



## Using data to inform education policy

---

Luke Sibieta, Senior Research Economist

Institute for Fiscal Studies

Email: [luke\\_s@ifs.org.uk](mailto:luke_s@ifs.org.uk)

# Good data is important, but is not enough

- **Data is critical to teaching us about the world we live in**
  - What's the level of unemployment? How are average living standards changing? Which pupils have the highest and lowest levels of attainment?
- **Data can provide an evidence base to illustrate the most effective (and least effective) policy interventions.**

## Two key ingredients to evaluating policy interventions

### 1. Rich data on participants and non-participants

### 2. Valid Counterfactual

- How would outcomes have evolved in the absence of the policy intervention?
- Randomised control trials, Natural Experiments are ideal

**Good Example:** IFS research examining academic achievement of summer-born children

# Wide-range of data available in UK

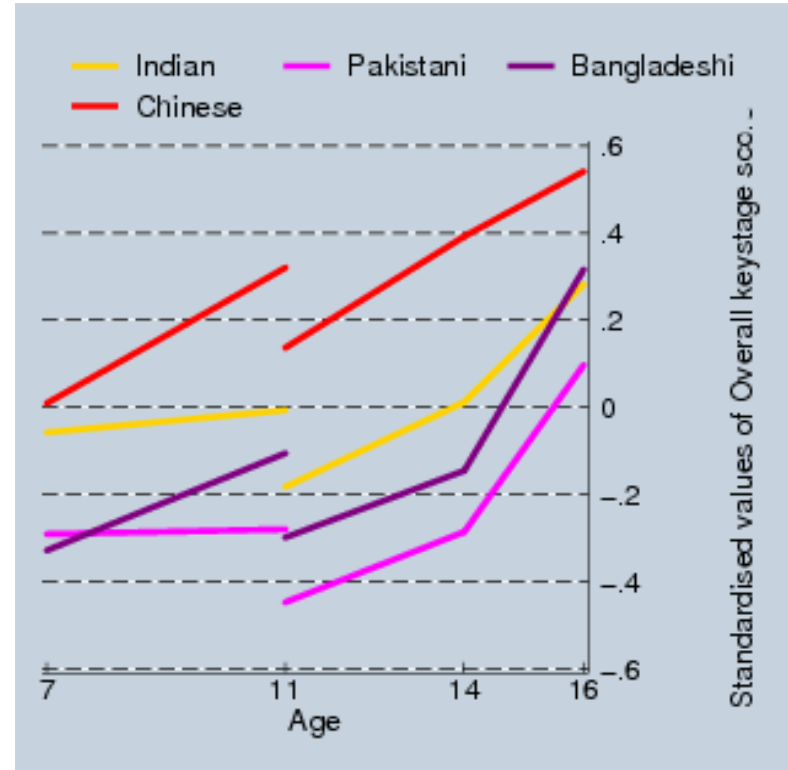
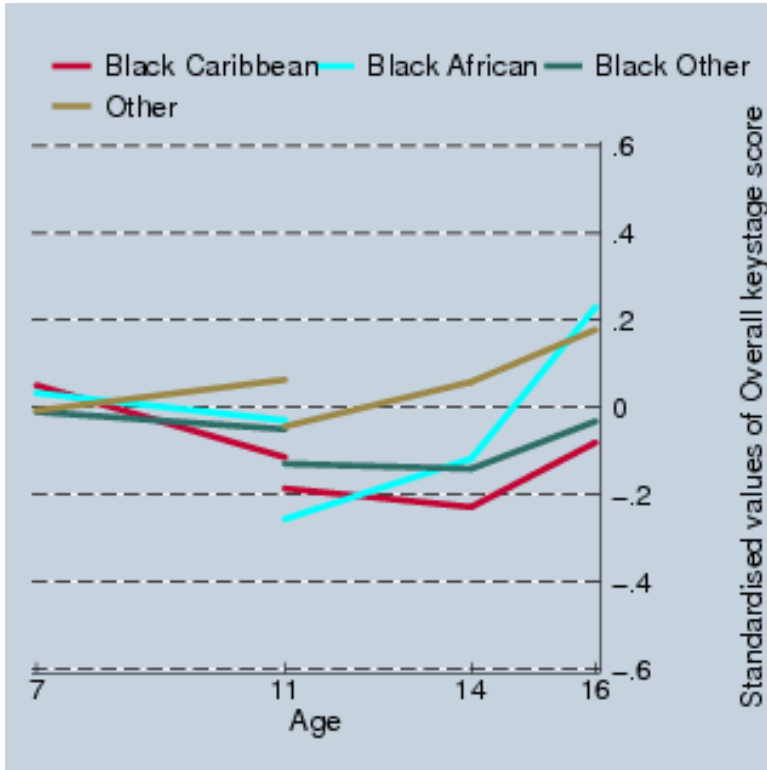
- **National Pupil Database**

- National test scores and teacher assessments (Key Stages 1 to 5) for all pupils state schools in England
- Pupil and School Characteristics
- Limited information on family background (eligibility for free school meals)
- Less data available for Wales, Scotland and Northern Ireland

# What is now possible because of the NPD?

- How do pupil characteristics affect pupil attainment?
- How do the dynamics of pupil attainment vary by ethnicity?
- How segregated are schools in England?
- How likely are pupils from different backgrounds to go to grammar schools?
- How much choice of schools do parents have?
- What is the impact of schools competition?
- What is the house price premium associated with ‘good’ schools?
- Are teacher assessments systematically biased?
- How effective are different types of schools?
- Have academies improved academic attainment?
- *Has policy x improved pupil attainment?*

# Example: Dynamics of Attainment by Ethnicity



Source: Briggs, Burgess and Wilson (2006)

# Wide-range of data available in UK

- **National Pupil Database**
  - National test scores and teacher assessments (Key Stages 1 to 5) for all pupils state schools in England
  - Pupil and School Characteristics
  - Limited information on family background (eligibility for free school meals)
  - Less data available for Wales, Scotland and Northern Ireland
- **School Workforce Census (2010 onwards)**
  - Characteristics of teachers and other staff in state schools in England (e.g. Pay, experience, qualifications)
- **Cohort Studies**
  - National Child Development Study (1958) and British Cohort Study (1970)
  - Millennium Cohort Survey (2000/01)
  - Longitudinal Study of Young People in England, LSYPE (1989) – halted at age 19
  - New LSYPE established for 14 year olds in 201x – same age as MCS!

# Pupil Premium: great data, but hard to evaluate

- **Aim:** narrow socio-economic achievement gap by providing extra funding for schools with more disadvantaged pupils (£2.5bn by 2014-15)
- We have great data on achievement of children from different backgrounds
- We have detailed data on funding and spending decisions of individual schools
- Very hard to assess if the pupil premium has narrowed the achievement gap
  - No way of knowing how the achievement gap would have evolved in the absence of the pupil premium as policy was rolled out nationally (i.e. no clear counterfactual)
  - Other policies are explicitly designed to reduce the achievement gap
- There is already extra funding for pupils from disadvantaged backgrounds and we don't really know its effect on achievement gap!

# EMA: good attainment data, good counterfactual, lack of long-run data

- EMA: extra cash paid to pupils from poorer households conditional on attending education after age 16 (closed to new applicants after 2010)
- Great data on participation and attainment of all pupils in England
- Good counterfactual: pilot areas randomly selected before national roll-out
- Increased participation rates of eligible 17-year olds from 54% to 61%
- No way of assessing long-run impact on outcomes in the labour market
- Detailed data exists on labour market outcomes, but is not linked to educational experience



# Access to Higher Education: lack of linked data

- Why do fewer pupils from disadvantaged backgrounds attend prestigious universities?
- Need good data on pupils school records, where they go to university and where they applied
- All this data exists individually
  - National Pupil Database, Higher Education Statistics Agency, UCAS
- NPD-HESA linked data is very hard to get hold of and can only be used for projects funded by DfE/BIS
- UCAS data linked to NPD-HESA is not publicly available for researchers

# Teachers: new data emerging, but no linked data...

- Academic consensus surrounding importance of good teachers
- Little evidence on how to attract the best teachers.
- Historically, data on teachers has been poor
- New School Workforce Census provides detailed data on teachers from 2010
- No way of linking individual teachers to pupil attainment
- Lack of linked data makes it very hard to conduct research on critical subject of teacher quality

# UK is behind the curve on pupil-teacher linked data

Chetty, Friedman and Rockoff (2011) use US data to conclude:

*“Replacing a teacher whose VA is in the bottom 5% with an average teacher would increase the present value of students’ lifetime income by more than \$250,000 for the average class- room in our sample.”*

Details of twin sources of data:

*“The first is a dataset on test scores and classroom and teacher assignments in grades 3-8 from a large urban school district in the U.S. These data cover more than 2.5 million students and 18 million tests for math and English (reading) spanning 1989-2009. The second is selected data from United States tax records spanning 1996-2010. These data contain information on student outcomes such as earnings, college attendance, and teenage births as well as parent characteristics”*

# Great example of linked data from North Carolina

## North Carolina Education Research Data Center:

- “the center has received data from the NC Department of Public Instruction (NCDPI) on every district, school, teacher, and student in the North Carolina public school system from the mid1990s to the present. ...By pooling these resources and standardizing data across datasets, the Data Center provides a wide variety of information. **The Data Center has created longitudinal student and teacher databases, which allow researchers to follow students and teachers over time and link their records across files.** For example, teacher's qualifications can be linked to school characteristics, and student academic performance can be associated with district attributes.”

# Concluding Thoughts

- UK has very good administrative data that has allowed researchers to provide new insights
- Education Endowment Fund set up to trial school-level interventions using Randomised Control Trials
- National Policies are not piloted in ways that allow researchers to evaluate their impact
- Lack of long-run and linked data hampers our ability to answer important research questions