

Institute for  
Fiscal Studies

---

# Evaluation design for Achieve Together

Ellen Greaves and Luke Sibieta

# Achieve Together

TeachFirst

tl teachingleaders  
Take a lead Make an impact Change lives



- Bring together three programmes in a school:
  - Teach First
  - Teaching Leaders
  - Future Leaders
- Intensive human capital investment
- Original motivation was also to encourage schools to work together and to engage the local community and organisation in school-improvement
  - Cluster-design
  - Difficult to evaluate quantitatively
- Evaluation and pilot funded by the Education Endowment Foundation (EEF)

# Outline

- The original design of the evaluation
- What went wrong
  - Design of the pilot
  - Recruitment (round 1)
  - Recruitment (round 2)
- Final design of the evaluation
- Lessons for evaluators

# Achieve Together

TeachFirst

 teachingleaders  
Take a lead Make an impact Change lives



- Two pilots:
  1. Area-based design
  2. School-level human capital investment

# Achieve Together

TeachFirst

tl teachingleaders  
Take a lead Make an impact Change lives



- Two pilots:
  1. Area-based design
    - One-cluster in Bournemouth
    - 4 primary schools and 6 secondary schools
    - Involvement of local community/organisations
    - Process evaluation
  2. School-level human capital investment

# Achieve Together

TeachFirst

tl teachingleaders  
Take a lead Make an impact Change lives



- Two pilots:
  1. Area-based design
  2. School-level human capital investment
    - School-level intervention
    - No co-ordination within clusters or involvement of external organisations
    - Quantitative evaluation and process evaluation

# Original evaluation design

- Randomised controlled trial
- Number of schools fixed by EEF: 24 treatment and 24 control
- Primary outcomes
  - Attainment at KS4
  - Attainment at Year 7 (focus of Achieve Together impact project)
- Secondary outcomes
  - Number of persistent absentees
  - Overall absence rate

# Original evaluation design

- Randomised controlled trial
- Number of schools fixed by EEF: 24 treatment and 24 control
- Primary outcomes
  - Attainment at KS4
  - Attainment at Year 7 (focus of Achieve Together impact project)
- Secondary outcomes
  - Number of persistent absentees
  - Overall absence rate
- Subgroups
  - Pupils eligible for free school meals
  - Pupils with low prior attainment
- “Business as usual” in control schools
  - Able to access one programme element of Achieve Together



# Power calculations

	0	0.1	0.2	0.3	0.4	0.5
Model 1	0.048	0.203	0.283	0.345	0.398	0.444
Model 2	0.052	0.220	0.306	0.373	0.430	0.480
Model 3	0.044	0.186	0.259	0.315	0.363	0.406

Note: These calculations represent the effect size that will be possible to detect using a two-sided hypothesis test with significance level of 5%, and with power against an alternative hypothesis of 80%. Model 1 reports the minimum detectable effect size when the variance of the outcome unexplained by attributes of the pupils (including prior attainment) is 60%. Model 2 reports a less optimistic scenario (70% unexplained), whilst Model 3 is more optimistic (50% unexplained).

# What went wrong: design of the pilot

- School-level RCT began to look clustered...
  - Cluster based recruitment
  - Co-ordination between schools
- Complicates and creates risks for evaluation:
  1. What can we learn from the evaluation?
  2. How will the power calculations be affected?

# What went wrong: design of the pilot

- School-level RCT began to look clustered...
  - Cluster based recruitment
  - Co-ordination between schools
- Complicates and creates risks for evaluation:
  1. What can we learn from the evaluation?
    - Is positive impact due to the human capital approach?
    - Or better co-ordination between schools?
    - Our findings would be inconclusive
  2. How will the power calculations be affected?

# What went wrong: design of the pilot

- School-level RCT began to look clustered...
  - Cluster based recruitment
  - Co-ordination between schools
- Complicates and creates risks for evaluation:
  1. What can we learn from the evaluation?
  2. How will the power calculations be affected?
    - At the extreme, we can think of the unit of treatment as the cluster
    - Uncertain risk for the minimum detectable effect size
    - Required treatment effect from power calculations with clustering at the school level already looked ambitious...
    - Clustering may increase the intra-cluster correlation and increase the challenge of detecting a significant effect

# What went wrong: recruitment (round 1)

- Target: 48
- Recruited: 13
- Problems for recruitment:
  - Time available
  - Uncertainty about staff availability
  - Uncertainty about school budget (for costly programme)
  - Risk of being allocated to control group
  - Clarity about the pilot
- The recruited schools began Achieve Together in September 2013

# What went wrong: recruitment (round 2)

- Target: 48
- Recruited: 15
- Problems for recruitment:
  - Time available
  - Uncertainty about staff availability
  - Uncertainty about school budget (for costly programme)
  - Risk of being allocated to control group
  - Clarity about the pilot
- The recruited schools will begin Achieve Together in September 2014

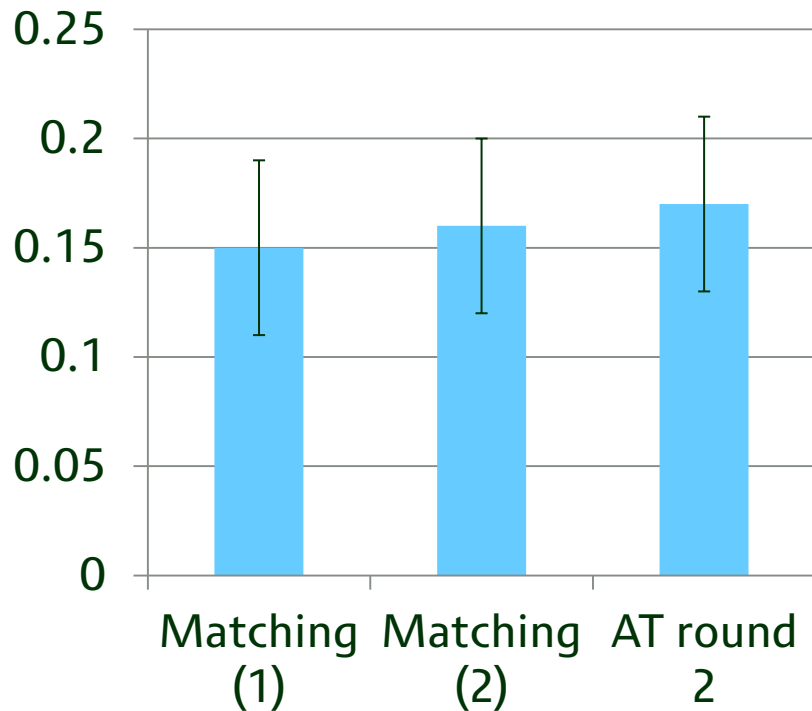
# Final evaluation design

- Non-experimental
  - Matching (“well-matched comparison group”)
    1. Similar in terms of observable characteristics
    2. Expressed a strong interest in Achieve Together
- How credible are the non-experimental estimates?
  - Depends on the factors that determine take-up and growth in pupil attainment - observable or unobservable?
- Assess the credibility of the non-experimental matching estimates
  - Achieve Together round 1 schools: compare matching estimates to a “gold standard” comparison group - schools that are similar in both observable and unobservable characteristics
  - Achieve Together round 2 schools

# Final evaluation design

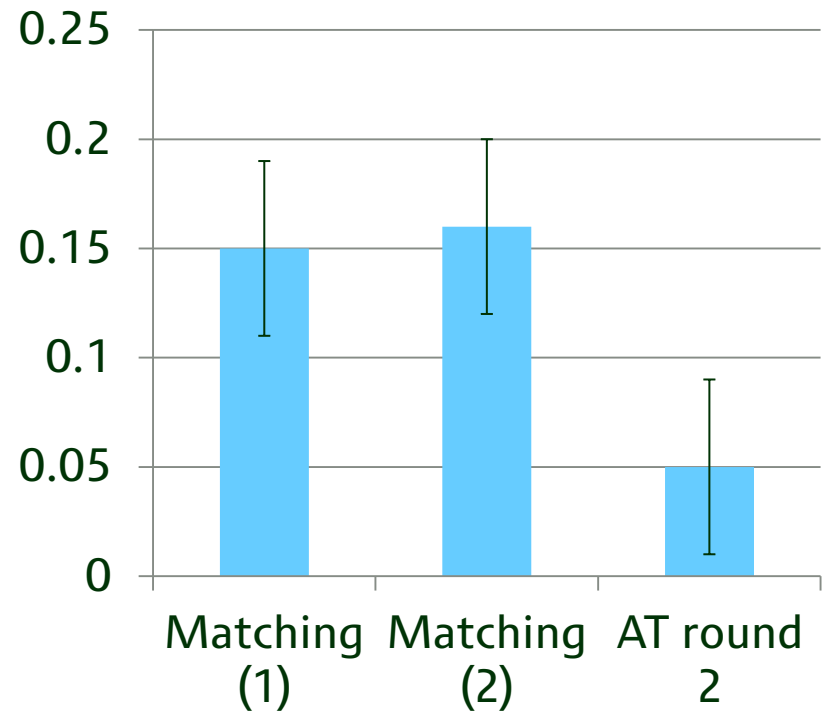
## Matching likely to be credible

### Effect size



## Matching unlikely to be credible

### Effect size





# Lessons for evaluators (1)

- Evaluators must have good communication with the project team
  - How are plans for the pilot developing?
  - What are the implications for the evaluation design?
  - Why is the evaluation important?
- Evaluators should be clear about the necessary requirements for the evaluation
  - What is expected of control schools?
    - Restrictions on “business as usual”
  - What is expected of treatment schools?
    - Additional testing
    - Involvement with process evaluation
  - What are non-negotiable elements of the evaluation

## Lessons for evaluators (2)

- Recruitment can be difficult!
  - What barriers does the evaluation impose and can these be reduced?
- Be creative
  - What evaluation design is feasible as circumstances change?
- Be selective!
  - What is the potential for a robust and informative evaluation?
  - What are the risks to the evaluation?