

## Using the tools of science to improve social policy

When drugs are launched, we expect rigorous testing, yet with government strategy we rely on anecdote or public mood when empirical study could offer better results



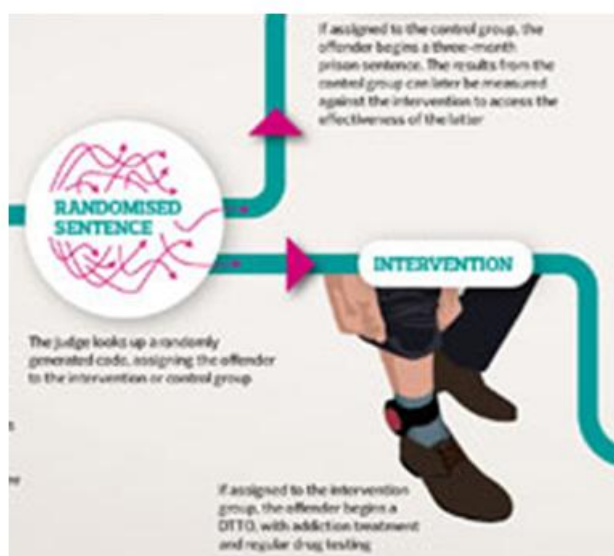
Education secretary Michael Gove: missed opportunity. Photograph: Eddie Mulholland/Rex Features

"All life is an experiment," wrote Ralph Waldo Emerson. "The more experiments you make the better." It's a maxim that is the stuff of science, the foundation stone of an approach to discovery that delivers reliable, if provisional, knowledge with incredible consistency.

Scientists observe the world, they develop ideas that may explain what they see and then, critically, they put them to the test in as dispassionate a fashion as possible. As the results of these experiments come in, we can start to separate good ideas from bad, and discard even beautiful hypotheses that fail to survive contact with the evidence. We can discover whether a medicine works, whether GM crops help or harm the environment, and whether the Higgs boson really exists.

The power of this experimental approach to knowledge has furnished us with understanding and technology that have shaped the modern world. It is also increasingly recognised by business, where successful companies like Google deliberately allow their staff the latitude to innovate and fail, so that they can learn from their mistakes.

Yet in another area of public life, experimental thinking is largely missing in action. If governments want to learn how best to teach our children, to cut crime or to rehabilitate offenders, they could use the rigorous methods of science to find out. Far too few of their policies, however, are examined by experiment before they are introduced.



Does prison work? [Click here to see the full flow chart](#). Graphic: Pete Guest

We rightly expect new drugs to be properly assessed by randomised controlled trials (RCTs) before they are taken to market, so we can be reasonably sure that they are effective and that they don't do more harm than good. For policy interventions that have just as much impact on people's lives, we are happy to accept much lower standards of evidence. Pilot projects are designed badly, if they are bothered with at all. Ideology, anecdote and the imagined public mood trump data time and again.

Neither, when a drug is licensed, is the experiment considered over. As tens or hundreds of thousands of patients start to take it, their experience is monitored consistently, and those that raise concerns, such as the painkiller Vioxx, are ultimately withdrawn. Government policies, however, go unrecognised as the mass experiments that they are.

Teaching techniques or sentencing guidelines are rolled out, unencumbered by genuine attempts to evaluate their success. If they're ever stopped, it's usually because of a popular backlash or an election. When was the last time you heard a minister say: "We've decided to scrap this because it just didn't work"?

Policy experiments, of course, involve people, and we can't set up a school or a prison in a lab and vary the conditions at will. But that doesn't mean it's impossible to design appropriate trials that can shed real light on what works and what fails, as the examples that follow show.

The alternative to rigorous, well-designed experiments in social policy isn't no experiments at all, it's experiments we run without bothering to collect any useful data. It isn't unethical or irresponsible to experiment with education or criminal justice. It's unethical not to.

## **THE SCHOOL DAY**

The body clocks of teenagers run hours behind those of adults and young children.

Photograph: Jamie Grill/Getty

### **The hypothesis**

The traditional school day starts between 8am and 9am, and many teachers believe that pupils do their



best work early in the morning. But research led by Russell Foster, a professor of circadian neuroscience at the University of Oxford, has suggested that this may not actually be the case for teenagers.

He has found that the body clocks of teenagers run several hours behind those of adults and younger children, perhaps explaining their propensity for late nights and lie-ins. This raised a tantalising possibility: could it be that starting the secondary school a little later might actually improve learning, by allowing pupils to study at a time of day when they are naturally more alert?

**The experiment** Foster's idea was ridiculed by the teaching unions, but Paul Kelley, then headteacher of Monkseaton high school in Tyneside, thought it worth investigating. In 2010, he persuaded his governors to allow him to push back the start of the school day from 9am to 10am. An experiment was under way.

In August 2011, after the first full school year using the new timetable, Monkseaton's year 11 pupils recorded the best GCSE results in the school's history. The proportion of pupils achieving at least five GCSEs at grades A\* to C rose by 19% on the previous year. Results were especially impressive in science and information and communications technology. Persistent absenteeism has also fallen by 27%. As things stand, this experiment proves little – as Kelley and Foster are the first to admit. It shows what's happened at a single school, over a single year – perhaps Monkseaton's year 11 was particularly bright, or perhaps the novelty of the new timetable, rather than the timetable itself, accounted for any benefits, which might thus fade over time. What it does reveal, however, is prima facie evidence that is worth following up properly. It would be relatively simple to run an RCT that would provide us with sound evidence. All the secondary schools in a particular region would be randomly assigned to start the school day at 9am or 10am. The exam results would then be tracked to see whether one group achieved statistically significant improvements in excess of the other.

## ACADEMIES

Without trials it is difficult to benchmark the performance of new academies against other schools. Photograph: Jim Wileman

### **The hypothesis**

Schools that are given academy status are made independent of the local authority, and have the opportunity to raise further funds from an individual or corporate sponsor. Academies can



vary admissions policies and the curriculum. Many academies have recorded better exam results than their predecessor schools, but there is controversy over whether they sustainably raise standards. Michael Gove, the education secretary, is convinced of their

value, and last year announced a plan to turn the 200 weakest primary schools into academies.

**The experiment** As with previous academy initiatives, Gove's policy hasn't been designed as an experiment that could be rigorously evaluated. The 200 weakest schools might well improve after the change, but as there is no way of benchmarking them against similar schools, it will be difficult to determine whether any differences result from the policy or some other factor.

It could be that standards would have risen anyway – the statistical phenomenon of regression to the mean makes it likely that underperforming schools will improve by chance alone. It could be that extra money, or the impetus of new governors, has an impact unrelated to structure. Without a good experimental design, it's impossible to know.

This is a particular shame because Gove's policy could easily have been introduced in a different way that would have given us some real answers. Indeed, the large number of schools he wants to change, and the clear selection criteria, would have been ideal for a proper experiment.

Carole Torgerson, a professor of education at Durham University, suggests that it could work like this: the 200 worst performing primary schools would have been identified in the same way as is happening now, but they wouldn't all have been transformed into academies at once. Rather, the schools would be assigned at random to receive academy status either immediately, or a year or two later.

This staggered RCT would have created a well-matched control group, against which the schools that became academies immediately could have been compared. It would therefore become possible to chalk up any improvements to the policy. And if the results looked good, all the schools would go on to receive a proven intervention in a timely fashion.

## DRUGS SENTENCING

It is well established many crimes are funding drug addiction.

Photograph:  
Mark Fagelson/Alamy

### **The hypothesis**

It is well established that many people convicted of crimes such as burglary are funding drug addiction. Treating such offenders, rather than incarcerating them, may therefore reduce recidivism.



Attracted by this, the Labour government introduced a new sentence in 1998, the drug treatment and testing order (DTTO). When a qualifying offender was convicted, he would take part in a mandatory treatment programme, with regular drug testing. A pilot project was deemed a success, and the policy was rolled out nationwide.

**The experiment** It was commendable that the Home Office decided to launch a pilot study of DTTOs before introducing them more widely. But Sheila Bird, a professor at the MRC biostatistics unit in Cambridge, showed that the pilots were so badly designed as to be virtually worthless. First, they included too few young offenders to achieve statistical significance. Second, the research wasn't randomised.

Random allocation of research subject to intervention and control groups is one of the most powerful tools for conducting trials of human subjects. It leaves minimal room for bias, and without it there always remains a possibility that any differences observed between subjects and controls may be the result of underlying differences between the two groups, rather than a true effect.

It would have been a simple matter to randomise the DTTO pilot. When a qualifying offender was convicted, the judge would pass the sentence that he or she felt appropriate. But before that sentence was actually carried out, the judge would use a random code to assign the offender either to the normal sentence or to a DTTO.

Both DTTO and control groups would then be followed up for differences in recidivism rates after their sentences were over. All that would have differed between the two groups was the sentence, which would therefore explain any different patterns of reoffending.

In the real pilots, the judges were left to decide who was to receive DTTOs, creating great potential for bias: they could easily have been tempted to cherry-pick more serious offenders for one arm of the trial or the other, according to their prejudices. No pharmaceutical company would have got away with running a trial this shoddy. Yet it was sufficient to change a criminal justice policy.

## FOREIGN AID

A charity study showed treating children medically improved performance more than supplying books. Photograph: Jerome Delay/AP

**The hypothesis** In the 1990s, a Dutch development charity called International Christelijk Steunfonds decided



to fund a programme to support education in Kenya. Previous research had suggested that providing African children with textbooks that they could not normally afford might improve their exam results, so the charity paid for 25 schools to receive sets of English, science and maths books. The charity, however, didn't just provide the books. It decided to run an experiment.

**The experiment** As Tim Harford describes in his book *Adapt*, ICS asked the Kenyan government not to select 25 schools that would receive the books, but to identify 100 schools that would be equally suitable. From these, 25 were selected at random. The books were delivered and exam results at the 25 intervention schools compared with those from the 75 similar schools without the extra teaching resources.

The textbooks, it turned out, made very little difference. ICS then tried another intervention – illustrated teaching flip-charts – in a similar randomised trial. Again, there was no significant effect.

So the charity tried a third approach, funding treatment for intestinal worms. This time, the trial followed a staggered design: 25 random schools received the treatment immediately, 25 after two years, and another 25 two years after that. This time, there was clear evidence: de-worming children unequivocally improved their learning, probably thanks to improved nutrition.

ICS had used the power of randomisation to identify how its limited resources could be spent most effectively. Few governments, alas, are as far-sighted.