Comparative research with national birth cohort studies

A case study for the US, the UK, Australia and Canada

Dr Liz Washbrook
University of Bristol

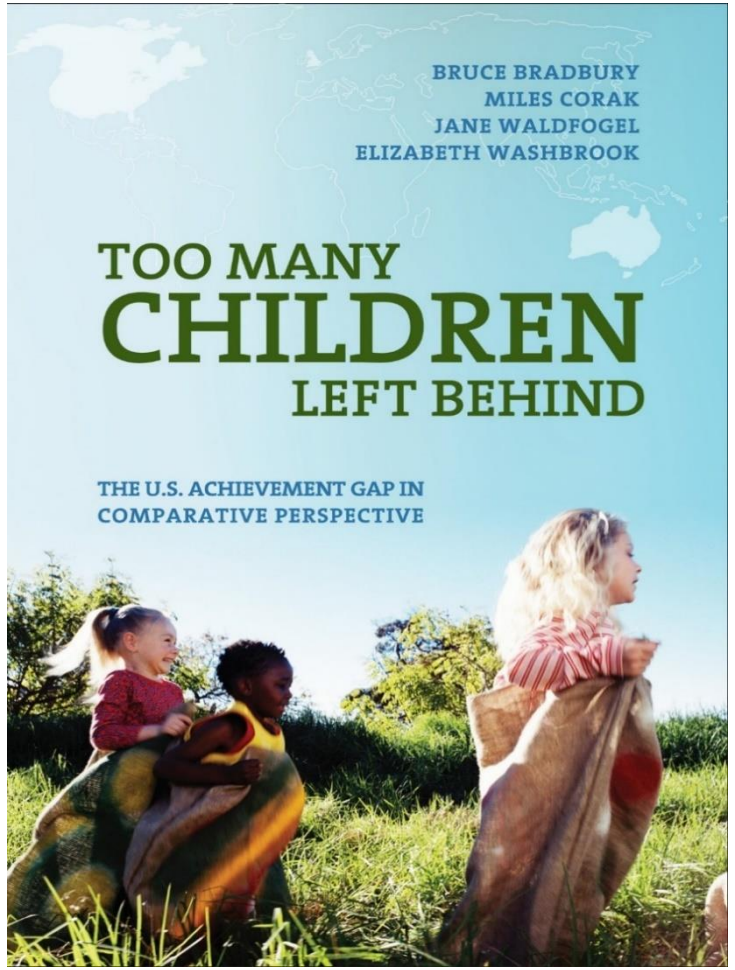
ILSA Workshop, NAEd, 17th June, 2016

What can we learn from comparison of socioeconomic status (SES) gaps in achievement across countries?

In a 2015 book for Russell Sage Foundation, Bruce Bradbury, Miles Corak, Jane Waldfogel, and I examine this question using cohort data from Australia, Canada, US, UK.

We used national longitudinal surveys in which children were all assessed at ages 5, 7/9, and 11 (at a minimum)

These datasets were not designed for the purpose of cross-national comparisons!



The cohorts

	US	UK	Australia	Canada
Dataset	ECLS-K (Early Childhood Longitudinal Study – Kindergarten Cohort)	MCS (Millennium Cohort Study)	LSAC-K (Longitudinal Study of Australian Children – Kindergarten Cohort)	NLSCY (National Longitudinal Study of Children and Youth)
Mean age at "age 5" assessment	5.7 years	5.2 years	4.9 years	4.9 years
Year of age 5 assessment	1998	2006	2004	1996-1999
Analysis sample size	8,370	11,762	3,940	4,346

Outcome measurement

- Instruments used to assess achievement and socio-emotional outcomes in children differ **across** countries at a given age, and **within** a country at different ages
- Example: measures of language/reading skills

	US	Australia
Age 5	ECLS-K kindergarten-1st grade reading test	PPVT receptive vocabulary test
Age 11	ECLS-K 5th grade reading test	NAPLAN national reading assessment

- To make comparisons, we have to argue that the instruments are measuring the same latent constructs (at least at a given age)

Implications for comparisons

- Outcomes are standardized within-country to mean zero, unit variance z-scores at each age
  - We cannot compare **levels** of achievement across countries at a given age (are US 3rd graders smarter than Canadian 3rd graders? ✗)
  - We cannot talk about **growth** in 'absolute' levels of skills as children age (do children learn more in elementary school in the UK than in Australia? ✗)
- What we can compare is how position in the national distribution at each age differs, on average, with SES
  - Does the **gap** in reading ability between low- and high-SES children differ across countries? ✓

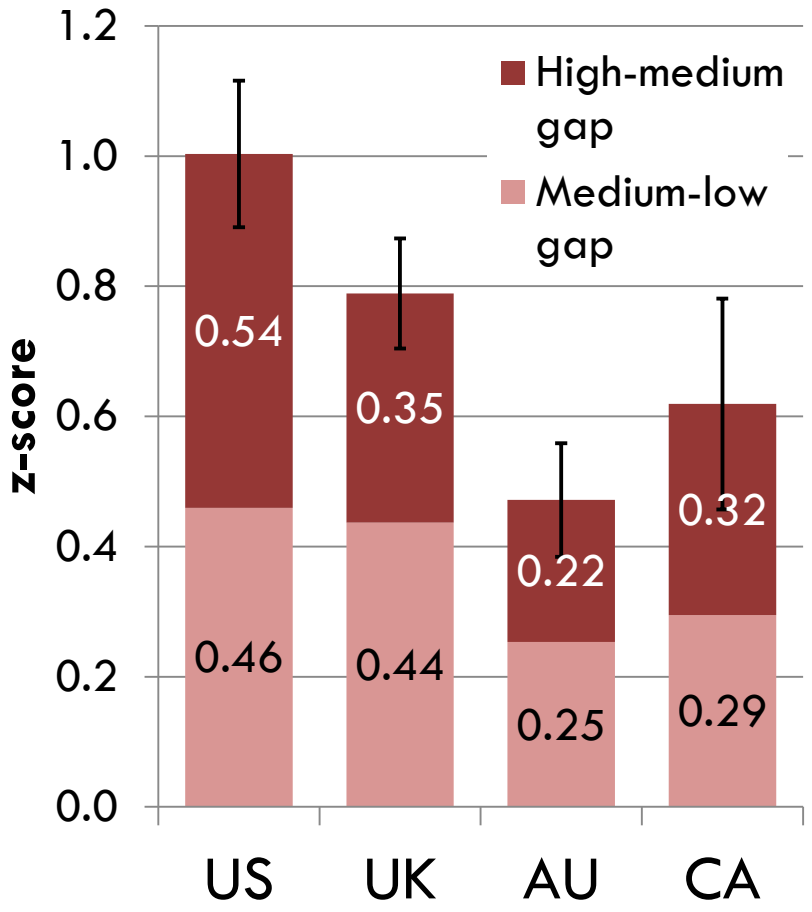
Given the inherent limitations – why do it? (1)

- Growing body of literature about the crucial importance of early life conditions in human development (e.g. work of James Heckman)
- No large scale international comparative studies assess children below the age of 10
- SES gaps at age 15 conflate two factors
  - Inequality in the skills children brought with them when they **started** school
  - Extent to which those inequalities widened or narrowed **during the school years**
- Disentangling the two is necessary for assessment of national school systems and lots of other policy questions
  - How do countries differ in the first aspect? In the second? Is something better possible?

Given the inherent limitations – why do it? (2)

- Cohort studies (potentially) allow for **longitudinal modelling of trajectories** as well as snapshots of SES gaps at different ages
- Suppose we want to decompose the SES gap in high school into two components: a part due to inequality at school entry and a part due to inequality in progress
- Questions such as this cannot be answered with data from repeated cross-sections of children at different ages, but only with longitudinal data

SES gaps in age 5 language outcomes



High SES = parent with 4-yr college or more

Medium SES = parent with some college

Low SES = parent with high school or less

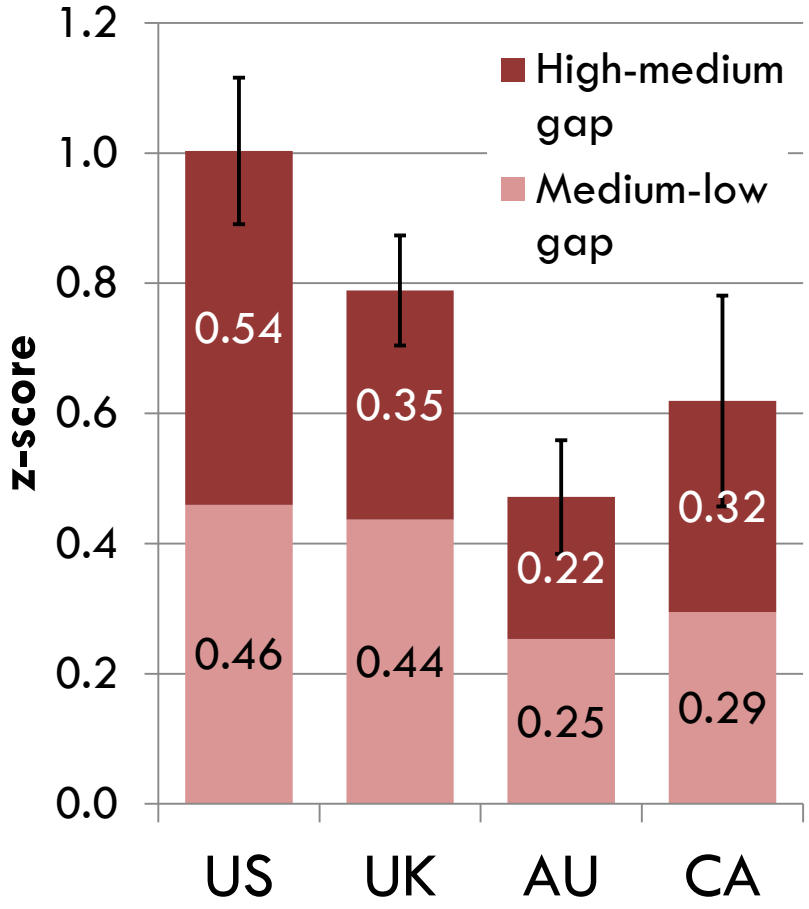
The overall height of each bar is the **total SES gap** – the difference in the mean scores of **high and low SES** children

The total gap is made up of:

- The gap between **medium and low SES** children (lighter bars)
- The gap between **high and medium SES** children (darker bars)

Error bars show 95% confidence intervals around the total high/low gap

Interpreting the SES gaps



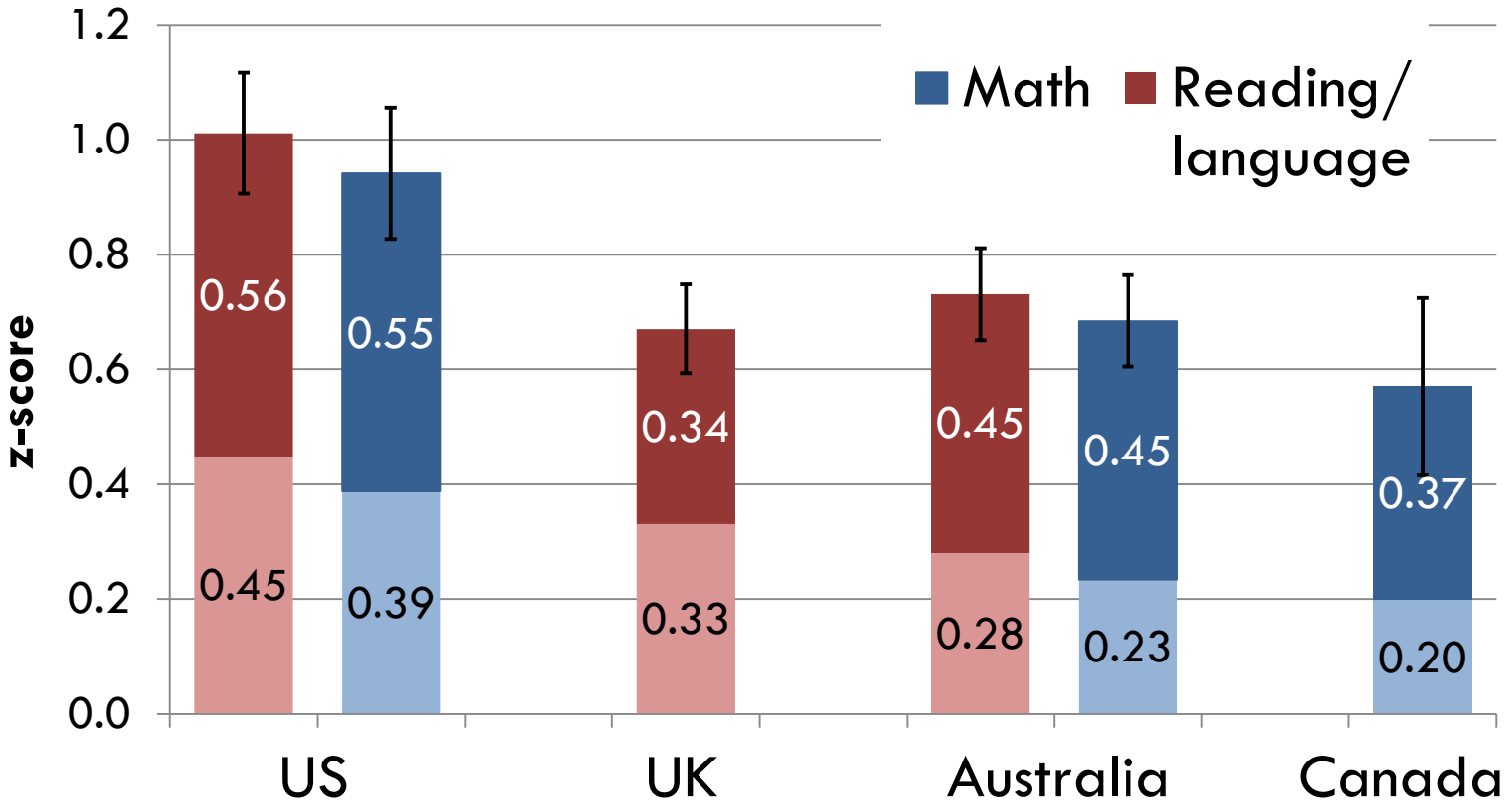
In what sense is a 1 SD gap in the US 'the same' as a 1 SD gap in Australia?

By definition, they are the same in a relative (positional) sense

But they will also be the same in an absolute sense if the **variance** in the underlying outcome is the same across countries (even if mean levels of achievement differ)

Are they the same? The best we can do is benchmark using the closest equivalents from truly comparative international studies (e.g. PIRLS 4th grade)

SES gaps in age 11 academic achievement outcomes



Light-shaded bars are medium/low SES gap. Dark-shaded bars are the high/medium SES gap. **The overall height of the bar is gap between high and low SES children.**

Thinking longitudinally…only about half of all US children were in the same achievement quartile in 8th grade as they were in kindergarten

	Quartile of reading score in 8th grade					
Quartile of reading score in kindergarten	**1 (Lowest)**	**2**	**3**	**4 (Highest)**		**Total (%)**
1 (Lowest)	**53**	25	15	7		100
2	29	**36**	22	13		100
3	14	24	**34**	28		100
4 (Highest)	4	15	29	**52**		100
All children	25	25	25	25		100

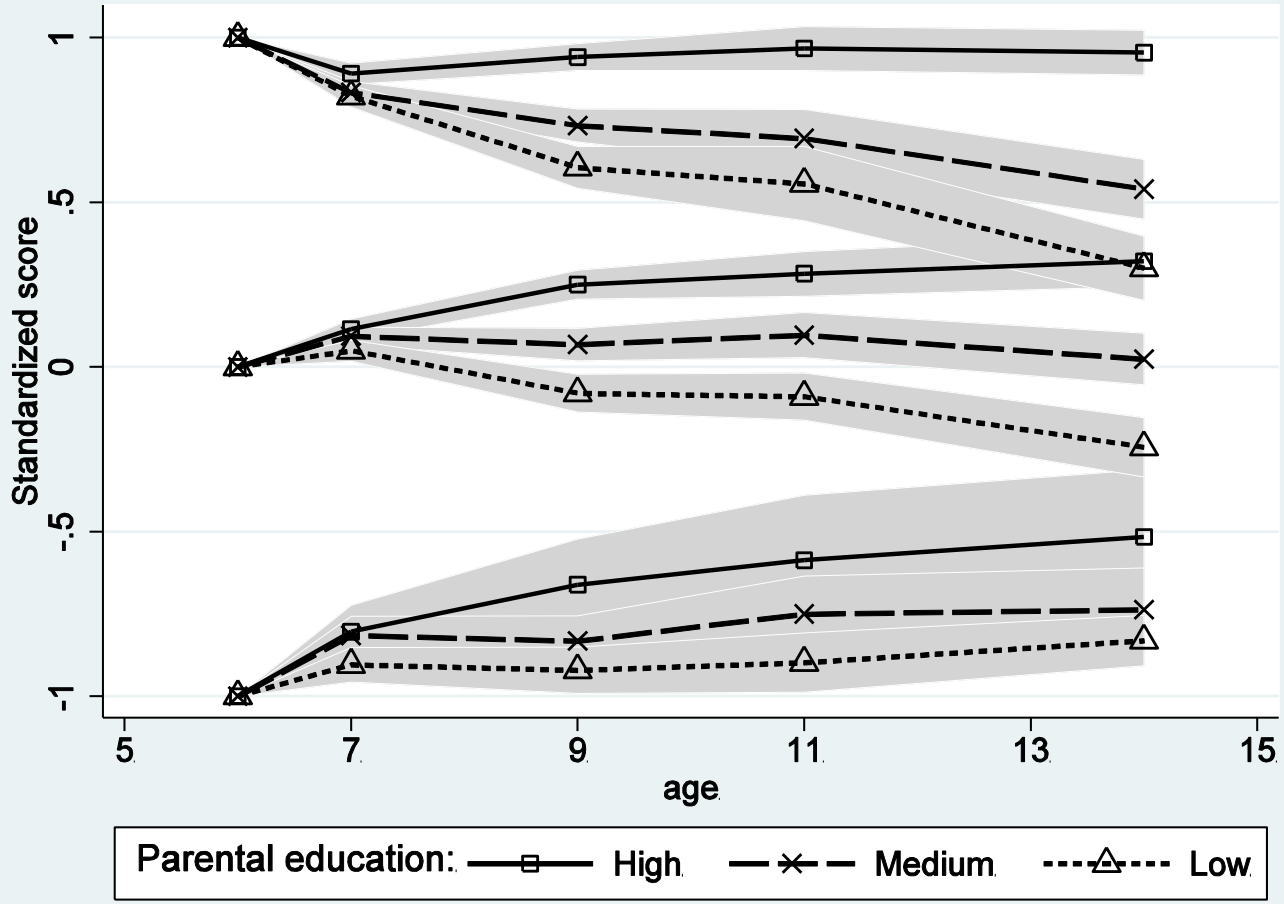
Modelling trajectories

- How much of the 8th grade SES gap reflects the fact low SES kids tended to start behind? How much that initially high-achieving low-SES kids differentially fall back? Or initially low-achieving low-SES kids differentially 'get stuck'?
- Modelling of trajectories raises additional methodological issues
  - Consequences of **measurement error** in initial scores (leads to overstatement of extent of poorer progress of low SES group vs their initial position)
  - Are instruments measuring **the 'same' latent construct at different ages**? (Is this even possible over long spans of childhood?) Differences in what is measured tend to have the same consequences as measurement error

Our approach

- Focus modelling of trajectories mainly on US data where (in theory) tests measure the evolution of the same skill set from 5 to 14
  - IRT modelling of 'theta' ability scores with overlapping test items at different ages
- Use an instrumental variable (IV) technique to correct for measurement error in 'initial' test score
  - Fall K test score used to correct Spring K test score (uniquely possible in US data because lag between first two measurements is only 6 months)
- A less restrictive method would be to use published **reliabilities** of initial tests (signal/noise ratio). Unfortunately this information can be very difficult to find!

Predicted trajectories in reading for high- (+1SD), medium- (0SD) and low- (-1SD) achieving children in Spring K, by SES



Shaded areas are 95% confidence intervals

We estimate that 60% of the 8th grade reading gap can be attributed to SES differences in Spring K

The 40% due to differential progress by SES reflects differences spread equally over the initial achievement distribution

Similar proportions appear to hold in the other countries, but comparability issues mean this is *very* tentative

Lessons learned

- Plea for longer panels
  - Two of the four studies we use have been discontinued, but new cohorts continue to start. Is the value added by additional waves under-appreciated?
- Plea for common instruments/greater cross-national communication
  - Gain in accuracy from improving psychometric properties with every new study balanced by loss of comparability
- Plea for accessible high-quality information on reliability of scales
- Attention to full spectrum of domains of child development (e.g. psychosocial as well as achievement outcomes)