25.3	36.6	22.5	13.5	*
Percent with Advanced Degree	44.8	38.0	49.5	(11.4)	*
Percent with Alternative Certification	19.5	17.6	21.7	(4.1)	*
Average Teacher Experience	12.4	11.1	13.7	(2.6)	*

1. TEACHER CHARACTERISTICS BY SCHOOL POVERTY

Examines the characteristics of teachers in the highest and lowest poverty schools.





2. STUDENT PRIOR ACHIEVEMENT BY TEACHER EXPERIENCE

Compares the prior-year test scores of students in early career teachers' classrooms with those of veteran teachers' students.

1. TEACHER CHARACTERISTICS BY SCHOOL POVERTY

Teacher Characteristics by School Poverty Level

	Agency Average	High Poverty Schools	Low Poverty Schools	Difference between High and Low Poverty Schools	
Percent Male	50.5	51.4	49.8	1.6	
Percent African American	12.9	19.6	8.5	11.1	*
Percent Latino	17.5	27.4	10.5	16.9	*
Percent White	55.7	39.2	65.7	(26.5)	*
Percent New Hires	27.7	36.0	22.5	13.5	*
Percent with Advanced Degree	44.8	38.0	49.5	(11.4)	*
Percent with Alternative Certification	19.9	17.6	21.7	(4.1)	*
Average Teacher Experience	12.4	11.1	13.7	(2.6)	*

*Difference is statistically significant at the 95 percent confidence level.

Notes: Sample includes teachers in the 2006-07 through 2010-11 school years, with 9,651 teacher years and 4,116 unique teachers. High (low) poverty schools are in the top (bottom) quartile of schools each year based on the share of students receiving free or reduced price lunch.

Purpose:
Examine the distribution of teachers across school characteristics.

Required analysis file variables:

```
tid
school_year
school_poverty_quartile
t_node
t_black
t_latino
t_white
t_newhire
t_adv_degree
t_certification_pathway
t_experience
```

Analysis-specific sample restrictions:

- Keep only employees whose job code is "teacher".
- Keep only years for which new line information is available.

Ask yourself:

- What supports could your agency offer to high-poverty schools with large shares of novice and early-career teachers?
- Does your agency have a strategy for placing effective teachers in high-needs schools? What initiatives would support this goal?

Potential further analyses:
Other teacher characteristics to add to the table if data are available include: attended a competitive postsecondary institution; master teacher, instructional mentor, other teacher leadership role. You may wish to explore to what extent these teacher characteristics are associated with value-added estimates of teacher effectiveness.

B. Placement

1. TEACHER CHARACTERISTICS BY SCHOOL POVERTY

// Step 1: Load the Teacher_Year_Analysis data file.

```
use "${analysis}\Teacher_Year_Analysis.dta", clear
isid tid school_year
```

// Step 2: Restrict the sample.

```
keep if t_is_teacher == 1
keep if !missing(school_poverty_quartile)
```

// Step 3: Review variables.

```
tab school_year school_poverty_quartile, mi row
bysort school_code school_year: gen tag = _n == 1
tab school_year school_poverty_quartile if tag, mi
drop tag
foreach var of varlist t_male t_black t_latino t_white t_newhire ///
       t_adv_degree t_certification_pathway {

       tab school_poverty_quartile `var', mi row
}
table school_poverty_quartile, c(mean t_experience)
```

// Step 4: Create binary variables for each school poverty quartile.

```
tab school_poverty_quartile, gen(school_poverty_q)
```

// Step 5: Create a binary variable for alternative certification.

```
gen t_alternative_certification = (t_certification_pathway > 1 & ///
       t_certification_pathway != .)
tab t_alternative_certification t_certification_pathway, mi
```

// Step 6: Get overall sample size.

```
summ tid
local teacher_years = string(r(N), "%9.0fc")
preserve
       bysort tid: keep if _n == 1
       summ tid
       local unique_teachers = string(r(N), "%9.0fc")
restore
```

// Step 7: Define row titles for the table.

```
local t_male "Percent Male"
local t_black "Percent African American"
local t_latino "Percent Latino"
local t_white "Percent White"
local t_newhire "Percent New Hires"
local t_adv_degree "Percent with Advanced Degree"
```

1. TEACHER CHARACTERISTICS BY SCHOOL POVERTY

```
local t_alternative_certification "Percent with Alternative Certification"
local t_experience "Average Teacher Experience"
```

// Step 8: Open output file.

```
tempvar tbl
file open `tbl' using "${graphs}\B1_Teacher_Char_by_School_Poverty.xls", ///
       write text replace
```

// Step 9: Write overall and column titles.

```
file write `tbl' "Teacher Characteristics by School Poverty Level"
file write `tbl' _newline
file write `tbl' _tab "Agency Average"
file write `tbl' _tab "High Poverty Schools"
file write `tbl' _tab "Low Poverty Schools"
file write `tbl' _tab "Difference between High and Low Poverty Schools"
file write `tbl' _newline
```

// Step 10: Start a loop through row variables.

```
foreach rowvar of varlist t_male t_black t_latino t_white t_newhire ///
       t_adv_degree t_alternative_certification t_experience {
```

// Step 11: Calculate quartile averages and difference.

```
       reg `rowvar' school_poverty_q4 school_poverty_q1, robust
       estimates store `rowvar'
       local highpov = _b[school_poverty_q4] + _b[_cons]
       local lowpov = _b[school_poverty_q1] + _b[_cons]
       local diff = _b[school_poverty_q4] - _b[school_poverty_q1]
```

// Step 12: Get significance for difference between top and bottom quartile.

```
       test school_poverty_q4 = school_poverty_q1
       gen star = ""
       replace star = "*" if r(p) < .05
```

// Step 13: Calculate agency average.

```
       quietly summ `rowvar', meanonly
       local agencyavg = r(mean)
```

// Step 14: Write values for high, low, difference, and significance for each row variable.

```
       file write `tbl' "``rowvar''"
       file write `tbl' _tab "`:di %9.3f round(`agencyavg', .001)'"
       file write `tbl' _tab "`:di %9.3f round(`highpov', .001)'"
       file write `tbl' _tab "`:di %9.3f round(`lowpov', .001)'"
       file write `tbl' _tab "`:di %9.3f round(`diff', .001)'"
       file write `tbl' _tab "`:di %3s star'"
```

1. TEACHER CHARACTERISTICS BY SCHOOL POVERTY

```
file write `tbl' _newline
drop star
}
```

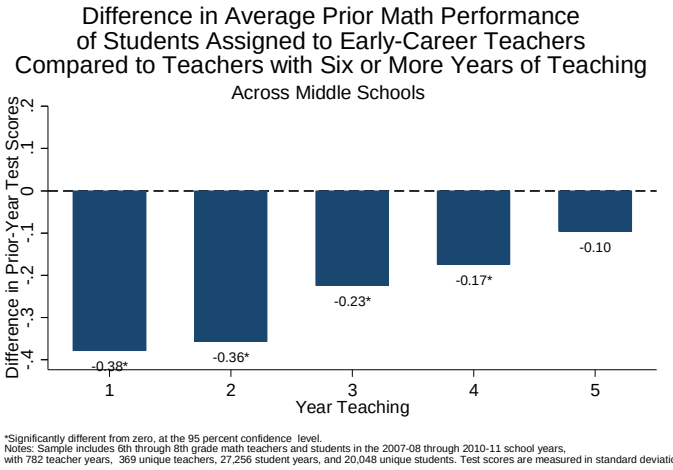
// Step 15: Write footnote including sample sizes.

```
#delimit ;
file write `tbl' "*Difference is statistically significant at the 95 percent confidence level.";
file write `tbl' _newline;
file write `tbl' "Notes: Sample includes teachers in the 2006-07 through 2010-11 school years,
with `teacher_years' teacher years and `unique_teachers' unique teachers.
High (low) poverty schools are in the top (bottom) quartile of schools each year based on the
share of students receiving free or reduced price lunch.";
#delimit cr
```

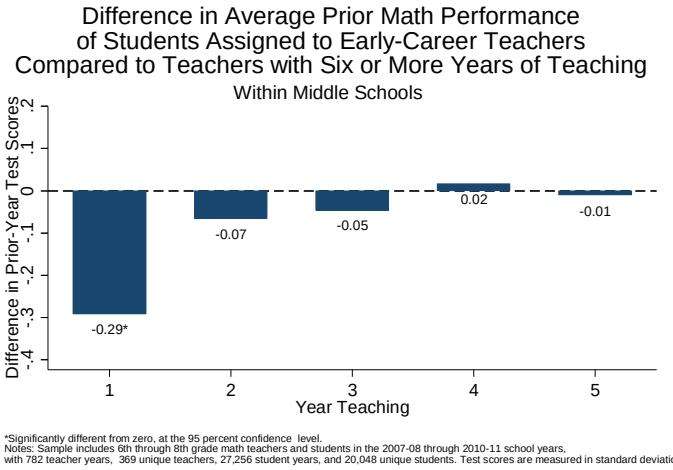
// Step 16: Close output file. Do table formatting in Excel.

```
file close _all
```

2. STUDENT PRIOR ACHIEVEMENT BY TEACHER EXPERIENCE



*Significantly different from zero, at the 95 percent confidence level.
Notes: Sample includes 6th through 8th grade math teachers and students in the 2007-08 through 2010-11 school years, with 782 teacher years, 369 unique teachers, 27,256 student years, and 20,048 unique students. Test scores are measured in standard deviations.



*Significantly different from zero, at the 95 percent confidence level.
Notes: Sample includes 6th through 8th grade math teachers and students in the 2007-08 through 2010-11 school years, with 782 teacher years, 369 unique teachers, 27,256 student years, and 20,048 unique students. Test scores are measured in standard deviations.

Purpose:

Examine how students are placed with teachers based on students' prior performance and teachers' experience.

Required analysis file variables:

```
sid
school_year
school_code
tid_math
cid_math
grade_level
std_scaled_score_math_tm1
t_experience
```

Analysis-specific sample restrictions:

- Restrict to classes included in value-added estimates.
- Keep only grades and years for which prior-year test scores are available.
- Keep only students with a single identified current-year core course and prior-year test scor in the given subject.
- If school-level restriction is chosen, keep only records for either elementary or middle schoool grades.

Ask yourself:

- To what degree are the placement patterns driven by the concentration of novice teachers and lower-performing students in certain schools?
- Do internal school politics influence placement patterns? Are there formal or informal arrangements that enable more senior teachers to choose their classroom assignments? Is there a norm within the agency that novice teachers need to "put in their time" with more difficult assignments? Do parents of higher-achieving students influence placements to well-known teachers?
- Are timing factors important? Are classroom rosters drawn up early in the summer? Are students who enroll late assigned to teachers hired just prior to the school year?
- Are within-school gaps concentrated in certain schools? Are there some schools in which novice teachers are actually assigned to higher-achieving students?

Potential further analyses:

- This graph could be created for individual schools.
- If meaningful placement patterns are observed in middle school, examine how courses teachers are assigned (e.g., remedial, honors, etc.) vary based on teaching experience.

2. STUDENT PRIOR ACHIEVEMENT BY TEACHER EXPERIENCE

// Step 1: Choose the subject (math or ela) and school level (elem or middle) for the analysis. Note: to make multiple charts at the same time, put loops for subject and level around the analysis and graphing code. To include all grade levels in the analysis, comment out the local level command below.

```
local subject math
local level middle
```

// Step 2: Load the Student_Teacher_Year_Analysis data file.

```
use "${analysis}\Student_Teacher_Year_Analysis.dta", clear
isid sid school_year
```

// Step 3: Restrict the sample. Keep grades and years for which prior-year test scores are available. Keep students with teachers with non-missing values for experience. Keep students with a single identified current-year core course and prior-year test score in the given subject. If school level restriction is chosen, keep only records from either elementary or middle school grades.

```
keep if school_year >= 2008 & school_year <= 2011
keep if grade_level >= 5 & grade_level <= 8
keep if t_is_teacher == 1
keep if !missing(t_experience)
keep if !missing(cid_`subject')
keep if !missing(std_scaled_score_`subject'_tm1)
if "`level'" == "elem" {
       keep if grade_level == 5
}
if "`level'" == "middle" {
       keep if grade_level >= 6 & grade_level <= 8
}
```

// Step 4: Review teacher and student variables.

```
tab school_year grade_level, mi
unique tid_`subject'
unique tid_`subject school_year
bysort tid_`subject' school_year: gen tag = (_n == 1)
tab t_experience if tag == 1, mi
drop tag
table t_experience, c(mean std_scaled_score_`subject'_tm1)
codebook cid_`subject' tid_`subject' school_year t_experience ///
       std_scaled_score_`subject' std_scaled_score_`subject'_tm1 ///
       grade_level school_code
```

2. STUDENT PRIOR ACHIEVEMENT BY TEACHER EXPERIENCE

// Step 5: Define dummy variables for teaching experience. Put teachers with six or more years of experience into one category.

```
replace t_experience = 6 if t_experience > 5 & !missing(t_experience)
tab t_experience, gen(exp)
```

// Step 6: Define grade-by-year variable for fixed effects.

```
egen grade_year = group(grade_level school_year)
```

// Step 7: Define grade-by-year-by-school variable for fixed effects.

```
egen grade_year_school = group(school_code grade_level school_year)
```

// Step 8: Define 5 x 4 matrix to store results

```
matrix results = J(5, 4, .)
matrix colnames results = across_school_coef across_school_se ///
       within_school_coef within_school_se
```

// Step 9: Do regression of prior student math score on teacher experience. Sixth plus year teachers are comparison group.

```
areg std_scaled_score_`subject'_tm1 exp1-exp5, robust cluster(cid_`subject') ///
       absorb(grade_year)
```

// Step 10: Get student and teacher sample sizes.

```
egen teacher_years = nvals(tid_`subject' subject_year) if e(sample)
summ teacher_years
local teacher_years = string(r(mean), "%9.0fc")
egen unique_teachers = nvals(tid_`subject') if e(sample)
summ unique_teachers

egen student_years = nvals(sid school_year) if e(sample)
summ student_years
local student_years = string(r(mean), "%9.0fc")
egen unique_students = nvals(sid) if e(sample)
summ unique_teachers
local unique_students = string(r(mean), "%9.0fc")
```

2. STUDENT PRIOR ACHIEVEMENT BY TEACHER EXPERIENCE

// Step 11: Store coefficients in the first column and and standard errors in the second column of the results matrix.

```
matrix results[1, 1] = _b[exp1]
matrix results[2, 1] = _b[exp2]
matrix results[3, 1] = _b[exp3]
matrix results[4, 1] = _b[exp4]
matrix results[5, 1] = _b[exp5]

matrix results[1, 2] = _se[exp1]
matrix results[2, 2] = _se[exp2]
matrix results[3, 2] = _se[exp3]
matrix results[4, 2] = _se[exp4]
matrix results[5, 2] = _se[exp5]
```

// Step 12: Add school fixed effects to prior model to generate within school comparison.

```
areg std_scaled_score_`subject'_tm1 exp1-exp5, robust cluster(cid_`subject') ///
       absorb(grade_year_school)
```

// Step 13: Store coefficients and standard errors in columns 3 and 4 of results matrix.

```
matrix results[1, 3] = _b[exp1]
matrix results[2, 3] = _b[exp2]
matrix results[3, 3] = _b[exp3]
matrix results[4, 3] = _b[exp4]
matrix results[5, 3] = _b[exp5]

matrix results[1, 4] = _se[exp1]
matrix results[2, 4] = _se[exp2]
matrix results[3, 4] = _se[exp3]
matrix results[4, 4] = _se[exp4]
matrix results[5, 4] = _se[exp5]
```

// Step 14: Clear data and replace with matrix contents. Add variable for year teaching.

```
clear
svmat results, names(col)
gen year_teaching = _n
```

2. STUDENT PRIOR ACHIEVEMENT BY TEACHER EXPERIENCE

// Step 15: Get and store significance. Concatenate coefficient and asterisk to use as value label.

```
foreach model in across_school within_school {
	gen `model'_sig = abs(`model'_coef / `model'_se)
}

foreach var of varlist across_school_sig within_school_sig {
	replace `var' = 0 if `var' < =1.96
	replace `var' = 1 if `var' > 1.96
	tostring `var', replace
	replace `var' = "" if `var' == "0"
	replace `var' = "*" if `var' == "1"
}

foreach model in across_school within_school {
	gen `model'_string = string(`model'_coef, "%9.2f")
	egen `model'_label = concat(`model'_string `model'_sig)
}
```

// Step 16: Define titles for subject and school level.

```
if "`subject'" == "math" {
	local subj_title "Math"
	local subj_foot "math"
}

if "`subject'"=="ela" {
	local subj_title "ELA"
	local subj_foot "English/Language Arts"
}

local gradespan "5th through 8th"

if "`level'" == "middle" {
	local level_title "Middle "
	local gradespan "6th through 8th"
}
if "`level'" == "elem" {
	local level_title "Elementary "
	local gradespan "5th"
}
```

// Step 17: Start loop through models to make and save across and within schools charts.

```
	foreach model in across within {
```

2. STUDENT PRIOR ACHIEVEMENT BY TEACHER EXPERIENCE

// Step 18: Define subtitle.

```
if "`model'" == "across" {
       local subtitle "Across `level_title'Schools"
}

if "`model'" == "within" {
       local subtitle "Within `level_title'Schools"
}
```

// Step 19: Make chart. Bar chart gives average score difference relative to 6th plus year teachers, while scatter plot places value and significance asterisk as marker label below bar. Marker symbol is invisible.

```
#delimit ;
twoway bar `model'_school_coef year_teaching,
       barwidth(.6) color(navy) finten(100) ||
scatter `model'_school_coef year_teaching,
       mlabel(`model'_school_label)
       msymbol(i)
       mlabpos(6)
       mlabcolor(black) ||,

       ytitle("Difference in Prior-Year Test Scores", size(medsmall))
       title("Difference in Average Prior `subj_title' Performance"
              "of Students Assigned to Early-Career Teachers"
              "Compared to Teachers with Six or More Years of Teaching", span)
       subtitle("`subtitle'", span)
       xtitle("Year Teaching", size(medsmall))
       xlabel(,labsize(medsmall))
       legend(off)

       yline(0, lpattern(dash) lcolor(black))
       yscale(range(-.4 .2))
       ylabel(-.4(.1).2, nogrid labsize(medsmall))
       ytick(-.4(.1).2)
       graphregion(color(white) fcolor(white) lcolor(white))
       plotregion(color(white) fcolor(white) lcolor(white) margin(5 5 2 0))

       note(" " "*Significantly different from zero, at the 95 percent confidence
level." "Notes: Sample includes `gradespan' grade `subj_foot' teachers and
students in the 2007-08 through 2010-11 school years," "with `teacher_years'
teacher years, `unique_teachers' unique teachers, `student_years' student
 years, and `unique_students' unique students. Test scores are measured in
standard deviations.", size(vsmall) span);

#delimit cr
```

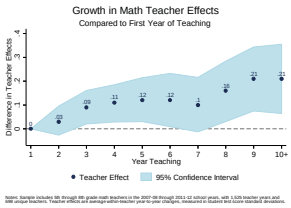
2. STUDENT PRIOR ACHIEVEMENT BY TEACHER EXPERIENCE

// Step 20: Save chart. If marker labels need to be moved by hand using Stata Graph Editor, re-save .gph and .emf files after editing.

```
graph export "${graphs}/B2_Prior_Ach_by_Exp_`subtitle'_`subj_title'.emf", replace
graph save "${graphs}/B2_Prior_Ach_by_Exp_`subtitle'_`subj_title'.gph", replace
}
```

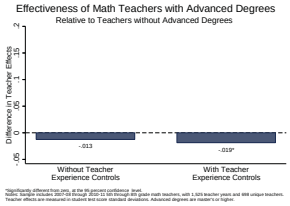
C. Development

Often, teachers are an agency's biggest investment. Once teachers have been recruited and placed in schools, their continued professional development benefits students and improves the success of the agency. Traditionally, agencies have incentivized two major forms of professional development: learning over time from experience and earning a graduate degree. This section of the diagnostic examines ways teachers develop during their careers and identifies whether agency incentives are aligned with gains in teacher effectiveness.



1. GROWTH IN TEACHER EFFECTS FOR EARLY CAREER TEACHERS

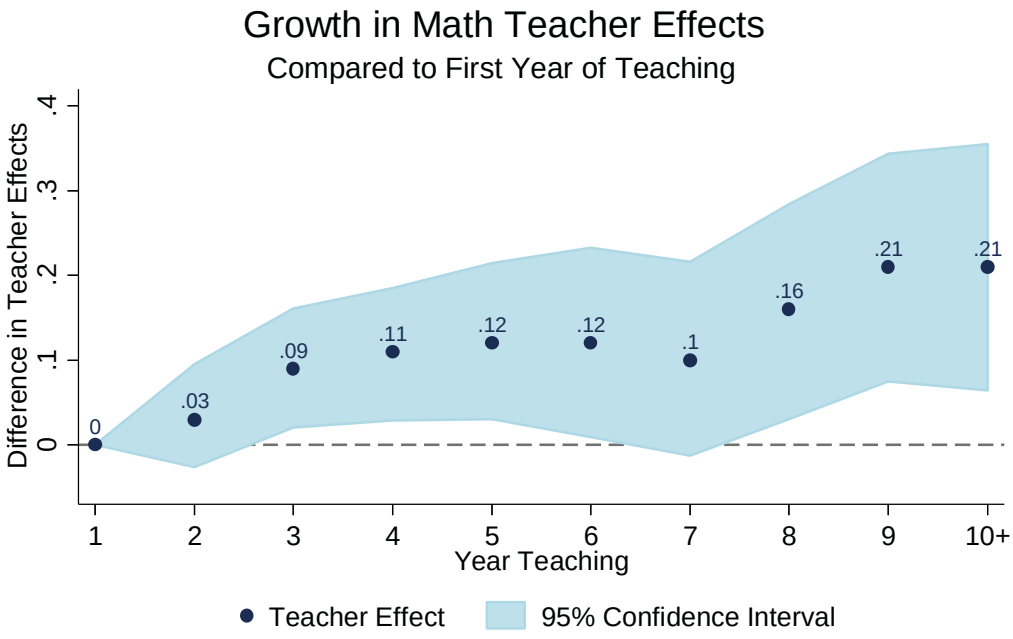
Describes the growth in effectiveness for early-career teachers.



2. DIFFERENCE IN TEACHER EFFECTS FOR TEACHERS WITH AND WITHOUT ADVANCED DEGREES

Examines the difference in effectiveness for teachers with and without advanced degrees.

1. GROWTH IN TEACHER EFFECTS FOR EARLY CAREER TEACHERS



Notes: Sample includes 5th through 8th grade math teachers in the 2007-08 through 2011-12 school years, with 1,525 teacher years and 698 unique teachers. Teacher effects are average within-teacher year-to-year changes, measured in student test score standard deviations.

Purpose:
Observe how teachers' effectiveness estimates change as they gain teaching experience.

Required analysis file variables:

```
sid
school_year
tid_math
cid_math
grade_level
t_is_teacher
t_experience
std_scaled_score_math
std_scaled_score_math_tm1
std_scaled_score_ela_tm1
```

(See full list of variables in Connect):
student characteristics vector
class characteristics vector
cohort characteristics vector

Analysis-specific sample restrictions:

- Keep only grades and years for which prior-year test scores are available.
- Keep only students with a single identified current-year core course and current and prior-year test scores in the given subject.
- If school-level restriction is chosen, keep only records for either elementary or middle school grades.

Potential further analyses:

- If your agency changed induction and/or early career programs and/or policies, conduct this analysis separately for teachers who were and were not affected by the program or policy.
- Examine other dimensions of teacher effectiveness over time (e.g., trends in student survey results).

Ask yourself:

- Teacher salary schedules often compensate teachers for their teaching experience. How does your agency compensate teachers for experience? What are some ways that salary schedules could better align to increases in student outcomes over time?
- What induction and early career supports do novice teachers have? Do they vary by school, level of instruction (elementary, middle, high), and/or content area? Are the growth trajectories you see most related to recruitment practices, early career supports, or both?

1. GROWTH IN TEACHER EFFECTS FOR EARLY CAREER TEACHERS

// Step 1: Choose the subject (math or ela) and school level (elem or middle) for the analysis. Note: To change from math to ELA, switch the subjects in the next two lines. To make multiple charts at the same time, put loops for subject and level around the analysis and graphing code. To include all grade levels in the analysis, comment out the local level command below.

```
local subject math
local alt_subject ela
*local level middle
```

// Step 2: Load the Student_Teacher_Year_Analysis data file.

```
use "${analysis}\Student_Teacher_Year_Analysis.dta", clear
isid sid school_year
```

// Step 3: Restrict the sample. Keep grades and years for which prior-year test scores are available. Keep students with teachers with non-missing values for experience. Keep students with a single identified core course and current and prior-year test scores in the given subject. If school level restriction is chosen, keep only records from either elementary or middle school grades.

```
keep if school_year >= 2008 & school_year <= 2011
keep if grade_level >= 5 & grade_level <= 8
keep if t_is_teacher == 1
keep if !missing(t_experience)
keep if !missing(cid_`subject')
keep if !missing(std_scaled_score_`subject', std_scaled_score_`subject'_tm1)
if "`level'" == "elem" {
       keep if grade_level == 5
}
if "`level'" == "middle" {
       keep if grade_level >= 6 & grade_level <= 8
}
```

// Step 4: Review teacher variables.

```
tab school_year
unique tid_`subject'
unique tid_`subject' school_year
tab t_experience t_novice, mi
bysort tid_`subject' school_year: gen tag = (_n == 1)
tab t_experience if tag == 1, mi
drop tag
```

// Step 5: Create dummy variables for each year of teaching experience, putting all teachers with 10 or more years of experience in one group.

```
replace t_experience = 10 if t_experience >= 10
tab t_experience, gen(exp)
```

1. GROWTH IN TEACHER EFFECTS FOR EARLY CAREER TEACHERS

// Step 6: Create variable for grade-by-year fixed effects.

```
egen grade_by_year = group(grade_level school_year)
```

// Step 7: Create variables for previous year's score squared and cubed.

```
gen std_scaled_score_`subject'_tm1_sq = std_scaled_score_`subject'_tm1^2
gen std_scaled_score_`subject'_tm1_cu = std_scaled_score_`subject'_tm1^3
```

// Step 8: Create indicator for whether student is missing prior achievement for alternate subject. Make a replacement variable that imputes score to zero if missing.

```
gen miss_std_scaled_score_`alt_subject'_tm1 = ///
       missing(std_scaled_score_`alt_subject'_tm1)
gen _IMPstd_scaled_score_`alt_subject'_tm1 = std_scaled_score_`alt_subject'_tm1
replace _IMPstd_scaled_score_`alt_subject'_tm1 = 0 ///
       if miss_std_scaled_score_`alt_subject'_tm1 == 1
```

// Step 9: Identify prior achievement variables to use as controls.

```
#delimit ;
local prior_achievement
       "std_scaled_score_`subject'_tm1
       std_scaled_score_`subject'_tm1_sq
       std_scaled_score_`subject'_tm1_cu
       _IMPstd_scaled_score_`alt_subject'_tm1
       miss_std_scaled_score_`alt_subject'_tm1";
#delimit cr
```

// Step 10: Identify other student variables to use as controls.

```
#delimit;
local student_controls
       "s_male
       s_black
       s_asian
       s_latino
       s_naam
       s_mult
       s_racemiss
       s_reducedlunch
       s_freelunch
       s_lunch_miss
       s_retained
       s_retained_miss
       s_gifted
       s_gifted_miss
       s_iep
       s_iep_miss
       s_ell
       s_ell_miss
```

1. GROWTH IN TEACHER EFFECTS FOR EARLY CAREER TEACHERS

```
       s_absence_high
       s_absence_miss";
#delimit cr
```

// Step 11: Review all variables to be included in the teacher effectiveness model. Class and cohort (grade/school/year) variables should include means of all student variables, and means, standard deviations, and percent missing for prior-year test scores for both main and alternate subject. Class and cohort size should also be included as controls.

```
codebook std_scaled_score_`subject' exp1-exp10
codebook `prior_achievement'
codebook `student_controls'
codebook _CL*`subject'*
codebook _CO*
codebook grade_by_year cid_`subject'
```

// Step 12: Estimate growth in teacher effectiveness relative to novice teachers, using within-teacher fixed effects.

```
areg std_scaled_score_`subject' exp2-exp10 `prior_achievement' `student_controls' ///
       _CL*`subject'* _CO* i.grade_by_year, absorb(tid_`subject') cluster(cid_`subject')
```

// Step 13: Store coefficients and standard errors.

```
forval year = 2/10 {
       gen coef_exp`year' = _b[exp`year']
       gen se_exp`year' = _se[exp`year']
}
```

// Step 14: Set values to zero for novice comparison teachers.

```
gen coef_exp1 = 0 if exp1 == 1
gen se_exp1 = 0 if exp1 == 1
```

// Step 15: Get teacher sample size.

```
egen teacher_years = nvals(tid_`subject' school_year) if e(sample)
summ teacher_years
local teacher_years = string(r(mean), "%9.0fc")
egen unique_teachers = nvals(tid_`subject') if e(sample)
summ unique_teachers
local unique_teachers = string(r(mean), "%9.0fc")
```

1. GROWTH IN TEACHER EFFECTS FOR EARLY CAREER TEACHERS

// Step 16: Collapse and reshape data for graph.

```
collapse (max) coef_exp* se_exp*
gen results = 1
reshape long coef_exp se_exp, i(results) j(year_teaching)
rename coef_exp coef
rename se_exp se
```

// Step 17: Generate confidence intervals of the estimated returns to experience.

```
gen conf_hi = coef + (se * 1.96)
gen conf_low = coef - (se * 1.96)
replace coef = round(coef,.01)
```

// Step 18: Define subject and school level titles for graph.

```
if "`subject'" == "math" {
       local subj_foot "math"
       local subj_title "Math"
}

if "`subject'" == "ela" {
       local subj_foot "English/Language Arts"
       local subj_title "ELA"
}

local gradespan "5th through 8th"

if "`level'" == "middle" {
       local subj_title "Middle School `subj_title'"
       local gradespan "6th through 8th"
}

if "`level'" == "elem" {
       local subj_title "Elementary School `subj_title'"
       local gradespan "5th"
}
```

// Step 19: Make chart.

```
#delimit ;
twoway rarea conf_hi conf_low year_teaching if year_teaching <= 10,
       sort
       color(ltblue) ||

       scatter coef year_teaching,
              mlab(coef) mlabposition(12) mcolor(dknavy) mlabcolor(dknavy)
              yline(0, lcolor(gs7) lpattern(dash))
              yscale(range(-.05(.05).3))
              ylabel(0(.1).4, labsize(medsmall) nogrid)
              ytick(0(.1).4) ||,
```

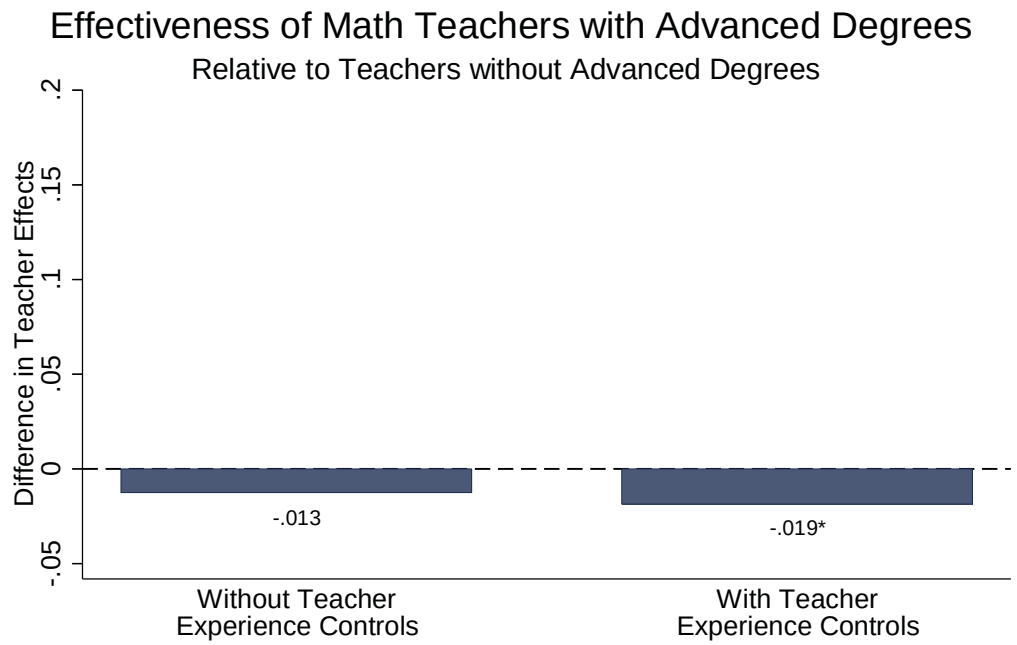
1. GROWTH IN TEACHER EFFECTS FOR EARLY CAREER TEACHERS

```
        graphregion(color(white) fcolor(white) lcolor(white))
        plotregion(color(white) fcolor(white) lcolor(white))

        title("Growth in `subj_title' Teacher Effects", span)
        subtitle("Compared to First Year of Teaching", span)
        ytitle("Difference in Teacher Effects", size(medsmall))
        legend(order(2 1 3)
        label(2 "Teacher Effect")
        label(1 "95% Confidence Interval"))
        legend(cols(2) symxsize(5) ring(1) region(lstyle(none) lcolor(none) color(none)))
        xtitle("Year Teaching")
        xtick(1(1)10)
        xscale(range(1(1)10))
        xlabel(1 "1" 2 "2" 3 "3" 4 "4" 5 "5" 6 "6" 7 "7" 8 "8" 9 "9" 10 "10+")

note(" " "Notes: Sample includes `gradespan' grade `subj_foot' teachers
in the 2007-08 through 2011-12 school years, with `teacher_years' teacher years and"
"`unique_teachers' unique teachers. Teacher effects are average within-teacher
year-to-year changes, measured in student test score standard deviations.", size(vsmall) span);

#delimit cr
```

// Step 20: Save chart. If marker labels need to be moved by hand using Stata Graph Editor, re-save .gph and .emf files after editing.

```
graph export "${graphs}/C1_Returns_to_Teaching_Experience_`subj_title'.emf", replace
graph save "${graphs}/C1_Returns_to_Teaching_Experience_`subj_title'.gph", replace
```

2. DIFFERENCE IN TEACHER EFFECTS FOR TEACHERS WITH AND WITHOUT ADVANCED DEGREES



*Significantly different from zero, at the 95 percent confidence level.
Notes: Sample includes 2007-08 through 2010-11 5th through 8th grade math teachers, with 1,525 teacher years and 698 unique teachers. Teacher effects are measured in student test score standard deviations. Advanced degrees are master's or higher.

Purpose:
Determine whether there are differences in effectiveness estimates for teachers with and without advanced degrees.

Required analysis file variables:

```
sid
school_year
tid_math
cid_math
grade_level
t_is_teacher
t_adv_degree
t_experience
std_scaled_score_math
std_scaled_score_math_tm1
std_scaled_score_ela_tm1
```

(See full list of variables in Connect):
student characteristics vector
class characteristics vector
cohort characteristics vector

Analysis-specific sample restrictions:

- Keep only grades and years for which prior-year test scores are available.
- Keep only students with a single identified current-year core course and current and prior-year test scores in the given subject.

Ask yourself:

- How are teachers compensated for education beyond a bachelor's degree in your agency? Does your agency subsidize master's degrees as well as reward teachers who have them with higher salaries? What are some ways in which your agency can strengthen the link between compensation and actual performance?
- What alternatives can your agency offer to the master's degree bonus? How can your agency customize professional development for the individual teachers' areas for improvement, and reward teachers for progress toward their professional goals and student outcomes?

Potential further analyses:

- If degree title is available, separate teachers into groups by type of degree (e.g., administrative, content area that does or does not match subject taught).

2. DIFFERENCE IN TEACHER EFFECTS FOR TEACHERS WITH AND WITHOUT ADVANCED DEGREES

// Step 1: Choose a subject for the analysis. Note: To change from math to ELA, switch the subjects in the next two lines. To generate ELA and math charts at the same time, enclose the analysis code within a loop.

```
local subject math
local alt_subject ela
```

// Step 2: Load the Student_Teacher_Year_Analysis data file.

```
use "${analysis}\Student_Teacher_Year_Analysis.dta", clear
isid sid school_year
```

// Step 3: Restrict the sample. Keep grades and years for which prior-year test scores are available. Keep students with teachers with non-missing values for experience and degree information. Keep students with a single identified core course and current and prior-year test scores in the given subject.

```
keep if school_year >= 2008 & school_year <= 2011
keep if grade_level >= 5 & grade_level <= 8
keep if t_is_teacher == 1
keep if !missing(t_adv_degree)
keep if !missing(cid_`subject')
keep if !missing(std_scaled_score_`subject', std_scaled_score_`subject'_tm1)
```

// Step 4: Review teacher variables.

```
tab school_year
unique tid_`subject'
unique tid_`subject' school_year
bysort tid_`subject' school_year: gen tag = #(_n == 1)
tab t_experience t_adv_degree if tag == 1, mi
drop tag
```

// Step 5: Create dummy variables for each year of teaching experience.

```
tab t_experience, gen(exp)
```

// Step 6: Create variable for grade-by-year fixed effects.

```
egen grade_by_year = group(grade_level school_year)
```

// Step 7: Create variables for previous year's score squared and cubed.

```
gen std_scaled_score_`subject'_tm1_sq = std_scaled_score_`subject'_tm1^2
gen std_scaled_score_`subject'_tm1_cu = std_scaled_score_`subject'_tm1^3
```

// Step 8: Create indicator for whether student is missing prior achievement for alternate subject. Make a replacement variable that imputes score to zero if missing.

```
gen miss_std_scaled_score_`alt_subject'_tm1 = ///
       missing(std_scaled_score_`alt_subject'_tm1)
```

2. DIFFERENCE IN TEACHER EFFECTS FOR TEACHERS WITH AND WITHOUT ADVANCED DEGREES

```
gen _IMPstd_scaled_score_`alt_subject'_tm1 = std_scaled_score_`alt_subject'_tm1
replace _IMPstd_scaled_score_`alt_subject'_tm1 = 0 ///
       if miss_std_scaled_score_`alt_subject'_tm1 == 1
```

// Step 9: Identify prior achievement variables to use as controls.

```
#delimit ;
local prior_achievement
       "std_scaled_score_`subject'_tm1
       std_scaled_score_`subject'_tm1_sq
       std_scaled_score_`subject'_tm1_cu
       _IMPstd_scaled_score_`alt_subject'_tm1
       miss_std_scaled_score_`alt_subject'_tm1";
#delimit cr
```

// Step 10: Identify other student variables to use as controls.

```
#delimit;
local student_controls
       "s_male
       s_black
       s_asian
       s_latino
       s_naam
       s_mult
       s_racemiss
       s_reducedlunch
       s_freelunch
       s_lunch_miss
       s_retained
       s_retained_miss
       s_gifted
       s_gifted_miss
       s_iep
       s_iep_miss
       s_ell
       s_ell_miss
       s_absence_high
       s_absence_miss";
#delimit cr
```

// Step 11: Review all variables to be included in the teacher effectiveness models. Class and cohort (grade/school/year) variables should include means of all student variables, and means, standard deviations, and percent missing for prior-year test scores for both main and alternate subject. Class and cohort size should also be included as controls.

```
codebook std_scaled_score_`subject' t_adv_degree exp*
codebook `prior_achievement'
codebook `student_controls'
codebook _CL*`subject'*
codebook _CO*
codebook grade_by_year cid_`subject'
```

2. DIFFERENCE IN TEACHER EFFECTS FOR TEACHERS WITH AND WITHOUT ADVANCED DEGREES

// Step 12: Estimate differences in teacher effectiveness between teachers with and without advanced degrees, without teacher experience controls.

```
reg std_scaled_score_`subject' t_adv_degree ///
       `student_controls' `prior_achievement' _CL*`subject'* _CO* ///
       i.grade_by_year, cluster(cid_`subject')
```

// Step 13: Store coefficient and standard error.

```
gen coef_noexp = _b[t_adv_degree]
gen se_noexp = _se[t_adv_degree]
```

// Step 14: Get teacher sample size for this model.

```
egen teachers_in_sample_noexp = nvals(tid_`subject') if e(sample)
summ teachers_in_sample_noexp
local teachers_in_sample_noexp = r(mean)
```

// Step 15: Estimate differences in teacher effectiveness between teachers with and without advanced degrees, with teacher experience controls.

```
reg std_scaled_score_`subject' t_adv_degree exp* ///
       `student_controls' `prior_achievement' _CL*`subject'* _CO* ///
       i.grade_by_year, cluster(cid_`subject')
```

// Step 16: Store coefficient and standard error.

```
gen coef_wexp = _b[t_adv_degree]
gen se_wexp = _se[t_adv_degree]
```

// Step 17: Get teacher sample size for this model and compare sample size for the two models.

```
egen teachers_in_sample_wexp = nvals(tid_`subject') if e(sample)
summ teachers_in_sample_wexp
local teachers_in_sample_wexp = r(mean)
assert `teachers_in_sample_wexp' == `teachers_in_sample_noexp'
```

// Step 18: Store teacher sample size for footnote.

```
egen teacher_years = nvals(tid_`subject' school_year) if e(sample)
summ teacher_years
local teacher_years = string(r(mean), "%9.0fc")
egen unique_teachers = nvals(tid_`subject') if e(sample)
summ unique_teachers
local unique_teachers = string(r(mean), "%9.0fc")
```

// Step 19: Collapse dataset for graphing.

```
collapse(max) coef* se*
```

2. DIFFERENCE IN TEACHER EFFECTS FOR TEACHERS WITH AND WITHOUT ADVANCED DEGREES

// Step 20: Get significance.

```
foreach spec in noexp wexp {
	gen sig_`spec' = abs(coef_`spec') - 1.96 * se_`spec' > 0
}
```

// Step 21: Reshape for graphing.

```
gen results = 1
reshape long coef_ se_ sig_, i(results) j(spec) string
rename coef_ coef
rename se_ se
rename sig_ sig
replace spec = "1" if spec == "noexp"
replace spec = "2" if spec == "wexp"
destring spec, replace
```

// Step 22: Make value labels with significance indicator.

```
tostring sig, replace
replace sig = "*" if sig == "1"
replace sig = ""  if sig == "0"
replace coef = round(coef,.001)
egen coef_label = concat(coef sig)
```

// Step 23: Define subject titles for graph.

```
if "`subject'" == "math" {
	local subject_foot "math"
	local subj_title "Math"
}

if "`subject'" == "ela" {
	local subject_foot "English/Language Arts"
	local subj_title "ELA"
}
```

// Step 24: Create a bar graph of the estimation results.

```
#delimit ;
graph twoway (bar coef spec,
		fcolor(dknavy) lcolor(dknavy) lwidth(0) barwidth(0.7))
	(scatter coef spec,
		mcolor(none) mlabel(coef_label) mlabcolor(black) mlabpos(12)
		mlabsize(small)),
	yline(0, style(extended) lpattern(dash) lwidth(medthin) lcolor(black))
	title("Effectiveness of `subj_title' Teachers with Advanced Degrees",
	span)
```

2. DIFFERENCE IN TEACHER EFFECTS FOR TEACHERS WITH AND WITHOUT ADVANCED DEGREES

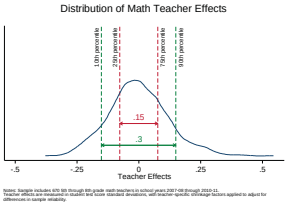
```
        subtitle("Relative to Teachers without Advanced Degrees", span)
        ytitle("Difference in Teacher Effects", size(medsmall))
        yscale(range(-.05 .2))
        ytick(-.05(.05).2)
        ylabel(-.05(.05).2, nogrid)
        xlabel("", notick)
        xtitle("")
        xlabel(1 `""Without Teacher" "Experience Controls""'
               2 `""With Teacher" "Experience Controls""', labsize(medsmall))
        legend(off)
        graphregion(color(white) fcolor(white) lcolor(white))
        plotregion(color(white) fcolor(white) lcolor(white) margin(5 5 2 0))
        note(" " "*Significantly different from zero, at the 95 percent confidence
level." "Notes: Sample includes 2007-08 through 2010-11 5th through 8th grade `subj_
foot' teachers, with `teacher_years' teacher years and `unique_teachers' unique
teachers." "Teacher effects are measured in student test score standard deviations. Advanced
degrees are master's or higher.", size(vsmall) span);

#delimit cr
```

// Step 25: Save chart. If marker labels need to be moved by hand using Stata Graph Editor, re-save .gph and .emf files after editing.

```
graph export "${graphs}/C2_Teacher_Effects_Advanced_Degree_`subj_title'.emf", replace
graph save "${graphs}/C2_Teacher_Effects_Advanced_Degree_`subj_title'.gph", replace
```

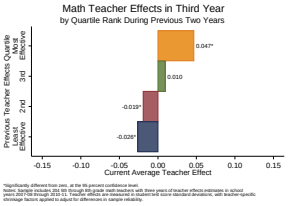
D. EVALUATION

Evaluating teachers serves two purposes: identifying areas where professional development is likely to benefit teachers, and identifying effective and ineffective teachers for career decisions. Evaluations can involve administrator observation, peer observation, collections of classroom preparation materials and artifacts, student surveys, and student achievement data. In the human capital diagnostic, SDP estimates teachers' effectiveness at raising student achievement using value-added methodology to evaluate teacher performance. Your agency may use a difference student growth based measure. A good measure of teacher effectiveness will have sufficient variation and consistency. That is, teacher effectiveness ratings are spread out across the range of possible values enough to observe differences across groups, and teachers' ratings are fairly well correlated over time. The Evaluation section of the diagnostic examines the extent to which effectiveness estimates meet these criteria.



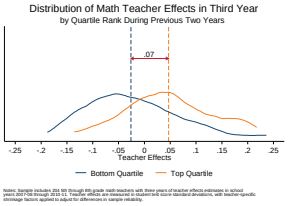
1. DISTRIBUTION OF TEACHER EFFECTS

Shows the overall distribution of a teacher effectiveness measure.



2. AVERAGE TEACHER EFFECTS IN THIRD YEAR BY QUARTILE RANK DURING PREVIOUS TWO YEARS

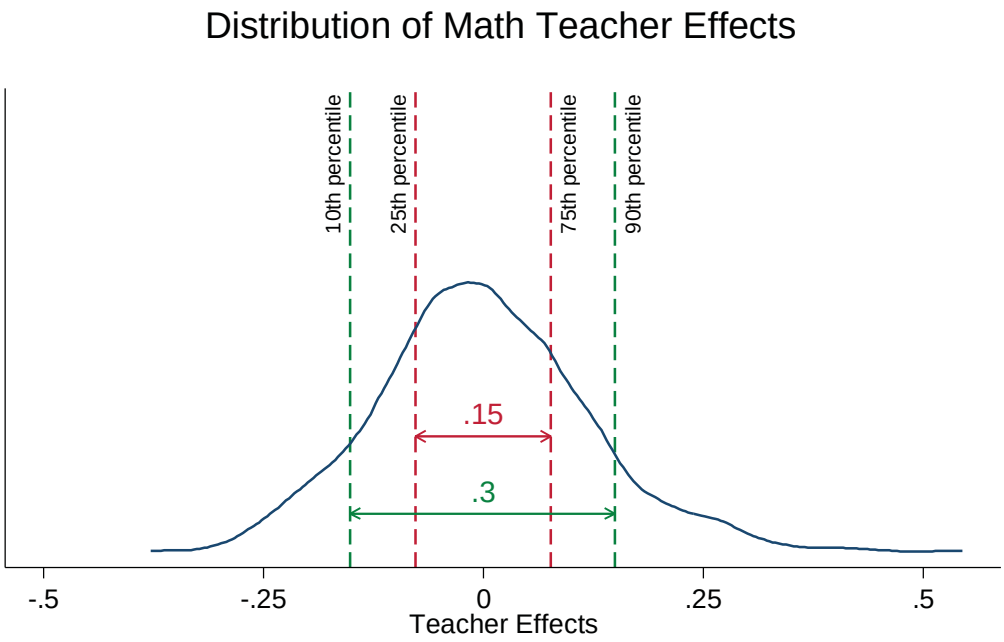
Examines whether two years of teacher effectiveness measures are predictive of average teacher effectiveness in a third year.



3. DISTRIBUTION OF TEACHER EFFECTS IN THIRD YEAR BY QUARTILE RANK DURING PREVIOUS TWO YEARS

Examines the distribution of teacher effectiveness in a third year for teachers ranked most and least effective in the prior two years.

1. DISTRIBUTION OF TEACHER EFFECTS



Notes: Sample includes 670 5th through 8th grade math teachers in school years 2007-08 through 2010-11.
Teacher effects are measured in student test score standard deviations, with teacher-specific shrinkage factors applied to adjust for differences in sample reliability.

Purpose:
Examine the distribution of teacher effectiveness estimates.

Required analysis file variables:

```
tid
school_year
t_is_teacher
school_lvl
tre_math
```

Analysis-specific sample restrictions:

- Keep only records for teachers with effectiveness estimates in the given subject.
- If school_level restriction is chosen, keep only records for elementary or middle school teachers.

Ask yourself:

- What measures of teacher effectiveness does your agency currently use? Compared to a measure that has little variation, what are the advantages of a measure of teacher effectiveness that has a lot of variation when making decisions about professional development, promotions to teacher leadership, and retention?
- What dimensions of teacher effectiveness do value-added estimates measure? What else is important to know about teacher quality? How are other dimensions of teacher quality measured in your agency?
- In what ways can knowledge of a teacher's value-added score be used to guide improvement? What kind of training would administrators and teachers need to be able to use the data to improve student achievement?

Potential further analyses:
If more than one measure of teacher effectiveness is recorded in your agency, produce a kdensity graph for each one. Correlate two measures of teacher effectiveness (e.g., value-added estimates and classroom observation ratings). Create a scatterplot with the categories of performance on one axis and value-added ratings on the other. Examine the range of value-added estimates within each performance category.

1. DISTRIBUTION OF TEACHER EFFECTS

// Step 1: Choose the subject (math or ela) and school level (elem or middle) for the analysis. Note: to make multiple charts at the same time, put loops for subject and level around the analysis and graphing code. To include all grade levels in the analysis, comment out the local level command below.

```
local subject math
*local level middle
```

// Step 2: Load the Teacher_Year_Analysis file containing value-added estimates.

```
use "${analysis}/Teacher_Year_Analysis.dta", clear
isid tid school_year
```

// Step 3: Restrict the sample. Keep years for which teacher effects value added estimates are available. Keep only employees who are teachers. Keep only records for which teachers have pooled teacher effects estimates (pooled estimates use information from all available years for each teacher). If school level restriction is chosen, keep only records from either elementary or middle schools.

```
keep if school_year >= 2008 & school_year <= 2011
keep if t_is_teacher == 1
keep if !missing(tre_`subject')
if "`level'" == "elem" {
       keep if school_lvl == "Elem"
}
if "`level'" == "middle" {
       keep if school_lvl == "Mid"
}
```

// Step 4: Review variables.

```
tab school_year
bysort tid: gen tag = _n == 1
summ tre_`subject' if tag == 1, detail
drop tag
```

// Step 5: Change data from teacher-year uniqueness level to teacher level by keeping only teacher id and pooled estimate and then dropping duplicate records.

```
keep tid tre_`subject'
duplicates drop
isid tid
```

// Step 6: Get sample size.

```
summ tid
local unique_teachers = string(r(N), "%9.0fc")
```

// Step 7: Get and store values for percentiles and percentile differences.

```
sum tre_`subject', detail
local p10_`subject' = r(p10)
local p25_`subject' = r(p25)
```

1. DISTRIBUTION OF TEACHER EFFECTS

```
local p75_`subject' = r(p75)
local p90_`subject' = r(p90)

local gap90 = round(`p90_`subject'' - `p10_`subject'', .01)
local gap75 = round(`p75_`subject'' - `p25_`subject'', .01)
```

// Step 8: Set positions for text labels, lines, and arrows.

```
local gphtxt_10           = `p10_`subject'' - .02
local gphtxt_25           = `p25_`subject'' - .02
local gphtxt_75           = `p75_`subject'' + .02
local gphtxt_90           = `p90_`subject'' + .02

local lnht75        = 1.5
local lnht90        = 0.5
local txtht75       = `lnht75' + .3
local txtht90       = `lnht90' + .3
```

// Step 9: Define subject and school level titles.

```
if "`subject'" == "math" {
       local subj_foot "math"
       local subj_title "Math"
}

if "`subject'" == "ela" {
       local subj_foot "English/Language Arts"
       local subj_title "ELA"
}

local gradespan "5th through 8th"

if "`level'" == "middle" {
       local subj_title "Middle School `subj_title'"
       local gradespan "6th through 8th"
}

if "`level'" == "elem" {
       local subj_title "Elementary School `subj_title'"
       local gradespan "5th"
}
```

// Step 10: Make chart.

```
#delimit ;
twoway (pcarrowi `lnht75' `p25_`subject'' `lnht75' `p75_`subject'',
              color(cranberry) mlwidth(medthin) lwidth(medthin))
       (pcarrowi `lnht75' `p75_`subject'' `lnht75' `p25_`subject'',
              color(cranberry) mlwidth(medthin) lwidth(medthin))
       (pcarrowi `lnht90' `p10_`subject'' `lnht90' `p90_`subject'',
              color(green) mlwidth(medthin) lwidth(medthin))
       (pcarrowi `lnht90' `p90_`subject'' `lnht90' `p10_`subject'',
```

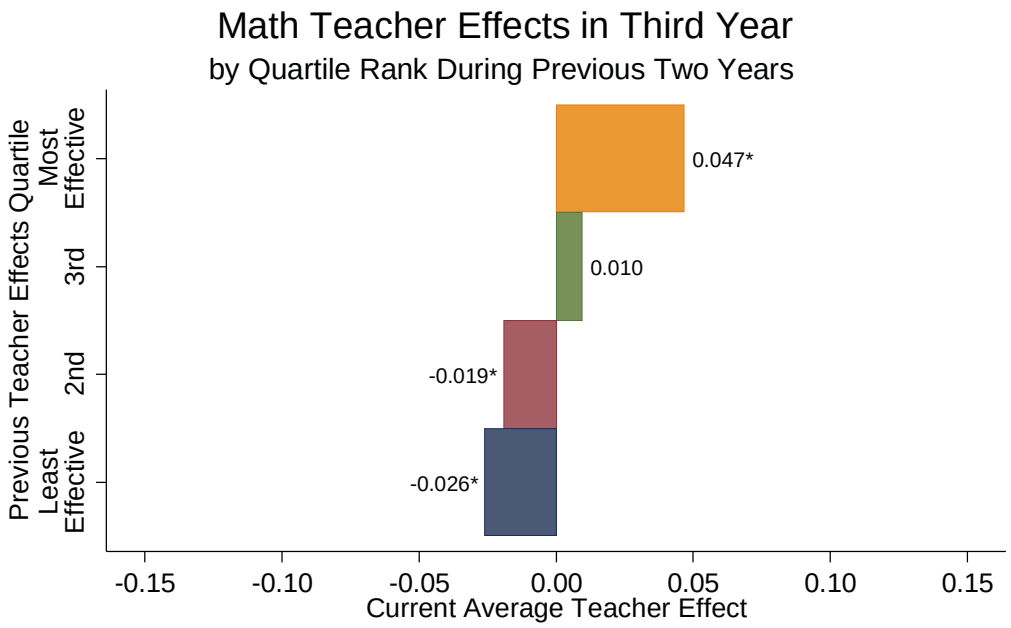
1. DISTRIBUTION OF TEACHER EFFECTS

```
            color(green) mlwidth(medthin) lwidth(medthin))
      (kdensity tre_`subject', color(navy) area(1)
            xline(`p25_`subject'', lpattern(dash) lcolor(cranberry))
            xline(`p75_`subject'', lpattern(dash) lcolor(cranberry))
            xline(`p10_`subject'', lpattern(dash) lcolor(green))
            xline(`p90_`subject'', lpattern(dash) lcolor(green))),

      title("Distribution of `subj_title' Teacher Effects" " ", span)
      xtitle("Teacher Effects", size(medsmall))

      xscale(range(-.5(.25).5))
      xlabel(-.5(.25).5, labsize(medsmall))
      yscale(range(0(1)6))
      ylabel(none)
      text(5 `gphtxt_10' "10th percentile", orientation(vertical) size(small))
      text(5 `gphtxt_25' "25th percentile", orientation(vertical) size(small))
      text(5 `gphtxt_75' "75th percentile", orientation(vertical) size(small))
      text(5 `gphtxt_90' "90th percentile", orientation(vertical) size(small))
      text(`txtht75' 0 "`gap75'", size(medium) color(cranberry))
      text(`txtht90' 0 "`gap90'", size(medium) color(green))
      legend(off)
      graphregion(color(white) fcolor(white) lcolor(white))
      plotregion(color(white) fcolor(white) lcolor(white) margin(5 5 2 0))
note(" " "Notes: Sample includes `unique_teachers' `gradespan' grade `subj_foot'
teachers in school years 2007-08 through 2010-11. Teacher effects are" "measured in test
score standard deviations, with teacher-specific shrinkage factors applied to adjust for
differences in sample reliability.", size(vsmall) span);
#delimit cr
```

// Step 11: Save chart.

```
graph save "${graphs}\D1_Overall_Teacher_Effects_`subj_title'.gph" , replace
graph export "${graphs}\D1_Overall_Teacher_Effects_`subj_title'.emf" , replace
```

2. AVERAGE TEACHER EFFECTS IN THIRD YEAR BY QUARTILE RANK DURING PREVIOUS TWO YEARS



*Significantly different from zero, at the 95 percent confidence level.
Notes: Sample includes 204 5th through 8th grade math teachers with three years of teacher effects estimates in school years 2007-08 through 2010-11. Teacher effects are measured in student test score standard deviations, with teacher-specific shrinkage factors applied to adjust for differences in sample reliability.

Purpose:
To show the extent to which prior estimates of a teacher's effectiveness predict effectiveness in future years.

Required analysis file variables:

```
tid
school_year
t_is_teacher
school_lvl
curr2year_tre_math
current_tre_math
```

Analysis-specific sample restrictions:

- Keep only records for teachers with three years of effectiveness estimates in the given subject.
- If school_level restriction is chosen, keep only records for elementary or middle school teachers.

Ask yourself:

- Value-added estimates are more likely to be accurate for teachers at the high and low ends of the distribution. What kinds of decisions would knowing which teachers consistently perform at the top and bottom ends of the distribution of value-added estimates help your agency make? How might this information be used for probationary teachers? For veteran teachers?

Potential further analyses:

If you have sufficient sample space, you can restrict the sample to teachers in their third year at the agency to understand how performance of probationary teachers in their first two years compares to performance in the third year.

2. AVERAGE TEACHER EFFECTS IN THIRD YEAR BY QUARTILE RANK DURING PREVIOUS TWO YEARS

// Step 1: Choose the subject (math or ela) and school level (elem or middle) for the analysis. Note: to make multiple charts at the same time, put loops for subject and level around the analysis and graphing code. To include all grade levels in the analysis, comment out the local level command below.

```
local subject math
*local level middle
```

// Step 2: Load the Teacher_Year_Analysis file containing value-added estimates.

```
use "${analysis}/Teacher_Year_Analysis.dta", clear
isid tid school_year
```

// Step 3: Restrict the sample. Keep years for which teacher effects value added estimates are available. Keep only records for which single-year teacher effectiveness estimates are available. Keep only employees who are teachers. If school level restriction is chosen, keep only records from either elementary or middle schools.

```
keep if school_year >= 2008 & school_year <= 2011
keep if t_is_teacher == 1
keep if !missing(current_tre_`subject')
if "`level'" == "elem" {
       keep if school_lvl == "Elem"
}
if "`level'" == "middle" {
       keep if school_lvl == "Mid"
}
```

// Step 4: Review variables.

```
tab school_year
summ current_tre_`subject', detail
summ curr2year_tre_`subject', detail
```

// Step 5: Identify the most recent year a teacher is present in the data and tag as "year 3."

```
egen max_school_year = max(school_year), by(tid)
gen year3 = max_school_year == school_year
drop max_school_year
tab year3, mi
```

// Step 6: Set time series structure and use lead operators to identify years 2 and 1.

```
tsset tid school_year
gen year1 = 0
gen year2 = 0
bysort tid: replace year2 = 1 if F.year3 == 1
bysort tid: replace year1 = 1 if F.year2 == 1
tab year2 year3, mi
tab year1 year3, mi
```

// Step 7: Keep a balanced panel which includes only teachers with observations for all 3 years.

2. AVERAGE TEACHER EFFECTS IN THIRD YEAR BY QUARTILE RANK DURING PREVIOUS TWO YEARS

```
bysort tid: egen balanced = max(year1)
keep if balanced == 1
drop balanced
unique tid
```

// Step 8: Assign teachers to quartiles based on two-year pooled teacher effects in year 2, and generate dummy variables for quartiles.

```
assert !missing(curr2year_tre_`subject') if year2 == 1
xtile quart_temp = curr2year_tre_`subject' if year2 == 1, nq(4)
bysort tid: egen quart = max(quart_temp)
tab quart if year2 == 1, mi
tab quart, gen(quart)
```

// Step 9: Drop records for years 1 and 2, reducing data to one record per teacher.

```
keep if year3 == 1
isid tid
```

// Step 10: Get sample size.

```
sum tid
local unique_teachers = string(r(N), "%9.0fc")
```

// Step 11: Get significance.

```
gen sig = 0
forval quartile = 1/4 {
       reg current_tre_`subject' quart`quartile', robust
       test _b[quart`quartile'] == 0
       gen sig`quartile' = r(p) < .05
       replace sig = sig`quartile' if quart`quartile' == 1
}
```

// Step 12: Collapse the data for graphing.

```
collapse (mean) current_tre_`subject' sig, by(quart)
```

// Step 13: Concatenate value labels and significance asterisks.

```
gen tre_str = string(current_tre_`subject', "%9.3f")
gen star = ""
forval quartile = 1/4 {
       replace star = "*" if quart == `quartile' & sig == 1
}

egen tre_label = concat(tre_str star), format(%9.3f)
```

// Step 14: Define subject and school level titles.

2. AVERAGE TEACHER EFFECTS IN THIRD YEAR BY QUARTILE RANK DURING PREVIOUS TWO YEARS

```
if "`subject'" == "math" {
       local subj_foot "math"
       local subj_title "Math"
}
if "`subject'" == "ela" {
       local subj_foot "English/Language Arts"
       local subj_title "ELA"
}

local gradespan "5th through 8th"

if "`level'" == "middle" {
       local subj_title "Middle School `subj_title'"
       local gradespan "6th through 8th"
}
if "`level'" == "elem" {
       local subj_title "Elementary School `subj_title'"
       local gradespan "5th"
}
```

// Step 15: Make chart.

```
#delimit ;
twoway (bar current_tre_`subj' quart if quart == 4,
              horizontal fcolor(dkorange) lcolor(dkorange) lwidth(0))
       (bar current_tre_`subj' quart if quart == 3,
              horizontal fcolor(forest_green) lcolor(forest_green) lwidth(0))
       (bar current_tre_`subj' quart if quart == 2,
              horizontal fcolor(maroon) lcolor(maroon) lwidth(0))
       (bar current_tre_`subj' quart if quart == 1,
              horizontal fcolor(dknavy) lcolor(dknavy) lwidth(0))
       (scatter quart current_tre_`subj' if current_tre_`subj' >= 0,
              mlabel(tre_label) msymbol(i) mlabpos(3) mlabcolor(black) mlabgap(.2))
       (scatter quart current_tre_`subj' if current_tre_`subj' < 0,
              mlabel(tre_label) msymbol(i) mlabpos(9) mlabcolor(black) mlabgap(.2)),
       title("`subj_title' Teacher Effects in Third Year", span)
       subtitle("by Quartile Rank During Previous Two Years", span)
       xtitle("Current Average Teacher Effect", size(medsmall))
              xscale(range(-0.15 (.05) 0.15))
              xlabel(-0.15 (.05) 0.15, format(%9.2f) labsize(medsmall))
              ytitle("Previous Teacher Effects Quartile", size(medsmall))
              yscale(range(1(1)4))
              ylabel(1 `""Least" "Effective""' 2 "2nd" 3 "3rd" 4 `""Most" "Effective""',
              labsize(medsmall) nogrid)
       legend(off)
       graphregion(color(white) fcolor(white) lcolor(white))
       plotregion(color(white) fcolor(white) lcolor(white) margin(5 5 2 2))

note(" " "*Significantly different from zero, at the 95 percent confidence level."
"Notes: Sample includes `unique_teachers' `gradespan' grade `subj_foot' teachers
with three years of teacher effects estimates in school years 2007-08" "through 2010-11.
Teacher effects are measured in test score standard deviations, with teacher-specific
shrinkage factors applied to adjust" "for differences in sample reliability.",
```

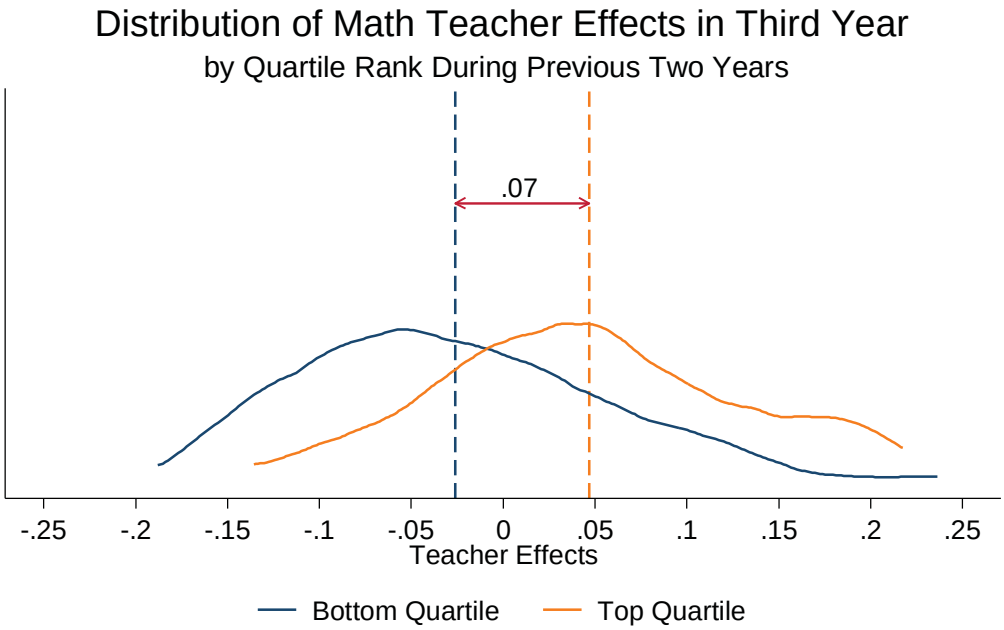
2. AVERAGE TEACHER EFFECTS IN THIRD YEAR BY QUARTILE RANK DURING PREVIOUS TWO YEARS

```
      span size(vsmall));

#delimit cr
```

// Step 16: Save chart.

```
graph save "${graphs}\D2_Predictive_Tchr_Effects_Avg_`subj_title'.gph" , replace
graph export "${graphs}\D2_Predictive_Tchr_Effects_Avg_`subj_title'.emf" , replace
```

3. DISTRIBUTION OF TEACHER EFFECTS IN THIRD YEAR BY QUARTILE RANK DURING PREVIOUS TWO YEARS



Notes: Sample includes 204 5th through 8th grade math teachers with three years of teacher effects estimates in school years 2007-08 through 2010-11. Teacher effects are measured in student test score standard deviations, with teacher-specific shrinkage factors applied to adjust for differences in sample reliability.

Purpose:
Examine the distribution of teacher effectiveness in a third year for teachers ranked most and least effective in the prior two years.

Required analysis file variables:

```
tid
school_year
t_is_teacher
school_lvl
curr2year_tre_math
current_tre_math
```

Analysis-specific sample restrictions:

- Keep only records for teachers with three years of effectiveness estimates in the given subject.
- If school_level restriction is chosen, keep only records for elementary or middle school teachers.

Ask yourself:

- How well do rankings in the first two years predict teacher effectiveness in the third year? Do you see distinct, differentiated peaks in the two distributions? How much overlap is there between the two distributions? How likely are teachers to change effectiveness rankings from year to year?

Potential further analyses:

- If you have sufficient sample size, you can restrict the sample to teachers in their third year at the agency to understand how performance of probationary teachers in their first two years compare to performance in the third year.

3. DISTRIBUTION OF TEACHER EFFECTS IN THIRD YEAR BY QUARTILE RANK DURING PREVIOUS TWO YEARS

// Step 1: Choose the subject (math or ela) and school level (elem or middle) for the analysis. Note: to make multiple charts at the same time, put loops for subject and level around the analysis and graphing code. To include all grade levels in the analysis, comment out the local level command below.

```
local subject math
*local level middle
```

// Step 2: Load the Teacher_Year_Analysis file containing value-added estimates.

```
use "${analysis}/Teacher_Year_Analysis.dta", clear
isid tid school_year
```

// Step 3: Restrict the sample. Keep years for which teacher effects value added estimates are available. Keep only records for which single-year teacher effectiveness estimates are available. Keep only employees who are teachers. If school level restriction is chosen, keep only records from either elementary or middle schools.

```
keep if school_year >= 2008 & school_year <= 2011
keep if t_is_teacher == 1
keep if !missing(current_tre_`subject')
if "`level'" == "elem" {
       keep if school_lvl == "Elem"
}
if "`level'" == "middle" {
       keep if school_lvl == "Mid"
}
```

// Step 4: Review variables.

```
tab school_year
summ current_tre_`subject', detail
summ curr2year_tre_`subject', detail
```

// Step 5: Identify the most recent year a teacher is present in the data and tag as "year 3."

```
egen max_school_year = max(school_year), by(tid)
gen year3 = max_school_year == school_year
drop max_school_year
tab year3, mi
```

// Step 6: Set time series structure and use lead operators to identify years 2 and 1.

```
tsset tid school_year
gen year1 = 0
gen year2 = 0
bysort tid: replace year2 = 1 if F.year3 == 1
bysort tid: replace year1 = 1 if F.year2 == 1
tab year2 year3, mi
tab year1 year3, mi
```

3. DISTRIBUTION OF TEACHER EFFECTS IN THIRD YEAR BY QUARTILE RANK DURING PREVIOUS TWO YEARS

// Step 7: Keep a balanced panel which includes only teachers with observations for all 3 years.

```
bysort tid: egen balanced = max(year1)
keep if balanced == 1
drop balanced
unique tid
```

// Step 8: Assign teachers to quartiles based on two-year pooled teacher effects in year 2, and generate dummy variables for quartiles.

```
assert !missing(curr2year_tre_`subject') if year2 == 1
xtile quart_temp = curr2year_tre_`subject' if year2 == 1, nq(4)
bysort tid: egen quart = max(quart_temp)
tab quart if year2 == 1, mi
tab quart, gen(quart)
```

// Step 9: Drop records for years 1 and 2, reducing data to one record per teacher.

```
keep if year3 == 1
isid tid
```

// Step 10: Get sample size.

```
       sum tid
       local unique_teachers = string(r(N), "%9.0fc")
```

// Step 11: Get quartile means and the difference between means for quartiles 1 and 4.

```
forval quartile = 1/4 {
       summ current_tre_`subject' if quart == `quartile'
       local mean_q`quartile' = r(mean)
}
local diff = round(`mean_q4' - `mean_q1', .01)
```

// Step 12: Set positions for lines and text on chart.

```
local lnht = 8.5
local txtht = `lnht' + .5
local diff_pl = `mean_q1' + `diff' * .5
```

// Step 13: Define subject and school level titles.

```
if "`subject'" == "math" {
       local subj_foot "math"
       local subj_title "Math"
}
if "`subject'" == "ela" {
       local subj_foot "English/Language Arts"
       local subj_title "ELA"
}
```

3. DISTRIBUTION OF TEACHER EFFECTS IN THIRD YEAR BY QUARTILE RANK DURING PREVIOUS TWO YEARS

```
local gradespan "5th through 8th"

if "`level'" == "middle" {
       local subj_title "Middle School `subj_title'"
       local gradespan "6th through 8th"
}
if "`level'" == "elem" {

       local subj_title "Elementary School `subj_title'"
       local gradespan "5th"
}
```

// Step 14: Make chart.

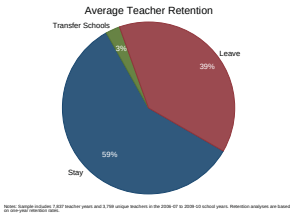
```
#delimit ;
twoway (pcarrowi `lnht' `mean_q4' `lnht' `mean_q1',
              color(cranberry) mlwidth(medthin) lwidth(medthin))
       (pcarrowi `lnht' `mean_q1' `lnht' `mean_q4',
              color(cranberry) mlwidth(medthin) lwidth(medthin))
       (kdensity current_tre_`subject' if quart == 1,
              lcolor(navy) area(1)
              xline(`mean_q1', lpattern(dash) lcolor(navy)))
       (kdensity current_tre_`subject' if quart == 4,
              lcolor(orange) area(1) lwidth(medium)
              xline(`mean_q4', lpattern(dash) lcolor(orange))),
       text(`txtht' `diff_pl' "`diff'", placement(0))
       title("Distribution of `subj_title' Teacher Effects in Third Year", span)
       subtitle("by Quartile Rank During Previous Two Years", span)
       xtitle("Teacher Effects", size(medsmall))
              xscale(range(-.25(.05).25))
              xlabel(-.25(.05).25, labsize(medsmall))
              ytitle("",)
              yscale(range(0(2)12))
              ylabel(none)
       legend(order(3 4) rows(1) label(3 "Bottom Quartile")
              label(4 "Top Quartile"))
       legend(symxsize(5) ring(1) size(medsmall)
              region(lstyle(none) lcolor(none) color(none)))
       graphregion(color(white) fcolor(white) lcolor(white))
       plotregion(color(white) fcolor(white) lcolor(white) margin(5 5 2 0))
note(" " "Notes: Sample includes `unique_teachers' `gradespan' grade `subj_foot'
teachers with three years of teacher effects estimates in school" "years 2007-08 through
2010-11. Teacher effects are measured in student test score standard deviations, with
teacher-specific" "shrinkage factors applied to adjust for differences in sample
reliability.", span size(vsmall));
#delimit cr
```

// Step 15: Save chart.

```
graph save "${graphs}\D3_Predictive_Tchr_Effects_Dist_`subj_title'.gph" , replace
graph export "${graphs}\D3_Predictive_Tchr_Effects_Dist_`subj_title'.emf" , replace
```

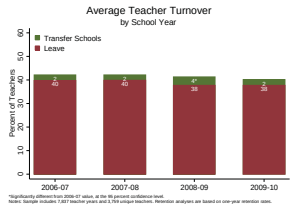
E. RETENTION

Schools invest substantial resources recruiting, developing, and retaining high-quality teachers. The analyses in the final section of the human capital pathway reveal how retention patterns vary by school characteristics and by teacher effectiveness categories. Because novice teachers typically experience high rates of turnover, retention analyses often focus on the retention patterns of novice teachers.



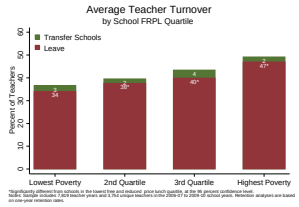
1. AVERAGE ANNUAL TEACHER RETENTION

Describes the overall annual shares of teachers who stay in the same school, transfer, and leave teaching in the agency.



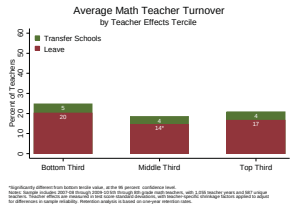
2.TEACHER TURNOVER BY SCHOOL YEAR

Describes the shares of teachers who transfer and leave over time.



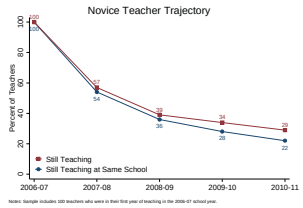
3. TEACHER TURNOVER BY SCHOOL POVERTY QUARTILE

Examines the extent to which retention patterns differ according to school poverty characrteristics.



4. TEACHER TURNOVER BY TEACHER EFFECTIVENESS TERCILE

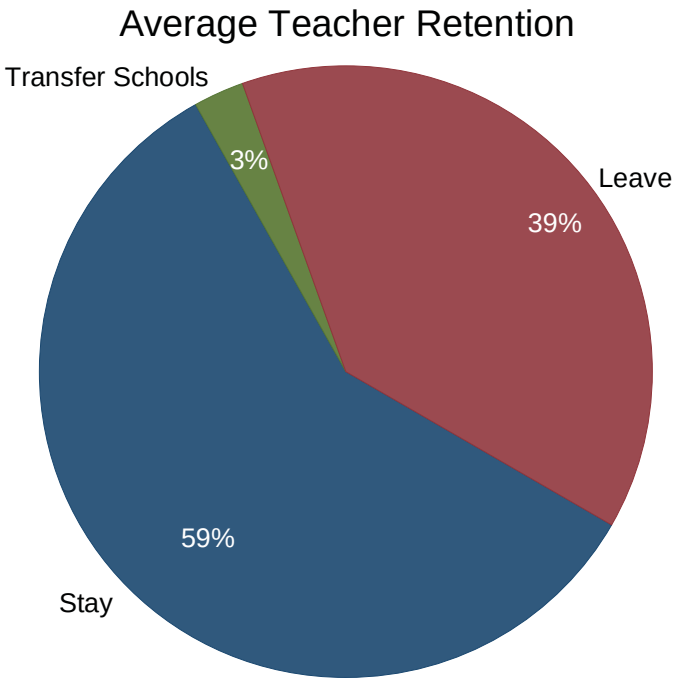
Examines whether the most and least effective teachers are being differentially retained.



5. NOVICE TEACHER RETENTION TRAJECTORY

Describes the retention trajectory of a cohort of novice teachers.

1. AVERAGE ANNUAL TEACHER RETENTION



Notes: Sample includes 7,837 teacher years and 3,759 unique teachers in the 2006-07 to 2009-10 school years. Retention analyses are based on one-year retention rates.

Purpose:
Examine basic novice teacher retention patterns for years in the agency.

Required analysis file variables:

```
tid
school_year
t_is_teacher
t_stay
t_transfer
t_leave
```

Analysis-specific sample restrictions:

- Keep only employees whose job code is "teacher".
- Keep only years for which next-year retention status can be calculated.

Ask yourself:

- Are transfer and attrition rates what you expected?
- What is the optimal balance for your agency between retaining experienced teachers and recruiting, hiring, and training new teachers?

Potential further analyses:
You can add a category for retirees, if you have explicit retirement data, or for "likely retirees," teachers who left teaching and were above a cutoff age.

1. AVERAGE ANNUAL TEACHER RETENTION

// Step 1: Load data.

```
use "${analysis}\Teacher_Year_Analysis.dta", clear
isid tid school_year
```

//Step 2: Restrict sample. Keep only teachers in years for which next-year retention status can be calculated.

```
keep if t_is_teacher == 1
keep if school_year >= 2007 & school_year <= 2010
assert !missing(t_stay, t_transfer, t_leave)
```

// Step 3: Review variables.

```
assert t_leave + t_transfer + t_stay == 1
tab school_year t_stay, mi
tab school_year t_transfer, mi
tab school_year t_leave, mi
```

// Step 4: Get sample size.

```
summ tid
local teacher_years = string(r(N), "%9.0fc")
preserve
      bysort tid: keep if _n == 1
      summ tid
      local unique_teachers = string(r(N), "%9.0fc")
restore
```

// Step 5: Collapse data and calculate shares.

```
collapse (mean) t_stay t_transfer t_leave (count) tid
foreach var of varlist t_stay t_transfer t_leave {
      replace `var' = `var' * 100
}
```

// Step 6: Make chart.

```
#delimit ;
graph pie t_stay t_transfer t_leave,
      angle0
            (330)
      title
      ("Average Teacher Retention", span)
            pie
```

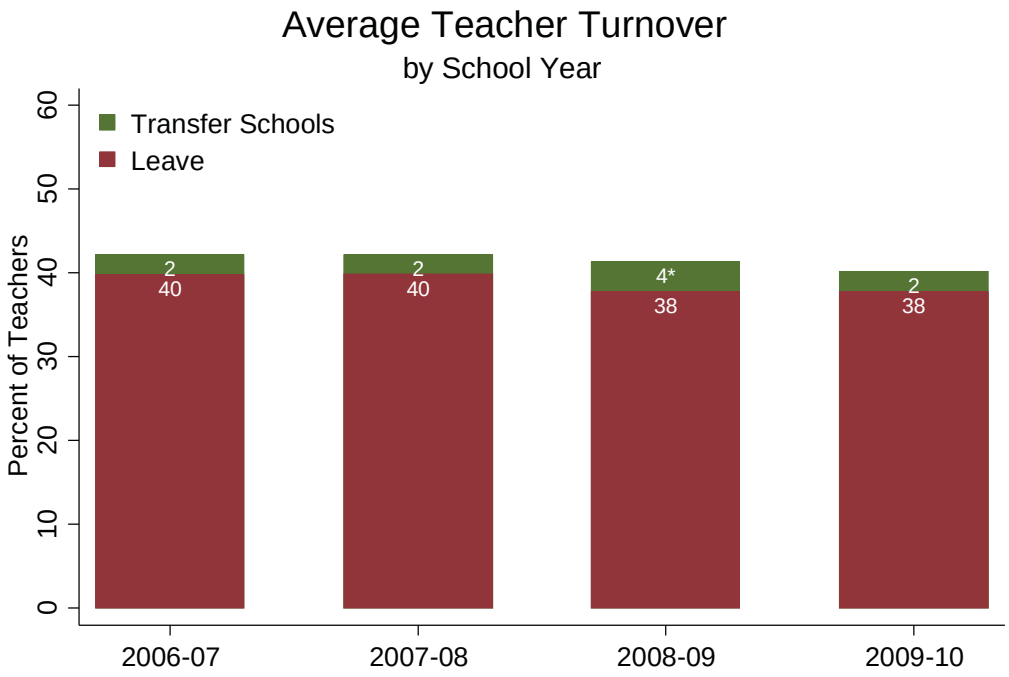
1. AVERAGE ANNUAL TEACHER RETENTION

```
            (1, color(navy))
      pie
            (2, color(forest_green))
      pie
            (3, color(maroon))
      pie
            (4, color(dkorange))
      plabel
            (1 percent, gap(5) format("%2.0f") color(white) size(medsmall) placement(3))
      plabel
            (2 percent, gap(5) format("%2.0f") color(white) size(medsmall) placement(0))
      plabel
            (3 percent, gap(5) format("%2.0f") color(white) size(medsmall) placement(3))
      plabel
            (4 percent, gap(5) format("%2.0f") color(white) size(medsmall) placement(3))
      plabel
            (1 "Stay", color(black) size(medsmall) placement(9) gap(20))
      plabel
            (2 "Transfer Schools", color(black) size(medsmall) placement(9) gap(20))
      plabel
            (3 "Leave", color(black) size(medsmall) placement(4) gap(20))

            legend
                  (off)

      graphregion(color(white) fcolor(white) lcolor(white)) plotregion(color(white)
            fcolor(white) lcolor(white))
note(" " "Notes: Sample includes `teacher_years' teacher years and `unique_
teachers'   unique teachers in the 2006-07 to 2009-20 school years. Retention
analyses are based" "on one-year retention rates.", span size(vsmall)) ;
#delimit cr
```

// Step 7: Save chart.

```
graph save "$graphs\E1_Average_Teacher_Retention.gph", replace
graph export "$graphs\E1_Average_Teacher_Retention.emf", replace
```

2.TEACHER TURNOVER BY SCHOOL YEAR



*Significantly different from 2006-07 value, at the 95 percent confidence level.
Notes: Sample includes 7,837 teacher years and 3,759 unique teachers. Retention analyses are based on one-year retention rates.

Purpose:
Describe the share of teachers who transfer and leave over time.

Required analysis file variables:

```
tid
school_year
t_is_teacher
t_stay
t_transfer
t_leave
```

Analysis-specific sample restrictions:

- Keep only employees whose job code is "teacher".
- Keep only years for which next-year retention status can be calculated.

Ask yourself:

- How has turnover changed over time?
- What factors might account for the trends that I see?

Potential further analyses:

- Instead of examining turnover by school year, you can use a graph of this type to look at turnover by other categories, including by district, school type, or school. You can also add a loop to the graphing code to generate multiple charts that examine trends over time for a different geographic or other categories.

2.TEACHER TURNOVER BY SCHOOL YEAR

// Step 1: Load data.

```
use "${analysis}\Teacher_Year_Analysis.dta", clear
isid tid school_year
```

// Step 2: Restrict sample. Keep only teachers in years for which next-year retention status can be calculated.

```
keep if t_is_teacher == 1
keep if school_year >= 2007 & school_year <= 2010
assert !missing(t_stay, t_transfer, t_leave)
```

// Step 3: Review variables.

```
assert t_leave + t_transfer + t_stay == 1
tab school_year t_stay, mi row
tab school_year t_transfer, mi row
tab school_year t_leave, mi row
```

// Step 4: Get sample size.

```
summ tid
local teacher_years = string(r(N), "%9.0fc")
preserve
       bysort tid: keep if _n == 1
       summ tid
       local unique_teachers = string(r(N), "%9.0fc")
restore
```

// Step 5: Calculate significance indicator variables by year.

```
foreach var in t_leave t_transfer {
       gen sig_`var' = .
       xi: logit `var' i.school_year, robust
       forval year = 2008/2010 {
              replace sig_`var' = abs(_b[_Ischool_ye_`year'] / _se[_Ischool_ye_`year']) ///
                     if school_year == `year'
              replace sig_`var' = 0 if sig_`var' <= 1.96 & school_year == `year'
              replace sig_`var' = 1 if sig_`var' > 1.96 & school_year == `year'

       }

       replace sig_`var' = 0 if school_year == 2007

       }
```

2.TEACHER TURNOVER BY SCHOOL YEAR

// Step 6: Collapse and calculate shares.

```
collapse (mean) t_leave t_transfer sig_* (count) tid, by(school_year)
foreach var of varlist t_leave t_transfer {
	replace `var' = `var' * 100
}
```

// Step 7: Concatenate value and significance asterisk.

```
foreach var of varlist t_leave t_transfer {
	tostring(sig_`var'), replace
	replace sig_`var' = "*" if sig_`var' == "1"
	replace sig_`var' = "" if sig_`var' == "0"
	gen `var'_str = string(`var', "%9.0f")
	egen `var'_label = concat(`var'_str sig_`var')
}
```

// Step 8: Generate count variable and add variables cumulatively for graphing

```
gen count = _n
replace t_transfer = t_leave + t_transfer
```

// Step 9: Make chart.

```
#delimit ;
twoway bar t_transfer count,
	barwidth(.6) color(forest_green) finten(100) ||
		bar t_leave count,
	barwidth(.6) color(maroon) finten(100) ||
	scatter t_transfer count,
		mlabel(t_transfer_label)
		msymbol(i) msize(tiny) mlabpos(6) mlabcolor(white)
		mlabgap(.001) ||
	scatter t_leave count,
		mlabel(t_leave_label)
		msymbol(i) msize(tiny) mlabpos(6) mlabcolor(white)
		mlabgap(.001) ||,

	title("Average Teacher Turnover", span)
	subtitle("by School Year", span)
	ytitle("Percent of Teachers", size(medsmall))
	yscale(range(0(10)60))
	ylabel(0(10)60, nogrid labsize(medsmall))
	xtitle("")
	xlabel(1 "2006-07" 2 "2007-08" 3 "2008-09" 4 "2009-10", labsize(medsmall))
	legend(order(1 "Transfer Schools" 2 "Leave")
		ring(0) position(11) symxsize(2) symysize(2) rows(2) size(medsmall)
		region(lstyle(none) lcolor(none) color(none)))
	graphregion(color(white) fcolor(white) lcolor(white)) plotregion(color(white)
		fcolor(white) lcolor(white))
		note("*Significantly different from 2006-07 value, at the 95 percent confidence
	level." "Notes: Sample includes `teacher_years' teacher years and
```

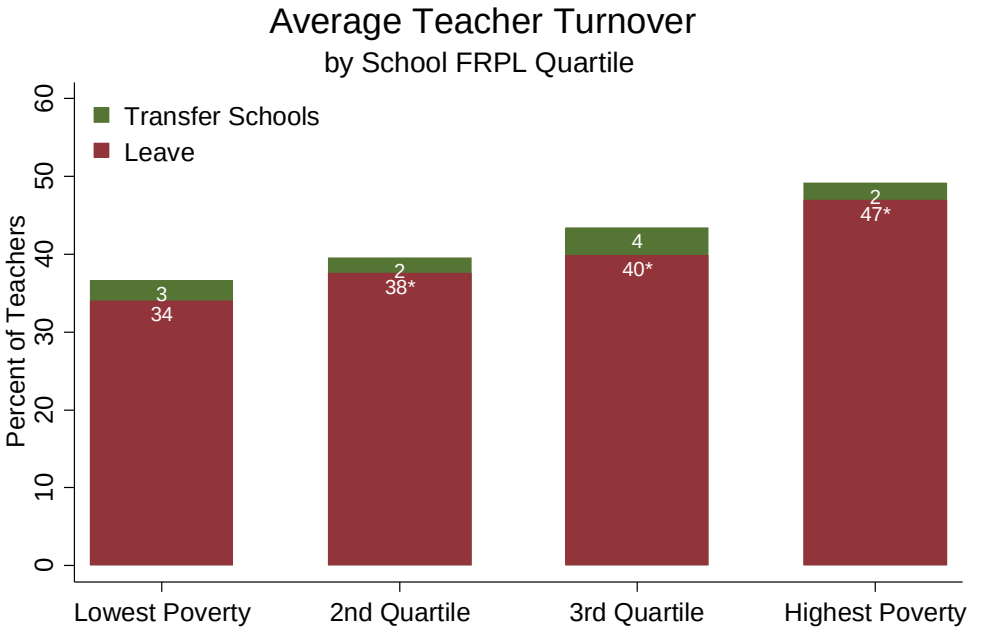
2.TEACHER TURNOVER BY SCHOOL YEAR

```
       `unique_teachers' unique teachers. Retention analyses are based on one-year retention
       rates.", span size(vsmall));
#delimit cr
```

// Step 10: Save chart.

```
graph save "$graphs\E2_Retention_by_School_Year.gph", replace
graph export "$graphs\E2_Retention_by_School_Year.emf", replace
```

3. TEACHER TURNOVER BY SCHOOL POVERTY QUARTILE



*Significantly different from schools in the lowest free and reduced price lunch quartile, at the 95 percent confidence level.
Notes: Sample includes 7,819 teacher years and 3,754 unique teachers in the 2006-07 to 2009-10 school years. Retention analyses are based on one-year retention rates.

Purpose:
Examine the extent to which retention patterns differ according to school poverty characteristics.

Required analysis file variables:

```
tid
school_year
t_is_teacher
t_transfer
t_leave
school_poverty_quartile
```

Analysis-specific sample restrictions:

- Keep only employees whose job code is "teacher".
- Keep only years for which next-year retention status can be calculated.

Ask yourself:

- How do turnover patterns vary for high-and-low poverty schools?
- What other factors (school and district size, urban vs. rural, school closings, etc.) might help account for the differences I see?
- Does your agency have an incentive program in place to increase recruiting and retention in high-need schools?

Potential further analyses:
You may want to use ranges of school free and reduced price lunch percentages, rather than quartiles, to make the chart easier to interpret, or use your agency's own classification of high-need schools. You can also use a graph of this type to examine teacher turnover by other school characteristics. For example, you could explore teacher turnover by school student minority share quartiles, or average test score quartiles.

3. TEACHER TURNOVER BY SCHOOL POVERTY QUARTILE

// Step 1: Load data.

```
use "${analysis}\Teacher_Year_Analysis.dta", clear
isid tid school_year
```

// Step 2: Restrict sample. Keep only teachers in years for which next-year retention status can be calculated. Keep records with non-missing values for school poverty quartile.

```
keep if t_is_teacher == 1
keep if school_year >= 2007 & school_year <= 2010
keep if !missing(school_poverty_quartile)
assert !missing(t_stay, t_transfer, t_leave)
```

// Step 3: Review variables.

```
assert t_leave + t_transfer + t_stay == 1
tab school_poverty_quartile t_stay, mi row
tab school_poverty_quartile t_transfer, mi row
tab school_poverty_quartile t_leave, mi row
```

// Step 4: Get sample sizes.

```
sum tid
local teacher_years = string(r(N), "%9.0fc")
preserve
	bysort tid: keep if _n == 1
	sum tid
	local unique_teachers = string(r(N), "%9.0fc")
restore
```

// Step 5: Calculate significance indicator variables by quartile.

```
foreach var of varlist t_leave t_transfer {
	gen sig_`var' = .
	xi: logit `var' i.school_poverty_quartile, robust
	forval quartile = 2/4 {
		replace sig_`var' = abs(_b[_Ischool_po_`quartile'] / ///
			_se[_Ischool_po_`quartile']) if school_poverty_quartile == `quartile'
		replace sig_`var' = 0 if sig_`var' <= 1.96 & ///
			school_poverty_quartile == `quartile'
		replace sig_`var' = 1 if sig_`var' > 1.96 & ///
			school_poverty_quartile == `quartile'
	}
	replace sig_`var' = 0 if school_poverty_quartile == 1
	}
```

// Step 6: Collapse and calculate shares.

```
collapse (mean) t_leave t_transfer sig_* (count) tid, by(school_poverty_quartile)
foreach var of varlist t_leave t_transfer {
```

3. TEACHER TURNOVER BY SCHOOL POVERTY QUARTILE

```
	replace `var' = `var' * 100
}
```

// Step 7: Concatenate value and significance asterisk.

```
foreach var of varlist t_leave t_transfer {
	tostring(sig_`var'), replace
	replace sig_`var' = "*" if sig_`var' == "1"
	replace sig_`var' = "" if sig_`var' == "0"
	gen `var'_str = string(`var', "%9.0f")
	egen `var'_label = concat(`var'_str sig_`var')
}
```

// Step 8: Generate count variable and add variables cumulatively for graphing.

```
gen count = _n
replace t_transfer = t_leave + t_transfer
```

// Step 9: Make chart.

```
#delimit ;
	twoway bar t_transfer count,
	barwidth(.6) color(forest_green) finten(100) ||

	bar t_leave count,
	barwidth(.6) color(maroon) finten(100) ||

	scatter t_transfer count,
		mlabel(t_transfer_label)
		msymbol(i) msize(tiny) mlabpos(6) mlabcolor(white) mlabgap(.001) ||

	scatter t_leave count,
		mlabel(t_leave_label)
		msymbol(i) msize(tiny) mlabpos(6) mlabcolor(white) mlabgap(.001) ||,
	title("Average Teacher Turnover", span)
	subtitle("by School FRPL Quartile", span)
	ytitle("Percent of Teachers", size(medsmall))
	yscale(range(0(10)60))
	ylabel(0(10)60, nogrid labsize(medsmall))
	xtitle("")
	xlabel(1 "Lowest Poverty" 2 "2nd Quartile" 3 "3rd Quartile" 4 "Highest Poverty",
		labsize(medsmall))
	legend(order(1 "Transfer Schools" 2 "Leave")
		ring(0) position(11) symxsize(2) symysize(2) rows(2) size(medsmall)
		region(lstyle(none) lcolor(none) color(none)))

	graphregion(color(white) fcolor(white) lcolor(white)) plotregion(color(white)
		fcolor(white) lcolor(white))
note("*Significantly different from schools in the lowest free and reduced
price lunch quartile, at the 95 percent confidence level." "Notes: Sample includes
`teacher_years' teacher years and `unique_teachers' unique teachers in the 2006-07
```

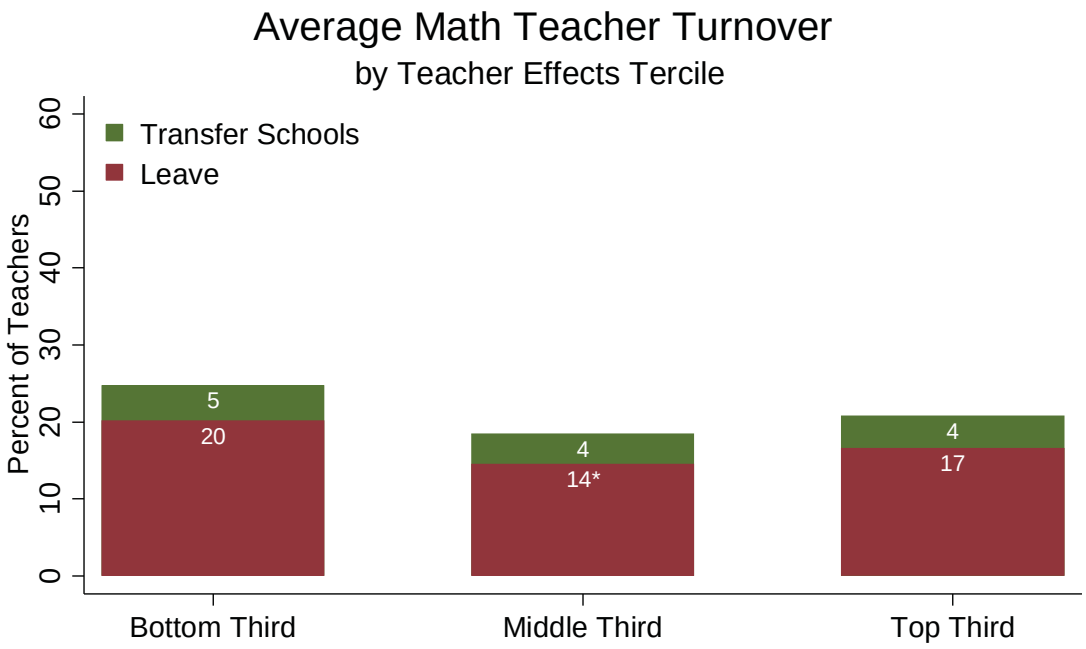
3. TEACHER TURNOVER BY SCHOOL POVERTY QUARTILE

```
to 2009-10 school years. Retention analyses are based" "on one-year retention rates.",
span size(vsmall));
#delimit cr
```

// Step 10: Save chart.

```
graph save "$graphs\E3_Retention_by_Poverty_Quartile.gph", replace
graph export "$graphs\E3_Retention_by_Poverty_Quartile.emf", replace
```

4. TEACHER TURNOVER BY TEACHER EFFECTIVENESS TERCILE



*Significantly different from bottom tercile value, at the 95 percent confidence level.
Notes: Sample includes 2007-08 through 2009-10 5th through 8th grade math teachers, with 1,055 teacher years and 587 unique teachers. Teacher effects are measured in test score standard deviations, with teacher-specific shrinkage factors applied to adjust for differences in sample reliability. Retention analysis is based on one-year retention rates.

Purpose:
Explore whether the agency is retaining its most effective teachers.

Required Analysis File Variables:

```
tid
school_year
t_is_teacher
school_lvl
t_transfer
t_leave
current_tre_math
```

Analysis-Specific Sample Restrictions:

- Keep only teachers with teacher effectiveness estimates.
- Keep only years for which next-year retention status can be calculated.
- If school-level restriction is chosen, keep only records for either elementary or middle school grades.

Ask Yourself:

- What does this analysis tell you about retention and transfer programs for the most and least effective teachers in your agency? Is your agency selectively retaining its most effective teachers?
- Is your agency working on policy initiatives to increase the difference in turnover between the most and least effective teachers?

Potential further analyses:
Because teacher effectiveness typically increases during the first few years of teaching, repeat this analysis for teachers with three or fewer years of experience, and for teachers wih more than three years of experience.

4. TEACHER TURNOVER BY TEACHER EFFECTIVENESS TERCILE

// Step 1: Choose the subject (math or ela) and school level (elem or middle) for the analysis. Note: to make multiple charts at the same time, put loops for subject and level around the analysis and graphing code. To include all grade levels in the analysis, comment out the local level command below.

```
local subject math
*local level middle
```

// Step 2: Load the Teacher_Year_Analysis file.

```
use "${analysis}/Teacher_Year_Analysis.dta", clear
isid tid school_year
```

// Step 3: Restrict the sample. Keep years for which both teacher effects value added estimates and next-year retention status are available. Keep only records for which one-year teacher effectiveness estimates are available. Keep employees who are teachers. If school level restriction is chosen, keep only records from either elementary or middle schools.

```
keep if school_year >= 2008 & school_year <= 2010
keep if t_is_teacher == 1
keep if !missing(current_tre_`subject')
if "`level'" == "elem" {
       keep if school_lvl == "Elem"
       }
if "`level'" == "middle" {
       keep if school_lvl == "Mid"
       }
```

// Step 4: Review variables.

```
assert t_leave + t_transfer + t_stay == 1
tab school_year
unique tid
codebook current_tre_`subject'
table t_stay, c(freq mean current_tre_`subject')
table t_leave, c(freq mean current_tre_`subject')
table t_transfer, c(freq mean current_tre_`subject')
```

// Step 5: Calculate effectiveness tercile using restricted sample.

```
xtile terc_current_tre_`subject' = current_tre_`subject', nq(3)
tab t_transfer terc_current_tre_`subject', mi
tab t_leave terc_current_tre_`subject', mi
```

// Step 6: Get sample sizes.

```
sum tid
local teacher_years = string(r(N), "%9.0fc")
preserve
       bysort tid: keep if _n == 1
       sum tid
```

4. TEACHER TURNOVER BY TEACHER EFFECTIVENESS TERCILE

```
	local unique_teachers = string(r(N), "%9.0fc")
restore
```

// Step 7: Calculate significance indicator variables by tercile.

```
foreach var of varlist t_leave t_transfer {
	gen sig_`var' = .
	xi: logit `var' i.terc_current_tre_`subject', robust
	forval quartile = 2/3 {
		replace sig_`var' = abs(_b[_Iterc_curr_`quartile'] / ///
			_se[_Iterc_curr_`quartile']) if terc_current_tre_`subject' == `quartile'
		replace sig_`var' = 0 if sig_`var' <= 1.96 ///
			& terc_current_tre_`subject' == `quartile'
		replace sig_`var' = 1 if sig_`var' > 1.96 ///
			& terc_current_tre_`subject' == `quartile
	}
	replace sig_`var' = 0 if terc_current_tre_`subject' == 1
}
```

// Step 8: Collapse and calculate shares.

```
collapse (mean) t_leave t_transfer sig_* (count) tid, by(terc_current_tre_`subject')
foreach var of varlist t_leave t_transfer {

	replace `var' = `var' * 100

}
```

// Step 9: Concatenate value and significance asterisk.

```
foreach var of varlist t_leave t_transfer {
	tostring(sig_`var'), replace
	replace sig_`var' = "*" if sig_`var' == "1"
	replace sig_`var' = "" if sig_`var' == "0"
	gen `var'_str = string(`var', "%9.0f")
		egen `var'_label = concat(`var'_str sig_`var')
	}

// Step 10: Generate count variable and add variables cumulatively for graphing.

gen count = _n
replace t_transfer = t_leave + t_transfer

// Step 11: Define titles for subject and school level.

if "`subject'" == "math" {
	local subj_title "Math"
	local subj_foot "math"
}
if "`subject'"=="ela" {
	local subj_title "ELA"
```

4. TEACHER TURNOVER BY TEACHER EFFECTIVENESS TERCILE

```
       local subj_foot "English/Language Arts"
}
       local gradespan "5th through 8th"
if "`level'" == "middle" {

       local level_title "Middle "

       local gradespan "6th through 8th"
}
if "`level'" == "elem" {
       local level_title "Elementary "
       local gradespan "5th"
}
```

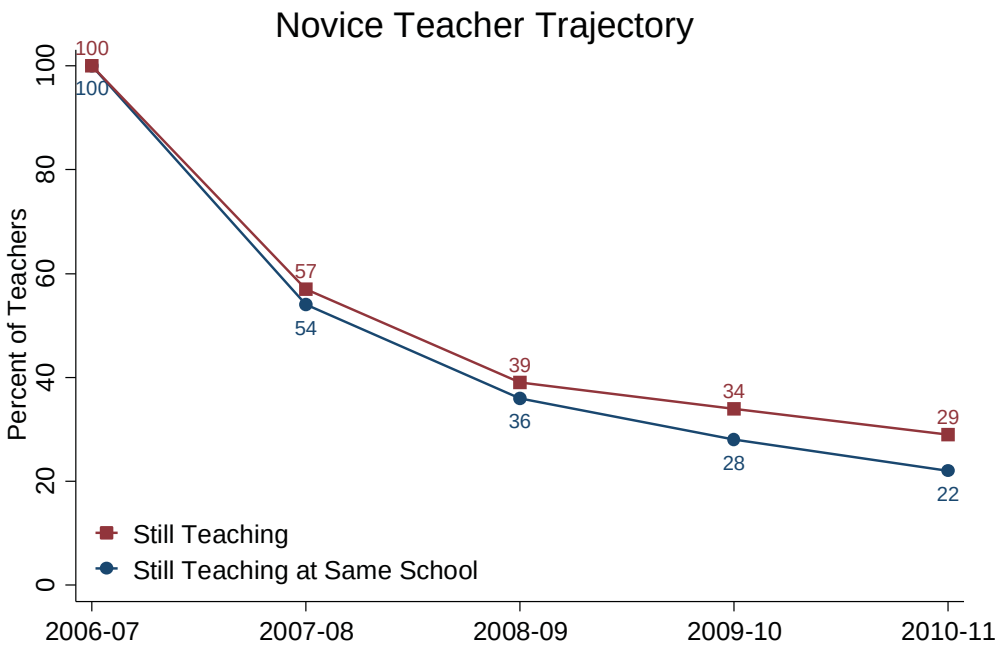
// Step 12: Make chart.

```
#delimit ;
twoway bar t_transfer count,
       barwidth(.6) color(forest_green) finten(100) ||
       bar t_leave count,
       barwidth(.6) color(maroon) finten(100) ||
       scatter t_transfer count,
              mlabel(t_transfer_label)
              msymbol(i) msize(tiny) mlabpos(6) mlabcolor(white) mlabgap(.001) ||
       scatter t_leave count,
              mlabel(t_leave_label)
              msymbol(i) msize(tiny) mlabpos(6) mlabcolor(white) mlabgap(.001) ||,
       title("Average `subj_title' Teacher Turnover", span)
       subtitle("by Teacher Effects Tercile", span)
       ytitle("Percent of Teachers", size(medsmall))
       yscale(range(0(10)60))
       ylabel(0(10)60, nogrid labsize(medsmall))
       xtitle(" ")
              xlabel(1 "Bottom Third" 2 "Middle Third" 3 "Top Third", labsize(medsmall))
       legend(order(1 "Transfer Schools" 2 "Leave")
              ring(0) position(11) symxsize(2) symysize(2) rows(2) size(medsmall)
              region(lstyle(none) lcolor(none) color(none)))
       graphregion(color(white) fcolor(white) lcolor(white)) plotregion(color(white)
              fcolor(white) lcolor(white))
       note(" " "*Significantly different from bottom tercile value, at the 95 percent
confidence level." "Notes: Sample includes 2007-08 through 2010-11 `gradespan' grade
`subj_foot' teachers, with `teacher_years' teacher years and `unique_teachers' unique"
"teachers. Teacher effects are measured in test score standard deviations, with
teacher-specific shrinkage factors applied to adjust" "for differences in sample
reliability. Retention analysis is based on one-year retention rates.",
span size(vsmall));
#delimit cr
```

// Step 13: Save chart.

```
graph save "$graphs\E4_Retention_by_Effectiveness_Tercile_`subj_title'.gph", replace
graph export "$graphs\E4_Retention_by_Effectiveness_Tercile_`subj_title'.emf", replace
```

5. NOVICE TEACHER RETENTION TRAJECTORY



Notes: Sample includes 100 teachers who were in their first year of teaching in the 2006-07 school year.

Purpose:
Describe the retention trajectory of a cohort of novice teachers.

Required analysis file variables:

```
tid
school_year
school_code
t_transfer
t_teacher
t_novice
```

Analysis-specific sample restrictions:

- Keep only years for which next-year retention status can be calculated.
- Keep only records for teachers who were novices in the first year of data.

Ask yourself:

- Does a sharp drop in retention occur in any one year? If so, what might be driving turnover at this stage of a teacher's career?
- What types of support does your agency provide to novice teachers, and for how many years do early career teachers receive additional support?

Potential further analyses:
This approach can be used to generate a number of interesting analyses. For example, you can examine retention for all teachers instead of just novices. If sample size permits, you can investigate novice retention by certification pathway or by effectiveness tercile. This latter analysis is one of SDP's **Human Capital Strategic Performance Indicators.**

5. NOVICE TEACHER RETENTION TRAJECTORY

```
// Load data.

use "${analysis}\Teacher_Year_Analysis.dta", clear
isid tid school_year

// Restrict sample to years for which next-year retention status can be observed and to teacher records with
non-missing novice indicators.

keep if school_year >= 2007 & school_year <= 2011
keep if t_is_teacher
keep if !missing(t_novice)

// Review variables.

tab school_year t_novice, mi

// Make indicator for membership in 2007 novice cohort.

gen t_novice_2007 = school_year == 2007 & t_novice == 1
bysort tid: egen max_t_novice_2007 = max(t_novice_2007)
drop t_novice_2007
rename max_t_novice_2007 t_novice_2007

// Restrict sample to 2007 novice cohort, dropping observations of teachers who reappear after leaving for one
or more school years.

keep if t_novice_2007 == 1
gen t_leave_year = school_year if t_leave == 1
bysort tid: egen min_t_leave_year = min(t_leave_year)
drop if school_year > min_t_leave_year

// Get sample size.

sum tid if school_year == 2007
local unique_teachers = string(r(N), "%9.0fc")

// Get initial school.

gen school_code_2007 = school_code if school_year == 2007
egen max_school_code_2007 = max(school_code_2007), by(tid)
replace school_code_2007 = max_school_code_2007
drop max_school_code_2007

// Define outcome variables.

gen still_same_school = school_code == school_code_2007
gen still_teach = 1
tab school_year still_same_school, mi
tab school_year still_teach, mi
```

5. NOVICE TEACHER RETENTION TRAJECTORY

```
// Collapse to sum variables of interest.

collapse (sum) still_same_school still_teach (count) tid, by(school_year)
gen cohort_count_2007 = tid if school_year == 2007
egen max_cohort_count_2007 = max(cohort_count_2007)
replace cohort_count_2007 = max_cohort_count_2007
drop max_cohort_count_2007

// Calculate outcome percentages by year.

foreach var in still_same_school still_teach {
       replace `var' = 100 * `var' / cohort_count_2007
       format `var' %9.0fc
}

// Make chart.

sort school_year
#delimit ;
twoway
scatter still_same_school school_year,
       connect(l)
       lcolor(navy)
       lpattern(solid)
       msymbol(circle)
       mcolor(navy)
       msize(medium)
       mlabel(still_same_school)
       mlabpos(6)
       mlabcolor(navy)
       mlabsize(small) ||
       scatter still_teach school_year,
       connect(l)
       lcolor(maroon)
       lpattern(solid)
       msymbol(square)
       mcolor(maroon)

              mlabsize(small)
              msize(medium)
              mlabel(still_teach)
              mlabpos(12)
              mlabcolor(maroon) ||,

       title("Novice Teacher Trajectory", span)
       ytitle("Percent of Teachers", size(medsmall))
       xtitle("")
       yscale(range(0(20)100))
       ylabel(0(20)100, nogrid format(%9.0f) labsize(medsmall))
              xscale(range(2007(1)2011))
              xlabel(2007 "2006-07" 2008 "2007-08" 2009 "2008-09" 2010 "2009-10" 2011 "2010-11",
```

5. NOVICE TEACHER RETENTION TRAJECTORY

```
        labsize(medsmall))
        legend(position(8) order(2 1) cols(1) symxsize(3) ring(0) size(medsmall)
               region(lstyle(none) lcolor(none) color(none))
               label(1 "Still Teaching at Same School")
               label(2 "Still Teaching"))
        graphregion(color(white) fcolor(white) lcolor(white)) plotregion(color(white)
               fcolor(white) lcolor(white))

note(" " "Notes: Sample includes `unique_teachers' teachers who were in their
first year of teaching in the 2006-07 school year.", span size(vsmall));
#delimit cr
```

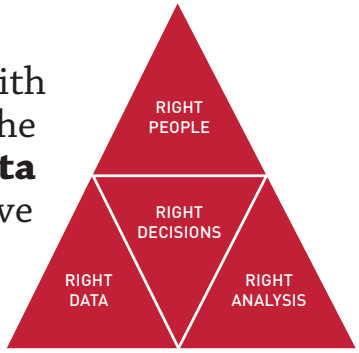
// Save chart.

```
graph save "$graphs\E5_Novice_Teacher_Retention_Trajectory.gph", replace
graph export "$graphs\E5_Novice_Teacher_Retention_Trajectory.emf", replace
```

The Strategic Data Project

OVERVIEW

The Strategic Data Project (SDP), housed at the Center for Education Policy Research at Harvard University, partners with school districts, school networks, and state agencies across the United States. **Our mission is to transform the use of data in education to improve student achievement.** We believe that with the right people, the right data, and the right analyses, we can improve the quality of strategic policy and management decisions.



SDP AT A GLANCE

56 AGENCY PARTNERS
34 SCHOOL DISTRICTS
12 STATE EDUCATION DEPARTMENTS
3 CHARTER SCHOOL ORGANIZATIONS
7 NONPROFIT ORGANIZATIONS

107 FELLOWS
65 CURRENT
42 ALUMNI

CORE STRATEGIES

1. Building a network of top-notch data strategists who serve as fellows for two years with our partners

2. Conducting rigorous diagnostic analyses of teacher effectiveness and college-going success using existing agency data

3. Disseminating our tools, methods, and lessons learned to the education sector broadly

SDP DIAGNOSTICS

SDP's second core strategy, conducting rigorous diagnostic analyses using existing agency data, focuses on two core areas: (1) college-going success and attainment for students, and (2) human capital (primarily examining teacher effectiveness).

The diagnostics are a set of analyses that frame actionable questions for education leaders. By asking questions such as "How well do students transition to postsecondary education?" or "How successfully is an agency recruiting effective teachers?" we support education leaders to develop a deep understanding of student achievement in their agency.

ABOUT THE SDP TOOLKIT FOR EFFECTIVE DATA USE

SDP's third core strategy is to disseminate our tools, methods, and lessons learned to education agencies broadly. This toolkit is meant to help analysts in all education agencies collect data and produce meaningful analyses in the areas of college-going success and teacher effectiveness. Notably, the analyses in this release of our toolkit primarily support questions related to college-going success. The data collection (Identify) and best practices (Adopt) stages of the toolkit, however, are applicable to any sort of diagnostic and convey general data use guidelines valuable to any analysts interested in increasing the quality and rigor of their analyses.

Center for Education Policy Research
HARVARD UNIVERSITY

CENTER FOR EDUCATION POLICY RESEARCH
STRATEGIC DATA PROJECT
50 CHURCH ST., 4TH FLOOR, CAMBRIDGE, MA 02138
VOX 617.496.1563
FAX 617.495.2614
WWW.GSE.HARVARD.EDU/SDP