Uninformative educational trials and their causes

Matthew Inglis, Loughborough University

Evaluating *educational innovations with rigorous, large-scale research must be a good idea – but why do we then find such small effects?*

The last decade has seen a major change in educational research funding in the UK. The advent of the Education Endowment Foundation (EEF) means that a majority of the money spent on education research is now used to conduct randomised controlled trials (RCTs) of educational interventions. Prior to the publication of the first EEF trials, such studies were almost unheard of: in 2012 only 3% of articles published in the eight major mathematics education journals reported RCTs. Has this change in focus been a success? With the aim of finding out, Hugo Lortie-Forgues and I recently conducted a review of all RCTs commissioned by the EEF and the NCEE, a US-based funder that also commissions educational RCTs.

What is an educational RCT? The basic structure is simple. An educational designer proposes some kind of intervention – perhaps a programme of one-on-one tuition, or a particular educational game – which they believe will raise student achievement. A large group of students is recruited to take part who are randomly allocated to either receive the intervention, or act as a control group and carry on with their normal activities. After the intervention is complete, both groups' educational achievement is assessed with some kind of outcome measure, perhaps a standardised test, and compared. If there is a difference between the groups, and if that difference could not plausibly be attributed to chance, then the researcher concludes that the intervention caused the difference.

In typical educational contexts things are slightly more complex. For one thing, children are usually taught in classes, so randomisation must take place at the class (or school) level rather than the individual level. Although this adds complexity, the use of appropriate statistical techniques permits causal conclusions to still be drawn.

RCTs are powerful. When we don't know whether or not a proposed intervention is effective (i.e. causes higher achievement), then a well conducted RCT with positive results can help us decide. However, results are not always positive, and not necessarily because the intervention is ineffective. To explain why, we need the concept of an effect size. This is essentially just a measure of the difference in outcome between the intervention and control groups. A positive effect size suggests that the intervention is effective (compared to whatever the control group was doing, usually 'business as usual'), an effect size of zero suggests that it is ineffective, and a negative effect size suggests that it is actively harmful. The effect size we obtain from an RCT, with its one particular group of participants, is merely an estimate of the 'true' effect size: the figure we would obtain if we ran the study on every member of the population of interest (an impossible task).

It is the true effect size that we care about, as it is this effect size which allows us to draw conclusions about future uses of the intervention. Let's restrict ourselves to the case where an intervention is either effective or ineffective (not actively harmful), i.e. where the true effect size is either positive or zero. We'd like to use our RCT to decide which. To do this we can make some assumptions about the range of plausible positive effects an RCT testing an effective intervention might find, and calculate a statistic known as a *Bayes Factor*. This quantifies which of our two hypotheses the RCT's results are more consistent with. Interestingly, sometimes RCTs are equally consistent with both hypotheses. Such an RCT does not allow us to conclude whether the intervention is effective or ineffective. RCTs of this sort are *uninformative*: before any RCT is run we didn't know whether the intervention is effective or ineffective, after we've seen the results of an uninformative RCT we still don't know.

Clearly, uninformative RCTs are highly undesirable. The EEF spends around £500k per RCT, so it is obviously problematic if they do not produce new information. But what proportion of educational RCTs are uninformative? To investigate, Hugo Lortie-Forgues-Forgues and I reanalysed 141 large-scale educational RCTs commissioned by the EEF and NCEE. In total 1.2m children took part in these studies.

There were two main findings. First, most educational RCTs find small effects: the average difference between the intervention and control groups was just 0.06 standard deviations. One way of understanding this figure is to ask what the probability is that a randomly picked member of the intervention group has a higher score than a randomly picked member of the control group. With an effect size of 0.06 the answer is 51.7%, barely above the 50% chance level.

Second, and most importantly, we found that 40% of trials were uninformative. In other words, between a third and half of all large-scale educational trials did not permit a conclusion to be drawn about whether the intervention they were testing was effective or ineffective. This is an alarmingly high number: at £500k per trial it suggests that the EEF and NCEE have spent around £28m conducting uninformative trials.

Why are so many trials uninformative? And what can be done about it?

In our paper we discuss three main hypotheses:

1. Perhaps the interventions which RCTs are testing are based on unreliable basic research.

2. We may not be effectively translating insights from reliable basic research into interventions that can be implemented at scale with fidelity.
3. RCTs themselves are typically designed to maximise their relevance to practitioners, but perhaps this comes at the cost of increasing the level of statistical noise in the design to too high a level.

Each of these accounts suggests a different change to practice: (i) and (iii) call for methodological reform, to basic research and RCT design respectively; (ii) calls for increased investment in educational design. Given the level of resource, both in terms of research funding and teacher/pupil time, that is currently being spent on educational RCTs, it is vital that we investigate why so many RCTs find small and uninformative results.

About the author

Matthew Inglis is Professor of Mathematical Cognition at Loughborough University. He completed undergraduate (BSc Mathematics) and postgraduate (MSc Mathematics Education, PhD Education) studies at the University of Warwick. Following a period as a Research Fellow at the Learning Sciences Research Institute in Nottingham, he took up a lectureship in Loughborough in 2008.

About *Researching Education* (ISSN 2752-3101)

'Researching Education' offers accessible summaries of recent research findings that concern educational practice. Articles are published online and written by the original research investigators. Subscription is open and free but most likely to appeal to practicing teachers. Subscribers receive periodic emails with topic headlines that link to the online article summaries. Subscribe here.

Reference for this article

Inglis, M (2020) Uninformative educational trials and their causes, Researching Education, 1(1). https://10.5281/zenodo.4279342