

BUILDING EVIDENCE INTO EDUCATION

BEN GOLDACRE

MARCH 2013



Background

Ben Goldacre is a doctor and academic who writes about problems in science and evidence based policy, with his Guardian column “Bad Science” for a decade, and the bestselling book of the same name. He is currently a Research Fellow in Epidemiology at London School of Hygiene and Tropical Medicine.

To find out more about randomised trials, and evidence based practice, you may like to read “Test, Learn, Adapt”, a Cabinet Office paper written by two civil servants and two academics, including Ben Goldacre:

<https://www.gov.uk/government/publications/test-learn-adapt-developing-public-policy-with-randomised-controlled-trials>

Table of contents

Background	3
Building evidence into education	7
How randomised trials work	8
Myths about randomised trials	10
Making evidence part of everyday life	15

Building evidence into education

I think there is a huge prize waiting to be claimed by teachers. By collecting better evidence about what works best, and establishing a culture where this evidence is used as a matter of routine, we can improve outcomes for children, and increase professional independence.

This is not an unusual idea. Medicine has leapt forward with evidence based practice, because it's only by conducting "randomised trials" - fair tests, comparing one treatment against another - that we've been able to find out what works best. Outcomes for patients have improved as a result, through thousands of tiny steps forward. But these gains haven't been won simply by doing a few individual trials, on a few single topics, in a few hospitals here and there. A change of culture was also required, with more education about evidence for medics, and whole new systems to run trials as a matter of routine, to identify questions that matter to practitioners, to gather evidence on what works best, and then, crucially, to get it read, understood, and put into practice.

I want to persuade you that this revolution could - and should - happen in education. There are many differences between medicine and teaching, but they also have a lot in common. Both involve craft and personal expertise, learnt over years of experience. Both work best when we learn from the experiences of others, and what worked best for them. Every child is different, of course, and every patient is different too; but we are all similar enough that research can help find out which interventions will work best overall, and which strategies should be tried first, second or third, to help everyone achieve the best outcome.

Before we get that far, though, there is a caveat: I'm a doctor. I know that outsiders often try to tell teachers what they should do, and I'm aware this often ends badly. Because of that, there are two things we should be clear on.

Firstly, evidence based practice isn't about telling teachers what to do: in fact, quite the opposite. This is about empowering teachers, and setting a profession free from governments, ministers and civil servants who are often overly keen on sending out edicts, insisting that their new idea is the best in town. Nobody in government would tell a doctor what to prescribe, but we all expect doctors to be able to make informed decisions about which treatment is best, using the best currently available evidence. I think teachers could one day be in the same position.

Secondly, doctors didn't invent evidence based medicine. In fact, quite the opposite is true: just a few decades ago, best medical practice was driven by things like eminence, charisma, and personal experience. We needed the help of statisticians, epidemiologists, information librarians, and experts in trial design to move forwards. Many doctors – especially the most senior ones - fought hard against this, regarding “evidence based medicine” as a challenge to their authority.

In retrospect, we've seen that these doctors were wrong. The opportunity to make informed decisions about what works best, using good quality evidence, represents a truer form of professional independence than any senior figure barking out their opinions. A coherent set of systems for evidence based practice listens to people on the front line, to find out where the uncertainties are, and decide which ideas are worth testing. Lastly, crucially, individual judgement isn't undermined by evidence: if anything, informed judgement is back in the foreground, and hugely improved.

This is the opportunity that I think teachers might want to take up. Because some of these ideas might be new to some readers, I'll describe the basics of a randomised trial, but after that, I'll describe the systems and structures that exist to support evidence based practice, which are in many ways more important. There is no need for a world where everyone is suddenly an expert on research, running trials in their classroom tomorrow: what matters is that most people understand the ideas, that we remove the barriers to “fair tests” of what works, and that evidence can be used to improve outcomes.

How randomised trials work

Where they are feasible, randomised trials are generally the most reliable tool we have for finding out which of two interventions works best. We simply take a group of children, or schools (or patients, or people); we split them into two groups at random; we give one intervention to one group, and the other intervention to the other group; then we measure how each group is doing, to see if one intervention achieved its supposed outcome any better.

This is how medicines are tested, and in most circumstances it would be regarded as dangerous for anyone to use a treatment today, without ensuring that it had been shown to work well in a randomised trial. Trials are not only

used in medicine, however, and it is common to find them being used in fields as diverse as web design, retail, government, and development work around the world.

For example, there was a longstanding debate about which of two competing models of “microfinance” schemes was best at getting people out of poverty in India, whilst ensuring that the money was paid back, so it could be re-used in other villages: a randomised trial compared the two models, and established which was best.

At the top of the page at Wikipedia, when they are having a funding drive, you can see the smiling face of Jimmy Wales, the founder, on a fundraising advert. He’s a fairly shy person, and didn’t want his face to be on these banners. But Wikipedia ran a randomised trial, assigning visitors to different adverts: some saw an advert with a child from the developing world (“she could have access to all of human knowledge if you donate...”); some saw an attractive young intern; some saw Jimmy Wales. The adverts with Wales got more clicks and more donations than the rest, so they were used universally.

It’s easy to imagine that there are ways around the inconvenience of randomly assigning people, or schools, to one intervention or another: surely, you might think, we could just look at the people who are already getting one intervention, or another, and simply monitor their outcomes to find out which is the best. But this approach suffers from a serious problem. If you don’t randomise, and just observe what’s happening in classrooms already, then the people getting different interventions might be very different from each other, in ways that are hard to measure.

For example, when you look across the country, children who are taught to read in one particularly strict and specific way at school may perform better on a reading test at age 7, but that doesn’t necessarily mean that the strict, specific reading method was responsible for their better performance. It may just be that schools with more affluent children, or fewer social problems, are more able to get away with using this (imaginary) strict reading method, and their pupils were always going to perform better on reading tests at age 7.

This is also a problem when you are rolling out a new policy, and hoping to find out whether it works better than what’s already in place. It is tempting to look at results before and after a new intervention is rolled out, but this can be very misleading, as other factors may have changed at the same time. For example, if you have a “back to work” scheme that is supposed to get people on

benefits back into employment, it might get implemented across the country at a time when the economy is picking up anyway, so more people will be finding jobs, and you might be misled into believing that it was your “back to work” scheme that did the job (at best, you’ll be tangled up in some very complex and arbitrary mathematical modelling, trying to discount for the effects of the economy picking up).

Sometimes people hope that running a pilot is a way around this, but this is also a mistake. Pilots are very informative about the practicalities of whether your new intervention can be implemented, but they can be very misleading on the benefits or harms, because the centres that participate in pilots are often different to the centres that don’t. For example, job centres participating in a “back to work” pilot might be less busy, or have more highly motivated staff: their clients were always going to do better, so a pilot in those centres will make the new jobs scheme look better than it really is. Similarly, running a pilot of a fashionable new educational intervention in schools that are already performing well might make the new idea look fantastic, when in reality, the good results have nothing to do with the new intervention.

This is why randomised trials are the best way to find out how well a new intervention works: they ensure that the pupils or schools getting a new intervention are the same as the pupils and schools still getting the old one, because they are all randomly selected from the same pool.

At around this point, most people start to become nervous: surely it’s wrong, for example, to decide what kind of education a child gets, simply at random? This cuts to the core of why we do trials, and why we gather evidence on what works best.

Myths about randomised trials

While there are some situations where trials aren’t appropriate - and where we need to be cautious in interpreting the results - there are also several myths about trials. These myths are sometimes used to prevent trials being done, which slows down progress, and creates harm, by preventing us from finding out what works best. Some people even claim that trials are undesirable, and even completely impossible, in schools: this is a peculiarly local idea, and there have been huge numbers of trials in education in other countries, such as the US. However, the specific myths are worth discussing.

Firstly, people sometimes worry that it is unethical to randomly assign children to one educational intervention or another. Often this is driven by an implicit belief that a new or expensive intervention is always necessarily better. When people believe this, they also worry that it's wrong to deprive people of the new intervention. It's important to be clear, before we get to the detail, that a trial doesn't necessarily involve depriving people of anything, since we can often run a trial where people are randomly assigned to receive the new intervention now, or after a six month wait. But there is a more important reason why trials are ethically acceptable: in reality, before we do a trial, we generally have no idea which of two interventions is best. Furthermore, new things that many people believe in can sometimes turn out, in reality, to be very harmful.

Medicine is littered with examples of this, and it is a frightening reality. For many years, it was common to treat everyone who had a serious head injury with steroids. This made perfect sense on paper: head injuries cause the brain to swell up, which can cause important structures to be crushed inside our rigid skulls; but steroids reduce swelling (this is why you have steroid injections for a swollen knee), so they should improve survival. Nobody ran a trial on this for many years. In fact, it was widely argued that randomising unconscious patients in A&E to have steroids or not would be unethical and unfair, so trials were actively blocked. When a trial was finally conducted, it turned out that steroids actually *increased* the chances of dying, after a head injury. The new intervention, that made perfect sense on paper, that everyone believed in, was killing people: not in large enough numbers to be immediately obvious, but when the trial was finally done, an extra two people died out of every hundred people given steroids.

There are similar cases from the world of education. The "Scared Straight" programme also made sense on paper: young children were taken into prisons and shown the consequences of a life of crime, in the hope that they would be more law abiding in their own lives. Following the children who participated in this programme into adult life, it seemed they were less likely to commit crimes, when compared with other children. But here, researchers were caught out by the same problem discussed above: the schools - and so the children - who went on the Scared Straight course were different to the children who didn't. When a randomised trial was finally done, where this error could be accounted for, we found out that the Scared Straight programme - rolled out at great expense, with great enthusiasm, good intentions, and huge optimism - was actively harmful, making children more likely to go to prison in later life.

So we must always be cautious about assuming that things which are new, or

expensive, are necessarily always better. But this is just one special case of a broader issue: we should always be clear when we are uncertain about which intervention is best. Right now, there are huge numbers of different interventions used throughout the country - different strategies to reduce absenteeism, or teach arithmetic, or reduce teenage pregnancies, or any number of other things - where there is no evidence to say which of the currently used methods is best. There is arbitrary variation, across the country, across a town, in what strategies and methods are used, and nobody worries that there is an ethical problem with this.

Randomisation, in a trial, adds one simple extra think to this existing variation: we need a group of schools, teachers, pupils, or parents, who are able to honestly say: “we don’t know which of these two strategies is best, so we don’t mind which we use. We want to find out which is best, and we know it won’t harm us.”

This is a good example of how gathering good evidence requires a culture shift, extending beyond a few individual randomised trials. It requires everyone involved in education to recognise when it’s time to honestly say “we don’t know what’s best here”. This isn’t a counsel of despair: in medicine, and in teaching, we know that most of what we do does *some* good (if we’re not better than nothing, then we’re all in big trouble!). The real challenge is in identifying what works the *best*, because when people are deprived of the best, they are harmed too. But this is also a reminder of how inappropriate certainty can be a barrier to progress, especially when there are charismatic people, who claim they know what’s best, even without good evidence.

Medicine suffered hugely with this problem, and as late as the 1970s there were infamous confrontations between people who thought it was important to run fair tests, and “experts”, who were angry at the thought of their expertise being challenged, and their favourite practices being tested. Archie Cochrane was one of the pioneers of evidence based medicine, and in his autobiography, he describes many battles he had with senior doctors, in glorious detail. In 1971, Cochrane was concerned that Coronary Care Units in hospitals might be no better than home care, which was the standard care for a heart attack at the time (we should remember that this was the early days of managing heart attacks, and the results from this study wouldn’t be applicable today). In fact, he was worried that hospital care might involve a lot of risky procedures that could even, conceivably, make outcomes worse for patients overall.

Because of this, Cochrane tried to set up a randomised trial comparing home

care against hospital care, against great resistance from the cardiologists. In fact, the doctors running the new specialist units were so vicious about the very notion of running a trial that when one was finally set up, and the first results were collected, Cochrane decided to play a practical joke. These initial results showed that patients in Coronary Care Units did worse than patients sent home; but Cochrane switched the numbers around, to make it look like patients on CCUs did better. He showed the cardiologists these results, which reinforced their belief that it was wrong of Cochrane to even dare to try running a randomised trial of whether their specialist units were helpful. The room erupted:

“They were vociferous in their abuse: “Archie,” they said “we always thought you were unethical. You must stop this trial at once.” ... I let them have their say for some time, then apologized and gave them the true results, challenging them to say as vehemently, that coronary care units should be stopped immediately. There was dead silence and I felt rather sick because they were, after all, my medical colleagues.

Similar confrontations are reported in many new fields, when people try subjecting ideas and practices to fair tests, in randomised trials. But being open and clear about the need for research - when there is no good evidence to help us choose between interventions - is also important because it helps make sure that research is done on relevant questions, meeting the needs of teachers, pupils and parents. When everyone involved in teaching knows a little about how research is done - and what previous research has found - then we can all have a better idea of what questions need to be asked next.

But before we get on to how this can happen, we should first finish the myths about trials. From now on, these are all cases where people *overstate* the benefits of trials.

For example, sometimes people think that trials can answer everything, or that they are the *only* form of evidence. This isn't true, and different methods are useful for answering different questions. Randomised trials are very good at showing *that* something works; they're not always so helpful for understanding *why* it worked (although there are often clues when we can see that an intervention worked well in children with certain characteristics, but not so well in others). “Qualitative” research - such as asking people open questions about their experiences - can help give a better understanding of how and why things worked, or failed, on the ground. This kind of research can also be useful for generating new questions about what works best, to be answered with trials. But

qualitative research is very bad for finding out whether an intervention has worked. Sometimes researchers who lack the skills needed to conduct or even understand trials can feel threatened, and campaign hard against them, much like the experts in Archie Cochrane's story. I think this is a mistake. The trick is to ensure that the right method is used to answer the right questions.

A related issue involves choosing the right outcome to measure. Sometimes people say that trials are impossible, because we can't capture the intangible benefits that come from education, like making someone a well rounded member of society. It's true that this outcome can be hard to measure, although that is an argument against any kind of measurement of attainment, and against any kind of quantitative research, not just trials. It's also, I think, a little far-fetched: there are lots of things we try to improve that are easy to measure, like attendance rates, teenage pregnancy, amount of exercise, performance on specific academic or performance tests, and so on.

However, we should return to the overly exaggerated claims sometimes made in favour of trials, and the need to be a critical consumer of evidence. A further common mistake is to assume that, once an intervention has been shown to be effective in a single trial, then it definitely works, and we should use it everywhere. Again, this isn't necessarily true. Firstly, all trials need to be run properly: if there are flaws in a trial's design, then it stops being a fair test of the treatments. But more importantly, we need to think carefully about whether the people in a trial of an intervention are the same as the people we are thinking of using the intervention on.

The Family Nurse Partnership is a programme that is well funded and popular around the world. It was first shown to be effective in a randomised trial in 1977. The trial participants were white mothers in a semirural setting upstate from New York, and people worried at the time that the positive results might have been exceptional, and occurred simply because the specific programme of social support that was offered had suited this population unusually well. In 1988, to check that the findings really were applicable to other settings, the same programme was assessed using a randomised trial in African-American mothers in inner city Memphis, and again found to be effective. In 1994, a third trial was conducted in a large population of Hispanic, African-American, and Caucasian mothers from Denver. After this trial also showed a benefit, people in the US were fairly certain that the programme worked, with fewer childhood injuries, increased maternal employment, improved "school readiness", and more.

Now, the Family Nurse Partnership programme is being brought to the UK,

but the people who originally designed the intervention have insisted that a randomised trial should be run here, to see if it really is effective in the very different setting of the UK. They have specifically stated that they expect to see less dramatic benefits here, because the basic level of support for young families in the UK is much better than that in the US: this means that the difference between people getting the FNP programme, and people getting the normal level of help from society, will be much smaller.

This is just one example of why we need to be thoughtful about whether the results of a trial in one population really are applicable to our own patients or pupils. It's also an illustration of why we need to make trials part of the everyday routine, so that we can replicate trials, in different settings, instead of blindly assuming we can use results from other countries (or even other schools, if they have radically different populations). It doesn't mean, however, that we can never trust the results of a trial. This is just another example of why it's useful to know more about how trials work, and to be a thoughtful consumer of evidence.

Lastly, people sometimes worry that trials are expensive and complicated. This isn't necessarily true, and it's important to be clear what the costs of a trial are being compared against. For example, if the choice is between running a trial, and simply charging ahead, implementing an idea that hasn't been shown to work - one that might be ineffective, wasteful, or even harmful - then it's clearly worth investing some time and effort in assessing its true impact. If the alternative is doing an "observational" study, which has all the shortcomings described above, then the analysis can be so expensive and complex - not to mention unreliable - that it would have been easier to randomise participants to one intervention or the other in the first place.

But the mechanics and administrative processes for running a trial can also be kept to a minimum with thoughtful design, for example by measuring outcomes using routine classroom data, that was being collected anyway, rather than running a special set of tests. More than anything, though, for trials to be run efficiently, they need to be part of the culture of teaching.

Making evidence part of everyday life

I'm struck by how much enthusiasm there is for trials and evidence based practice in some parts of teaching: but I'm also struck that much of this enthusiasm dies out before it gets to do good, because the basic structures needed to support evidence based practice are lacking. As a result, a small

number of trials are done, but these exist as isolated islands, without enough bridges joining the people and strands of work together. This is nobody's fault: creating an "information architecture" out of thin air is a big job, and it might take decades. The benefits, though, are potentially huge. Some individual randomised trials from the UK have produced informative results, for example, but these results are then poorly communicated, so they don't inform and change practice as well as they might.

Because of this, I've sketched out the basics of what education would need, as a sector, to embrace evidence based practice in a serious way. The aim - which I hope everyone would share - is to get more research done, involving as many teachers as possible; and to get the results of good quality research disseminated and put into practice. It's worth being clear, though, that this is a first sketch, and a call to arms. I hope that others will pull it apart and add to it. But I also hope that people will be able to act on it, because structures like these in medicine help capture the best value from the good work - and hard work - that is done all around the country.

Firstly - and most simply - it's clear that we need better systems for disseminating the findings of research to teachers on the ground. While individual studies are written up in very technical documents, in obscure academic journals, these are rarely read by teachers. And rightly so: most doctors rarely bother to read technical academic journals either. The British Medical Journal has brief summaries of important new research from around the world; and there is a thriving market of people offering accessible summary information on new "what works" research to doctors, nurses, and other healthcare professionals. The US government has spent vast sums of money on two similar websites for teachers: "Doing What Works", and the "What Works Clearing House". These are large, with good quality resources, and they are written to be relevant to teachers' needs, rather than dry academic games. While there are some similar resources in the UK, these are often short-lived, and on a smaller scale.

For these kinds of resources to be useful at all, they then need to land with teachers who know the basics of "how we know" what works. While much teacher training has reflected the *results* of research, this evidence has often been presented as a completed canon of answers. It's much rarer to find all young teachers being taught the basics of how different types of research are done, and the strengths and weaknesses of each approach on different types of question (although some individual teachers have taught themselves on this topic, to a very high level). Learning the basics of how research works is important, not

because every teacher should be a researcher, but because it allows teachers to be critical consumers of the new research findings that will come out during the many decades of their career. It also means that some of the barriers to research, that arise from myths and misunderstandings, can be overcome. In an ideal world, teachers would be taught this in basic teacher training, and it would be reinforced in Continuing Professional Development, alongside summaries of research.

In some parts of the world, it is impossible to rise up the career ladder of teaching without understanding how research can improve practice, and publishing articles in teaching journals. Teachers in Shanghai and Singapore participate in regular “Journal Clubs”, where they discuss a new piece of research, and its strengths and weaknesses, before considering whether they would apply its findings in their own practice. If the answer is no, they share the shortcomings in the study design that they’ve identified, and then describe any better research that they think should be done, on the same question.

This is an important quirk: understanding how research is done also enables teachers to generate new research questions. This, in turn, ensures that the research which gets done addresses the needs of everyday teachers. In medicine, any doctor can feed up a research suggestion to NIHR (the National Institute for Health Research), and there are organisations that maintain lists of what we don’t yet know, fed by clinicians who’ve had to make decisions, without good quality evidence to guide them. But there are also less tangible ways that this feedback can take place.

Familiarity with the basics of how research works also helps teachers get involved in research, and to see through the dangerous myths about trials being actively undesirable, or even “impossible” in education. Here, there is a striking difference with medicine. Many teachers pour their heart and soul into research projects which are supposed to find out whether something worked; but in reality the projects often turn out to be too small, being run by one person in isolation, in only one classroom, and lack the expert support necessary to ensure a robust design. Very few doctors would try and run a quantitative research project alone in their own single practice, without expert support from a statistician, and without help from someone experienced in research design.

In fact, most doctors participate in research by playing a small role in a larger research project which is coordinated, for example, through a research network. Many GPs are happy to help out on a research: they recruit participants from among their patients; they deliver whichever of two commonly used treatments

has been randomly assigned to their patient; and they share medical information for follow-up data. But they get involved by putting their name down with the Primary Care Research Network covering their area. Researchers interested in running a randomised trial in GP patients then go to the Research Network, and find GPs to work with.

This system represents a kind of “dating service” for practitioners and researchers. Creating similar networks in education would help join up the enthusiasm that many teachers have - for research that improves practice - with researchers, who can sometimes struggle to find schools willing to participate in good quality research. This kind of two-way exchange between researchers and teachers also helps the teacher-researchers of the future to learn more about the nuts and bolts of running a trial; and it helps to keep researchers out of their ivory towers, focusing more on what matters most to teachers.

In the background, for academics, there is much more to be said on details. We need, I think, academic funders who listen to teachers, and focus on commissioning research that helps us learn what works best, to improve outcomes. We need academics with quantitative research skills from outside traditional academic education departments - economists, demographers, and more - to come in and share their skills more often, in a multidisciplinary fashion. We need more expert collaboration with Clinical Trials Units, to ensure that common pitfalls in randomised trial design are avoided; we may also need - eventually - Education Trials Units, helping to support good quality research throughout the country.

But just as this issue stretches way beyond a few individual research projects, it also goes way beyond anything that one single player can achieve. We are describing the creation of a whole ecosystem from nothing. Whether or not it happens depends on individual teachers, researchers, heads, politicians, pupils, parents and more. It will take mischievous leaders, unafraid to question orthodoxies by producing good quality evidence; and it will need to land with a community that - at the very least - doesn't misunderstand evidence based practice, or reject randomised trials out of hand.

If this all sounds like a lot of work, then it should do: it will take a long time. But the gains are huge, and not just in terms of better evidence, and better outcomes for pupils. Right now, there is a wave of enthusiasm for good quality evidence, passing through all corners of government at the moment. This is the time to act. Teachers have the opportunity, I believe, to become an evidence based profession, in just one generation: embedding research into everyday

practice; making informed decisions independently; and fighting off the odd spectacle of governments telling teachers how to teach, because teachers can use the good quality evidence that they have helped to create, to make their own informed judgements.

There is also a roadmap. While evidence based medicine seems like an obvious idea today - and we would be horrified to hear of doctors using treatments without gathering and using evidence on which works best - in reality these battles were only won in very recent decades. Many eminent doctors fought viciously, as recently as the 1970s, against the very idea of evidence based medicine, seeing it as a challenge to their expertise. The case for change was made by optimistic young practitioners like Archie Cochrane, who saw that good evidence on what works best was worth fighting for.

Now we recognise that being a good doctor, or teacher, or manager, isn't about robotically following the numerical output of randomised trials; nor is it about ignoring the evidence, and following your hunches and personal experiences instead. We do best, by using the right combination of skills to get the best job done.