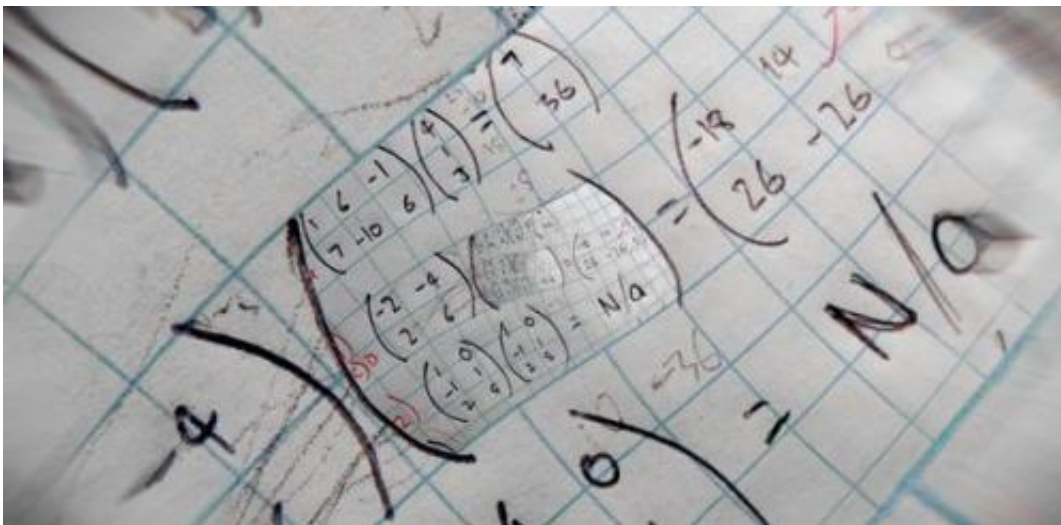


Getting schooled in the 'noise': learning about learning using big data

October 1 2013



Brits notoriously love bureaucracy, so it's perhaps unsurprising that the UK is a world leader in administrative data. With the digital era heralding a data revolution unlike anything in human history, education researchers such as Anna Vignoles are in a unique position to take advantage of this country's data deluge.

According to the January 2013 School Census, 8.2 million boys and girls trudged back to 24,328 schools across England after the Christmas holidays. Some are immigrants, some royalty. Some are Catholic, some Sikh, some Jewish. Some go to private schools. Some need free school

meals.

In the classroom, they are taught English, maths and science – as well as everything from the Reformation to colouring in. Their attainment levels are regularly predicted and monitored with varying degrees of accuracy – some will get straight As; some will get excluded. Most will fall somewhere in between.

And all of this information will be recorded, adding to the increasingly vast statistical reservoir of data as half a million kids complete the school system every year. Technology has allowed us to build information sets of unprecedented size and complexity, and this ocean of administrative info is the emerging coalface of education research in the era of 'big data'.

Professor Anna Vignoles, who joined Cambridge's Faculty of Education at the start of this year, is an economist specialising in the kind of quantitative research methods that have become possible in the wake of the big data explosion.

Buried in the 'noise' of these enormous administrative datasets lie important correlations awaiting detection, says Vignoles. When combined with qualitative methods, these data can provide [education researchers](#) with the kind of robust evidence needed to inform and evaluate education policy.

"The Department for Education first opened up their administrative data to the research community about 10 years ago, with fantastic results. A number of studies have used these data to look at school effectiveness or impact of resources," said Vignoles.

"It's partly through analysis of these data that we found the massive socio-economic gap at the point of entry into the [school](#) system actually

worsens through primary. A lot of universities now aim outreach activities at primary schools as a result.

"Combined with further education data, you have a timeline from the beginning of primary education right through to graduation, with the potential to go further by looking at figures like early career salaries," said Vignoles. "Just the admin data alone has transformed what we're able to ask in terms of understanding pupil achievement."

It's not just connections within certain datasets. By cross-referencing different data streams, researchers will be able to uncover connections and inferences between income, health, education, age, ethnicity and biomedical information.

One of the next projects for Vignoles, for which – along with her Faculty of Education colleague Professor Diane Reay – she is a co-investigator, will be the ambitious 'Life Study'. The project will start gathering data during pregnancy on everything from blood type to economic situation from a cohort of just under 100,000 mothers and their [children](#). It will amount to the largest study of its kind ever done and, when combined with existing datasets, will provide researchers with an unparalleled overview of the lives of British citizens.

"It's a massive undertaking," said Vignoles. "Clinics need to be built to provide space for interactions and testing, which is under way, and some field work will start next year.

"UK cohort surveys are probably the best in the world. We have data on children born in 1946, 1958, 1970, the millennium – and now going into 2015. There's no other country where you can have lifetime data on children born a generation apart going back to the forties."

The NHS is the spider at the centre of this data web, with its "highly

administered, centralised system" that produced the original longitudinal cohort datasets. "When the US attempted a very large-scale study of this kind it proved highly problematic because it's a very fragmented, private system," said Vignoles, "whereas the UK has centralised, manageable systems that provide rigorous administrative data which, when linked with survey data, is just incredibly powerful."

For Vignoles, Cambridge has got an important head start on other institutions when it comes to big data research in the social sciences: hardware. "You need to look at multiple cohorts when it comes to this kind of social science research, and the numbers get very big, very quickly. Cambridge's investment in the physical sciences means that we have the kind of huge computing capacity that you need – such as the Darwin supercomputer run by the High Performance Computing Service."

But hardware will only get us so far, says Vignoles. Social sciences in particular need to work hard at building the capacity in people – the skills and a "lack of fear" of the numbers, as many of these disciplines have traditionally focused on qualitative methods, and the big data explosion will require a "shift in emphasis".

"It takes years to accumulate the knowledge to analyse even one dataset, and those skills are becoming hugely important, whether you're a historian, economist, geneticist or sociologist. We need more people with the skills for this kind of work, and one of the big pushes here at Cambridge is to build quantitative capacity in terms of people."

Vignoles highlights the need for balance between both qualitative and quantitative research. "We have to be sensible in our recommendations of what should be measured, but also recognise why we're measuring it in the first place. There's an awful lot of evidence that children do better when they have assessment and feedback on what they do, and at the

system level you can't rely on everyone doing their thing and hope you will get a system that's efficient.

"Some degree of monitoring is part and parcel of any system. These big data sets and the number crunching involved offer efficient and effective ways to develop and improve the UK [education](#) system. Some of it's about accountability, but also understanding – you can't tell what's happening to a particular group without looking at it. We wouldn't have known about the success story of many non-white children in the system, for example, without the data analysis.

"I'm convinced that the projects we do going forward will not be discipline based. We will increasingly have to approach research questions with people from different backgrounds – that means multidisciplinary, it means multi-method, and I think the work will be much better for it. Cambridge, being so strong across a range of disciplines, can – and should – lead the world in such approaches."

Provided by University of Cambridge

Citation: Getting schooled in the 'noise': learning about learning using big data (2013, October 1) retrieved 12 May 2024 from <https://phys.org/news/2013-10-schooled-noise-big.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.