

---

# Best of Both Worlds: Hybrid Study Designs

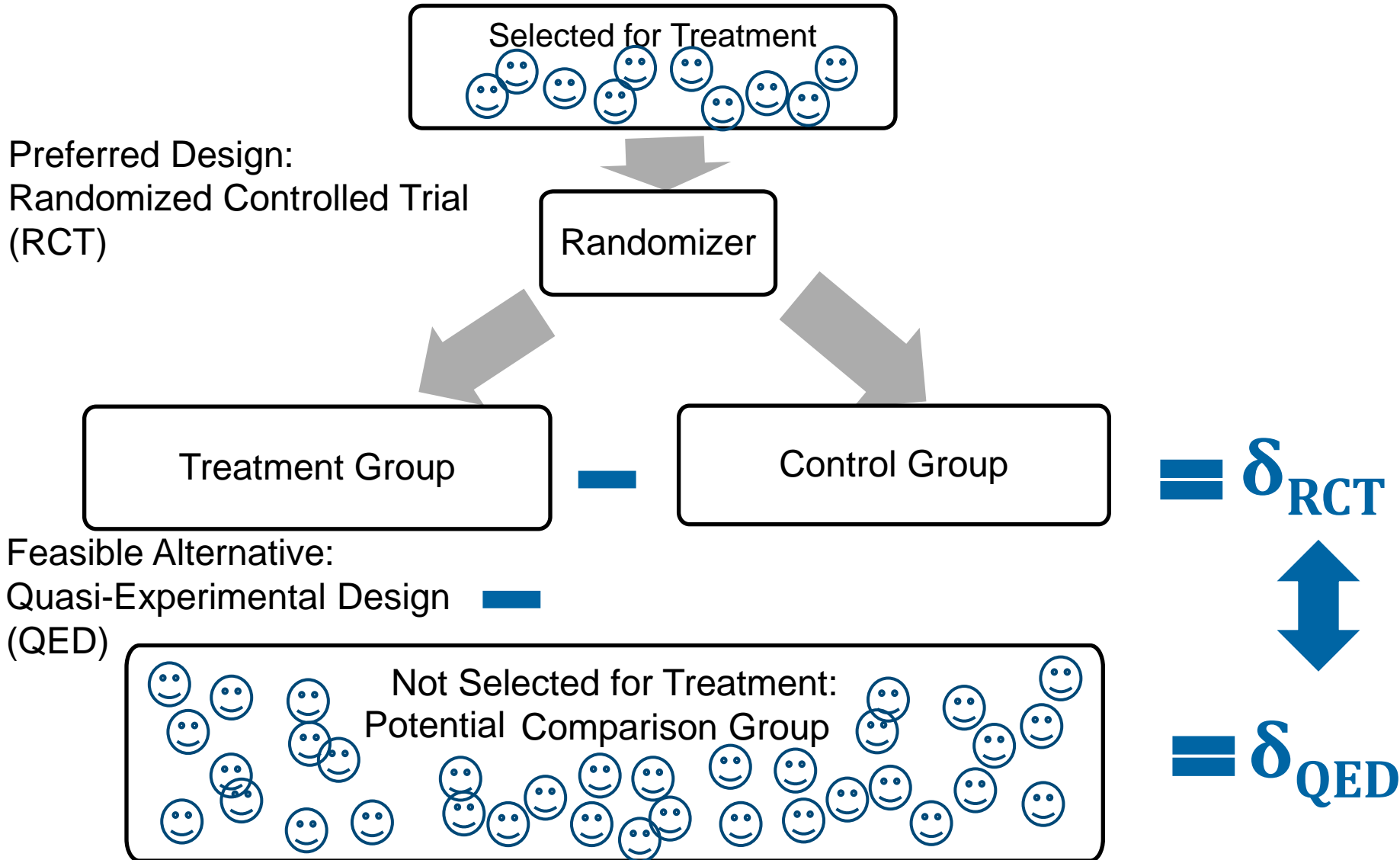
Presentation to the ACF/OPRE Methodological Advancement Meeting  
September 6, 2012

Steven Glazerman • Allison McKie Seifullah

**DRAFT – 8/29/12**

**MATHEMATICA**  
Policy Research

# Hybrid Design Illustrated



# Why Hybrid Designs?

---

- **RCT or QED?**

- Experimental design is generally preferred but not always feasible
- Different tools for different jobs
- Often have several research questions, requiring more than one design approach

- **Use overlapping designs**

- Estimate the same counterfactual using two different methods
- Compare the counterfactual estimates
- Gauge the sign and size of bias (subjective), assuming that one method starts off unbiased

---

# **Example: Evaluation of Chicago TAP: Background**

# What Is TAP?

---

- **Teacher Advancement Program**
- **Schoolwide intervention with:**
  - Performance-based pay
  - Career ladder (promotion to mentor or master teacher)
  - Professional development and mentoring (“cluster groups” and data-driven improvement)

# Chicago TAP

---

- **10 new schools per year starting in 2007–2008**
- **“Master” (called “lead”) teacher: \$15,000**
- **“Mentor” teacher: \$7,000**
- **Teacher performance awards**
  - Year 1: Target \$2,000 average payout
  - Years 2+: Target \$4,000 average payout
- **Principal performance award: \$5,000 maximum**
- **Performance awards for other school staff**
  - \$500 maximum in year 1
  - \$1,000 maximum in years 2+

# Research Questions

---

## **1. Implementation**

**What was implemented and how did it differ from business as usual?**

## **2. Impacts on student achievement**

**What impact did Chicago TAP have on math, reading, and science scores?**

## **3. Impacts on teacher retention**

**What impact did the program have on teachers' returning to the same school?**

# Selection Process: Becoming a Chicago TAP School

---

- **Self-selection**
  - Schools must apply
  - Faculty vote (75% must vote yes to participate)
- **District selection**
  - Eligibility criteria (75% of students must be eligible for free or reduced-price lunch)
  - Chicago Public Schools (CPS) staff screen the schools (“readiness” scores)
- **Random selection**
  - Finalists are assigned to implementation cohorts



# Data

---

- **Teacher surveys (every other year)**
- **Test-score data (annual)**
- **Teacher records (annual)**

# Summary of Our Findings

---

## ■ Implementation

- Non-TAP schools had considerable PD
- TAP had more mentoring and bonuses
- Payouts small, no teacher-level value-added

## ■ Student achievement

- No impacts on math or reading scores
- Science scores higher for TAP, but small samples

## ■ Teacher retention

- Positive impacts on teacher retention (early cohorts)
- Smaller/insignificant impacts for later cohorts

---

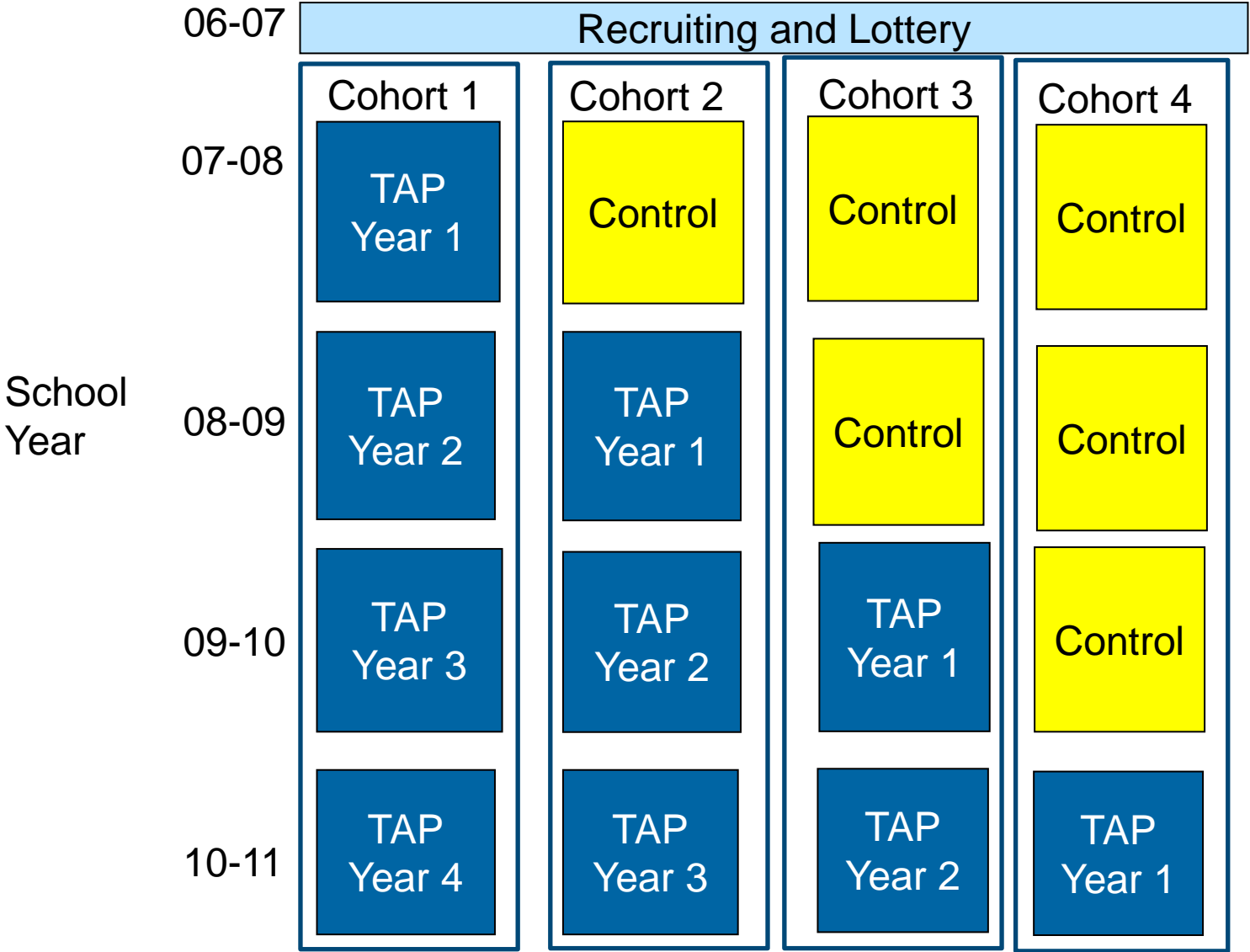
# Study Designs

# Two Design Approaches Used

---

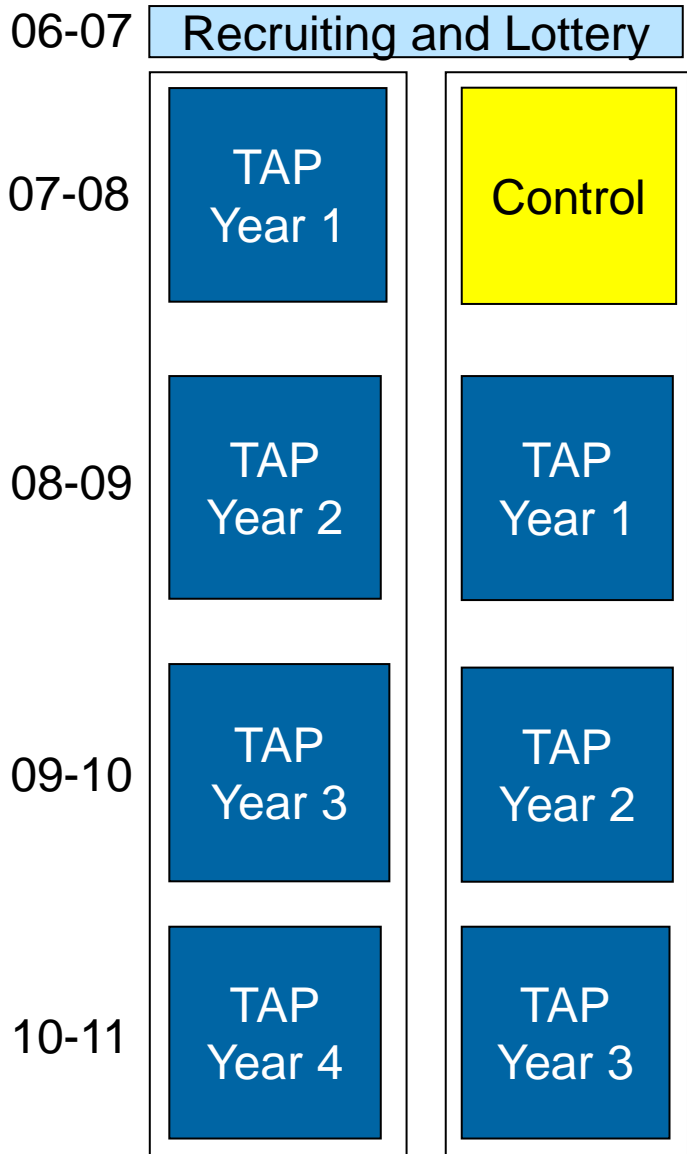
- **Experimental: Randomized rollout**
  - Assigned to delay implementation
  - Small but “clean”
  - Intention to treat (ITT) estimates
- **Quasi-experimental: Propensity-score matching**
  - Match on baseline school size, student demographics, student achievement, and teacher retention
  - Larger sample, always non-TAP
  - Potential bias from selection on unobservables

# Randomized Rollout Design (Ideal)

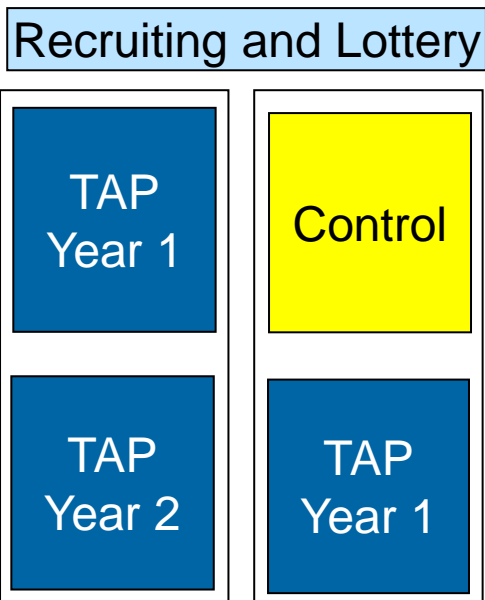


# Randomized Rollout Design (Actual)

## Cohorts 1 & 2



## Cohorts 3 & 4



# Hybrid Design

## Cohorts 1 & 2

06-07 Recruiting and Lottery **Non-TAP**

07-08 TAP Year 1 Control Matched Comparison

08-09 TAP Year 2 TAP Year 1 Matched Comparison

09-10 TAP Year 3 TAP Year 2 Matched Comparison

10-11 TAP Year 4 TAP Year 3 Matched Comparison

**Cohorts 3 & 4**  
Recruiting and Lottery **Non-TAP**

TAP Year 1 Control Matched Comparison

TAP Year 2 TAP Year 1 Matched Comparison

School Year

---

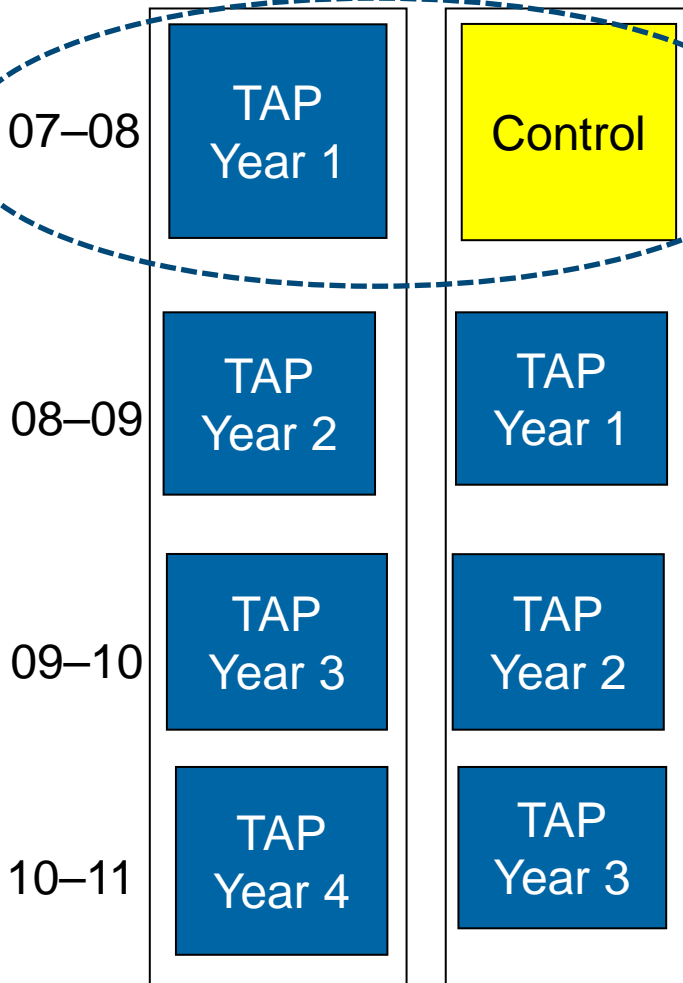
# Impacts on Students' Test Scores: Experimental Results



# Experimental Sample for First Year of Implementation

## Cohorts 1 & 2

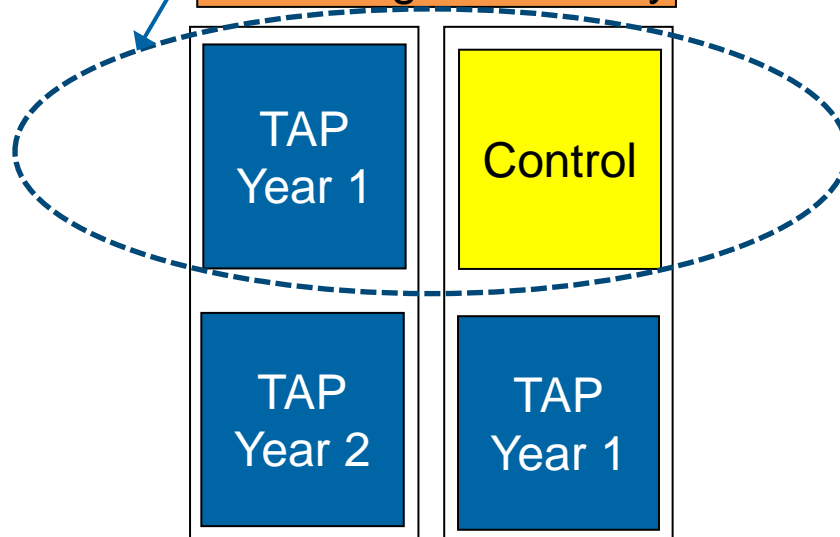
06-07 Recruiting and Lottery



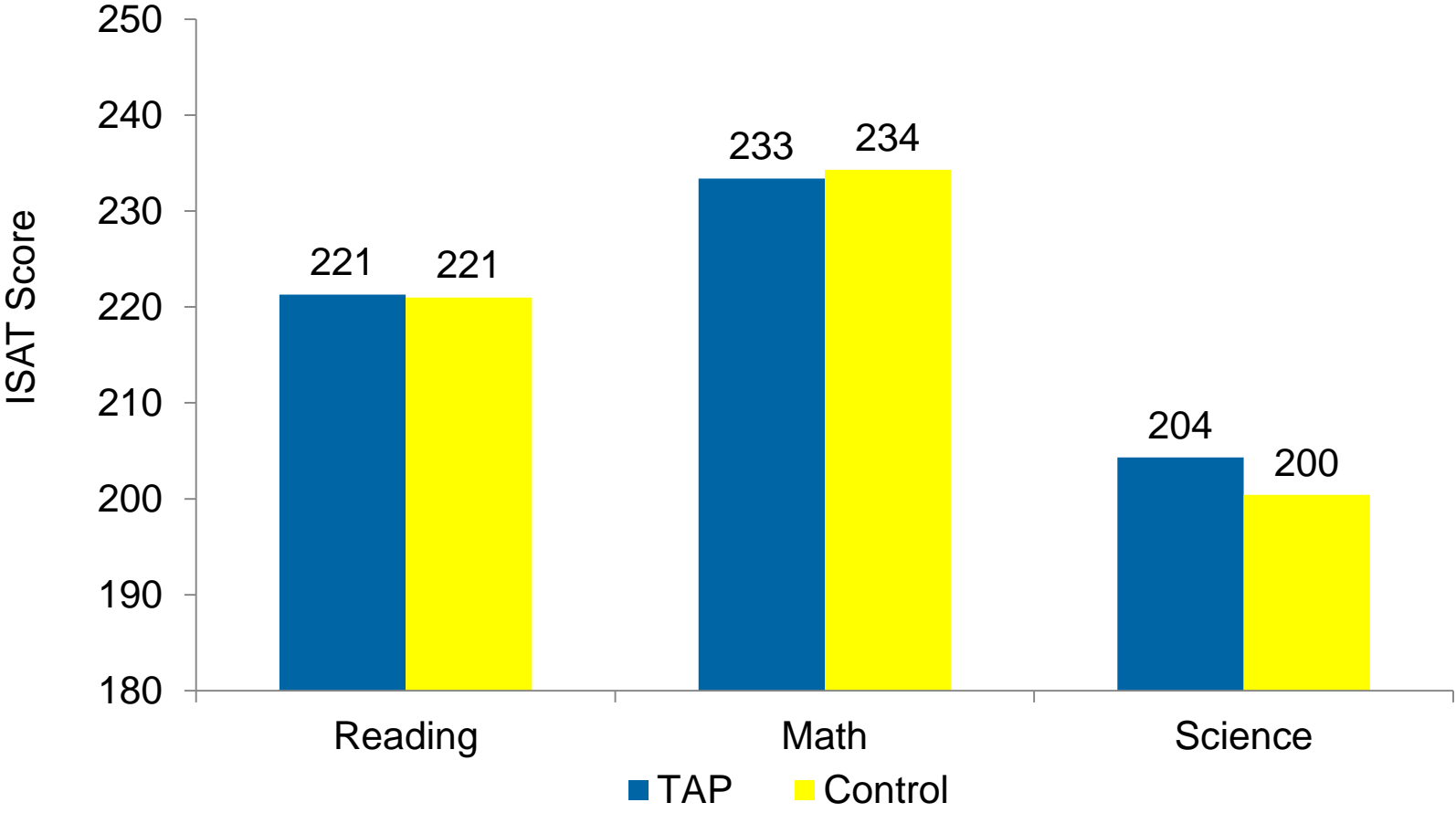
Experimental Sample

## Cohorts 3 & 4

Recruiting and Lottery

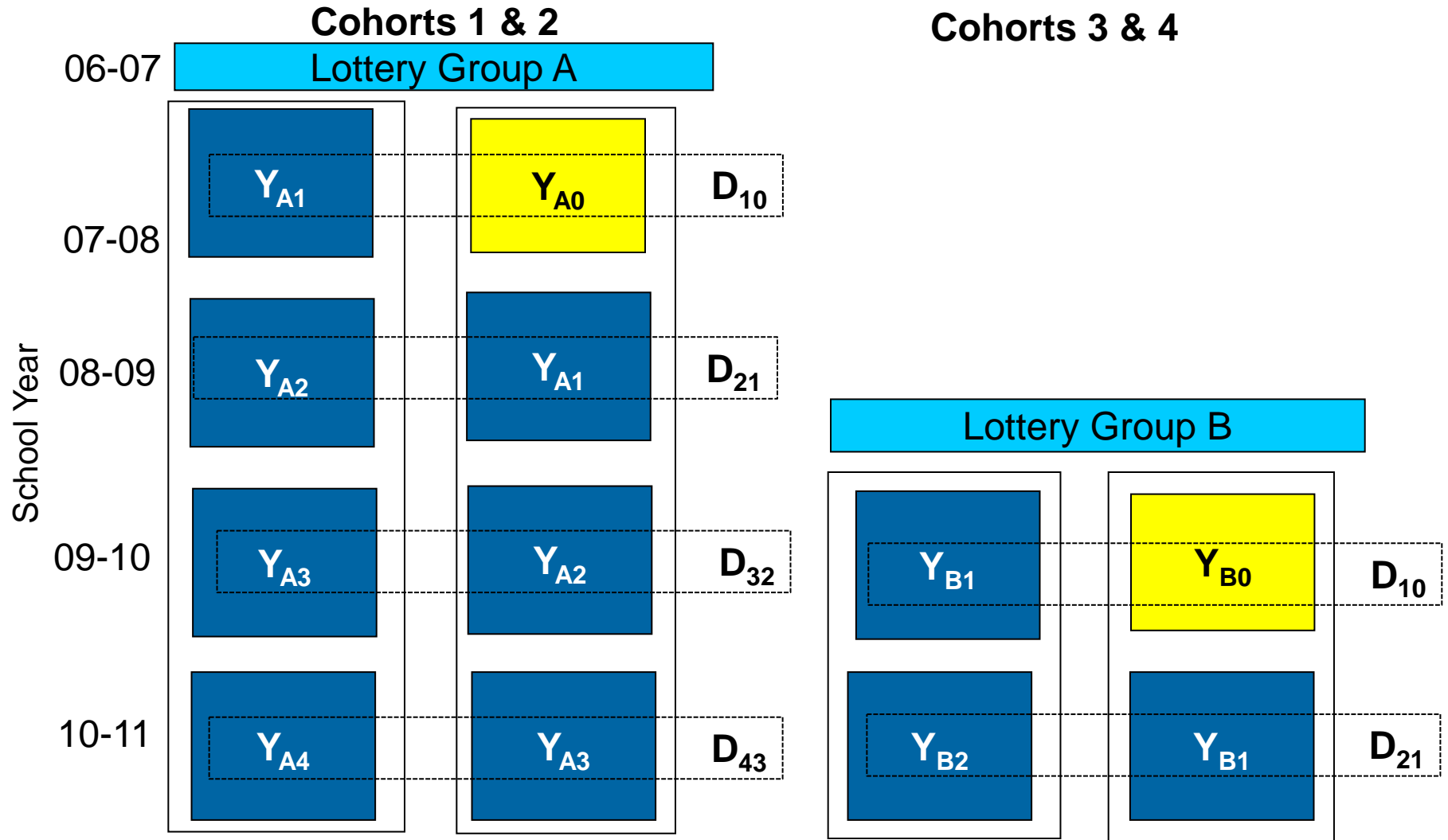


# Experimental Impact on Test Scores in First Year of Implementation



Note: Means are regression adjusted. Differences between the TAP and comparison groups are not statistically significant.

# Experimental Contrasts: Years 2–4



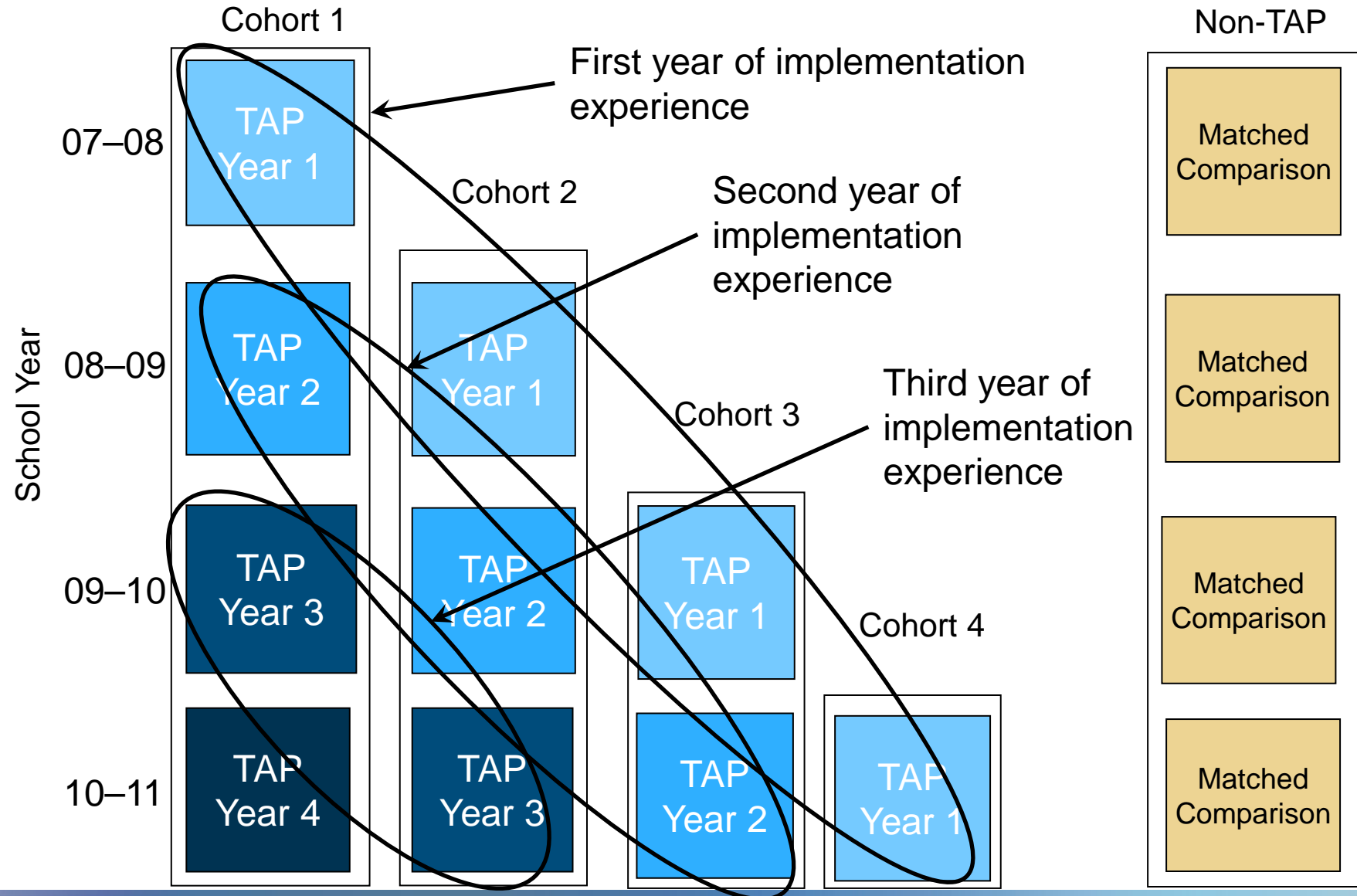
# Impacts of One Additional Year of Implementation

Implementation Years and Subject	More TAP	Less TAP	Difference
Second year vs. first year			
Reading	220.6	221.0	-0.5
Math	234.4	236.0	-1.6*
Science	203.1	204.1	-0.9
Third year vs. second year			
Reading	220.8	220.5	0.3
Math	234.7	235.5	-0.8
Science	200.1	199.4	0.7
Fourth year vs. third year			
Reading	222.4	223.5	-1.2
Math	236.7	238.7	-2.0
Science	203.8	205.4	-1.6

---

# Impacts on Student Test Scores: Quasi-Experimental Results

# Cohort, Year, and Experience Effects



# Did the Impact of TAP Grow as Schools Gained More Experience Implementing It?

TAP Effects by Year of Implementation	Reading	Math	Science
First year (all four cohorts)	-0.2	0.4	1.3
Second year (cohorts 1-3)	-0.1	-0.1	0.1
Third year (cohorts 1-2)	0.4	0.2	1.9
Fourth year (cohort 1 only)	-0.7	-1.3	2.1

**No.**

# Did the Impact of TAP Grow as CPS Gained More Experience Implementing It?

TAP Effects by School Year	Reading	Math	Science
Year 1: 2007-2008 (cohort 1)	-0.6	0.2	NA
Year 2: 2008-2009 (cohorts 1-2)	-0.4	0.3	-0.4
Year 3: 2009-2010 (cohorts 1-3)	1.3*	1.0	1.6
Year 4: 2010-2011 (all four cohorts)	-0.7	-0.5	0.3

**No.**



# Was the Impact of TAP Larger for Some Cohorts Than for Others?

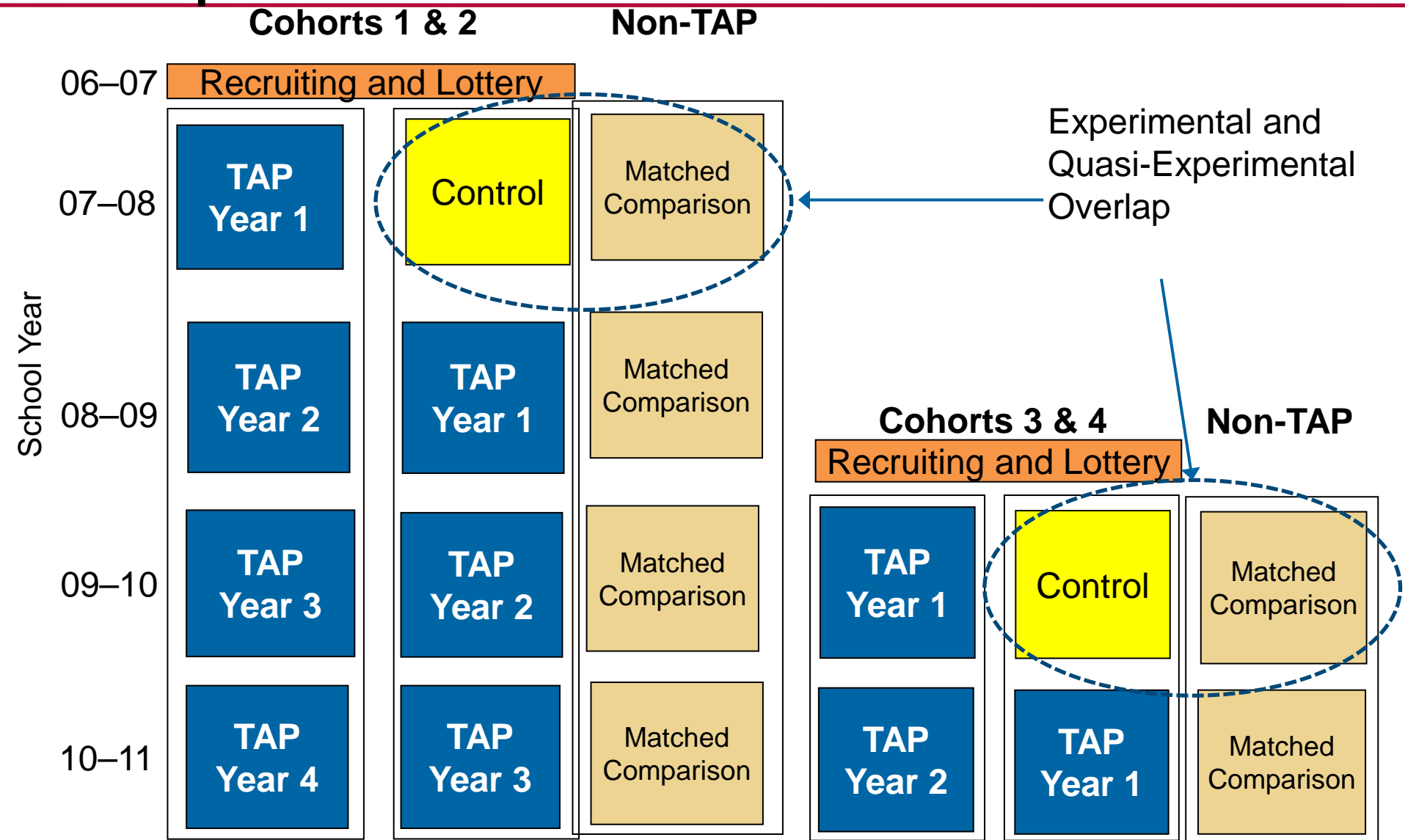
TAP Effects by Cohort	Reading	Math	Science
Cohort 1 (all four years)	-0.4	-0.4	1.7
Cohort 2 (2009-2011)	0.5	1.2*	0.7
Cohort 3 (2010-2011)	0.3	-0.4	1.3
Cohort 4 (2011 only)	-1.6*	-0.1	0.0

**Probably not.**

---

# Reconciling the Experimental and Quasi-Experimental Results

# Reconciling the Experimental and Quasi-Experimental Estimates



# Did the Experimental and Quasi-Experimental Results Match?

Matched-comparison mean minus randomized-control mean

Year and Matching Method	Reading	Math	Science
2007-2008 (cohorts 1 & 2)			
Nearest five neighbors	-0.6	-2.8*	NA
Caliper (radius = 0.025)	-0.4	-2.0*	NA
Kernel density	-0.4	-2.0*	NA
2009-2010 (cohorts 3 & 4)			
Nearest five neighbors	-0.1	-0.9	0.5
Caliper (radius = 0.025)	-0.1	-0.7	-0.2
Kernel density	0.1	0.3	2.1

# What Do the Overlapping Results Tell Us?

---

- **Two different findings**
  - Cohort 1/2: RCT > QED
  - Cohort 3/4: No detectable difference
- **Even if the cohort 1/2 results apply to the full sample, the story doesn't change**
- **Will always have to worry about generalizing: what is different about the selection mechanisms?**

# For More Information

---

- **Please contact**
  - **Steven Glazerman**
    - [sglazerman@mathematica-mpr.com](mailto:sglazerman@mathematica-mpr.com)
  - **Allison McKie Seifullah**
    - [aseifullah@mathematica-mpr.com](mailto:aseifullah@mathematica-mpr.com)