Causal inference and education data in Australia

Dr Nicholas Biddle[1] and Professor Robert Breunig[2]

1 ANU Centre for Social Research and Methods, Australian National University

2 Crawford School of Public Policy, Australian National University

There is a belief amongst some in the education research and policy fields that we know what works to improve educational outcomes, if only governments would fund what is needed. In our view, that belief is wrong.

We have a good idea in Australia about what tends to work in general, and some idea on what would appear to be less effective. For example, we know from Australian and overseas experience that those who have been exposed to high quality early childhood education have better outcomes than those who haven't. We know that there is a non-linear relationship with class size, with very large classes being detrimental, but very smaller classes not having high returns.

But the devil is almost always in the detail. What has been shown to work in one country may not work within Australia, or for particular population groups in Australia. For example, Conditional Cash Transfers have been shown to work in a number of developing countries, but have been less effective in developed countries (see Miller, Riccio et al. (2015) for an evaluation of a large-scale trial in New York City). We don't know, therefore, whether they would work in Australia, or for Indigenous children (for example).

The reason why we don't know as much as we should within the field of education on what policies and programmes are effective and which are not, is that we haven't invested in the type of rigorous trial and error approach that other countries have started to, or which are routine in other domains. Nor have we made use of the extensive administrative datasets to anywhere near their full capacity.

Many empirical research questions related to education are concerned principally with correlations. We may be interested in the geographic distribution of students from a low-income background, under the assumption that these students require a greater level of educational support. There are other research and policy questions, however, that are explicitly concerned with causality. Did a particular policy or does a particular behaviour directly affect a particular outcome. For these questions, observational data is less useful.

We can try and answer such questions by comparing the outcomes *ex post* of those individuals who were in one group (for example those who participated in a program) with those who were in a different group (those who didn't participate). Often times, we can use longitudinal data to look at the change in outcomes across the different groups. These comparisons are useful, but we cannot be sure whether the differences in outcomes are driven by the program itself (a causal effect), whether the outcomes affected participation (reverse causality), or whether there is a third variable or set of variables that affect both (unobserved heterogeneity).

We can try and recover this causal inference using econometric techniques (controlling for observable characteristics), theory/logic, baseline data or some other information about the program. However, such arguments and techniques are always going to be open to criticism or counter-claim. Perhaps there is some crucial characteristic that cannot be controlled for. Perhaps there is an alternative theory that suggests a different causal pathway. Even with baseline or longitudinal data, something else may have happened that affects outcomes between the two observation periods. The fundamental evaluation problem remains – we can't observe the same individual in two different states. Without careful design, we don't have a counterfactual.

The only way to overcome the evaluation problem is to make sure the two groups being compared are exactly the same at the base period, based on both observable and unobservable characteristics. That is, we set up our comparison groups *ex ante* rather than *ex post*ˆ. Or, following Torgerson and Torgerson (2008), 'we assemble a population for whom intervention is appropriate …; we then allocate the participants to two or more groups and apply the intervention(s) to the groups formed by randomisation; at some pre-specified time in the future we measure the groups in terms of their outcomes – if there is differences between the groups, and assuming that the difference and the sample size are sufficient, we can infer a causal relationship between our intervention and the group differences. '

Within education policy internationally, the number of rigorous randomised trials has reached a significant enough level that Fryer Jr (2016) was able to publish a review article of the results and implications of randomised field experiments in the areas of early childhood education, home-based interventions and school education. The author defined a field experiment as 'any intervention that uses a *verifiably* random procedure to assign participants to treatment and control groups in a non-laboratory environment.' Fryer Jr and colleagues were able to generate usable results from 196 of these in developed countries. From this review, the author was able to conclude that:

> Early childhood investments, on average, significantly increase achievement. Yet, experiments that attempt to alter the home environment in which children are reared in have shown very little success at increasing student achievement. Among school experiments, high-dosage tutoring and 'managed' professional development for teachers have shown to be effective. Ironically, high-dosage tutoring of adolescents seems to be as effective – if not more effective – than early childhood investments. This argues against the growing view that there is a point at which investments in youth are unlikely to yield significant returns … Lastly, charter schools can be effective avenues of achievement-increasing reform, though the evidence on other market-based approaches such as vouchers or school choice have less demonstrated success.

Whilst highlighting the increase in trials (and evidence) in recent years, Fryer Jr (2016) was also able to conclude that 'In the 1960s we saw the Perry preschool experiment and the income maintenance experiments, in the 1970s the Abecedarian project was initiated, and in the 1980s there was Project STAR, the Tennessee class size experiment. The data from these randomized experiments alone were used for decades to investigate many interesting questions about how to best produce human capital.' However, he also made clear that the number of trials has increased exponentially (in his words), pointing out that 'In 2000, 14 percent of reviewed education publications on What Works Clearinghouse met their standards without reservations, a distinction given only to well-designed studies that have comparison groups deter- mined through a random process. By 2010, that number had tripled to over 46 percent.'

The same could not be said for trials that focus on Australian conditions and Australian populations. Not a single one of the trials reported in Fryer Jr (2016) has been conducted in Australia. When Harrison, Goldfeld et al. (2012) wrote an article for the Australian Institute of Health and Welfare's Closing the Gap Clearinghouse, one of their main conclusions was that 'There have been no rigorous trials or evaluations of early childhood programs in Australia, particularly programs for Indigenous and at-risk children.' While this was four years ago now, it is fair to say that this situation hasn't changed much in the intervening years.

This current enquiry into education data is right to focus on the broad array of data required for an evidence-based education policy in Australia. However, there needs to be recognition of the specific data needs for addressing causal questions. And there are many questions that, if we were honest with ourselves as a policy and research community, we don't have answers to.

In the field of Early Childhood Education (ECE):

- Does participation in ECE improve outcomes for children?
    - Does the number of hours make a difference?
    - Does the institution a child attends make a difference?
- Did the introduction of the National Quality Framework affect the benefits of ECE?
    - Is it teacher qualifications, teacher ratios, something else?
- Do boys/girls, Indigenous/non-Indigenous, advantaged/disadvantaged, etc., children have different benefits/costs of ECE?
- Does the structure of the family tax benefits/childcare subsidies affect participation in ECE?
- Would financial incentives, non-financial incentives, nudges, etc. increase participation rates in ECE?

With regards to school education:

- What affects the supply of highly skilled individuals (graduates and undergraduates) into the teaching profession?
- Does the provision of information (e.g. My School) increase the per cent of children who don't attend their local school?
    - Does this school choice improve/worsen the outcomes of those children who do/do not attend a different school
- Does the child's experience at school (racism, bullying, etc.) affect their outcomes?
- Does attendance at an academically selective school improve outcomes? Does it worsen outcomes for those who remain in the comprehensive school system?
- Does attendance at a boarding school increase the chance of completion or post-school attendance of Indigenous children from regional/remote areas?
- Does the hours of homework set for a child affect their outcomes/wellbeing?
- What are the effects of linking welfare to school attendance

And in post-school education:

- Does advice given to students affect the decision to undertake university or VET education?
    - Does it matter when that advice is given and who gives the advice?
- Does working whilst studying affect outcomes?
- Are there types of training that increase the chances that someone who is long-term unemployed will obtain employment?
- Does the type of instruction (face-to-face/online/flipped) affect acquisition of knowledge
- Does having lecturers with a focus on research affect student learning?
- Are there interventions that minimise the effects of gender, ethnicity, socioeconomic background, etc., on participation or completion of post-school education?

Within almost all of the above questions, there are specific policies and programs that need to be evaluated for their effectiveness. In late 2015, the ANU hosted a visit from Dr Benjamin L Castleman from the University of Virginia. Dr Castleman is an expert on behaviourial insights into educational access and attainment in the United States. In his lecture to researchers and policy makers, he shared his experiences and findings from conducting randomised controlled trials in education research. There were three main conditions for effective education policy design identified by Castleman:

1. A recognition that there are many educational questions that we don't have an answer for;
2. Building causal inference into the design of education policies, rather than as an afterthought; and
3. Availability of data from administrative records at the unit-record level.

The key point with regards to causal inference and education policy is that the availability of administrative data substantially reduces the cost of trials. The policies themselves may be costly or not. However, with relatively easy access to administrative data, learning from those programs in a way that supports better education policy has very few additional costs.

References

Fryer Jr, R. G. (2016). The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments, National Bureau of Economic Research.

Harrison, L., S. Goldfeld, E. Metcalfe and T. Moore (2012). Early learning programs that promote children's developmental and educational outcomes, AIHW.

Miller, C., J. Riccio, N. Verma, S. Nuñez, N. Dechausay and E. Yang (2015). "Testing a conditional cash transfer program in the US: the effects of the family rewards program in New York City." IZA Journal of Labor Policy **4**(1): 1-29.

Torgerson, D. J. and C. Torgerson (2008). Designing randomised trials in health, education and the social sciences: an introduction, Palgrave Macmillan.