

**NPD User Group meeting**  
Bristol, Wednesday 19 September 2018

**Venue:** Room 4.10, School of Education, Uni. of Bristol, 35 Berkeley Square, Bristol, BS8 1JA

**Organisers:** George Leckie and Becky Allen; **Funders:** Department for Education

**Website:** <http://www.bristol.ac.uk/cmpo/npd-user-group/>

**10:30 – 11:00 Registration**

11:00 – 12:00 **Neil McIvor and Richard Machen, DfE; Nick O'Donnell and Sarah Fisher, ONS**  
*Modernising 3rd party access to DfE data.*

12:00 – 12:30 **Henry Morris and Mark Williams, PiC**  
*The value of for-profit organisations using the NPD*

**12:30 – 13:15 Lunch**

13:15 – 13:45 **Dave Thomson, Education Datalab**  
*Linking NPD with Longitudinal Educational Outcomes (LEO)*

13:45 – 14:15 **Emily Emmot, UCL Institute of Education**  
*Children in Need Dataset within the NPD*

14:15 – 14:45 **Marta Lapsley, Greater London Authority**  
*Supporting school place planning in London local government*

**14:45 – 15:00 Break**

15:00 – 15:30 **Steve Strand, University of Oxford**  
*Ethnic disproportionality in the identification of Social, Emotional and Mental Health (SEMH) Needs: A national longitudinal cohort age 4-11*

15:30 – 16:00 **Yi Liu, University of Exeter**  
*The effects of the academisation of English schools on educational trajectories of children with special educational needs*

16:00 – 16:30 **Ben Styles, National Foundation for Educational Research**  
*Using the NPD to optimise education RCT design – the Literacy Octopus*

16:30 **End**

**Posters displayed throughout the day:**

**Matthey Jay, UCL GOS Institute of Child Health**

*Evaluating the Children Act: a legal epidemiological investigation of the educational outcomes of looked after children and children in need*

**Margaret Antony, Royal Holloway, University of London**

*School restructuring & efficiency: analysing the impact of the academies act 2010 on primary education in England*

**John Freeman, National Consortium for Examination Results**

*The NCER Children Looked After Project*

## PRESENTATIONS

**Neil McIvor and Richard Machen**  
**Department for Education**

**Nick O'Donnell and Sarah Fisher**  
**Office for National Statistics**

*Modernising 3rd party access to DfE data.*

At the 2017 user group, DfE outlined plans to undertake a programme of user research with a view to developing a 'safe setting' model for research access to DfE data including NPD. This presentation will outline the findings of that, the broader challenges at play, and progress in delivering new ways of working. The session will include lots of time for discussion of these issues and more generally to allow the audience to pose questions on other aspects of the NPD to the DfE team.

**Henry Morris and Mark Williams**  
**Founder & CEO**  
**PiC**  
**henry@pic.is and mark@pic.is**

*The value of for-profit organisations using the NPD*

As a private, for-profit organisation using NPD data to effect positive social change, PiC is concerned by moves to restrict access to the NPD. In this talk, the team will set out their concerns, present how they are using the NPD for social impact and offer a case in favour of private involvement with the NPD.

### Content

Introducing PiC - Introduction to PiC as a social enterprise using data to help professional bodies, employers and universities be more representative and fair in their selection using big data. Specifically, an overview of our story and the sorts of organisations we work with.

Current commercial arrangements - An overview of the current commercial arrangements regarding the NPD. Specifically, the process of application, the criteria that are applied to applications and how it is easier for commercial organisations to work with the likes of FFT.

Criticisms of using NPD data commercially - A look at the campaign by Defend Digital Me and more general concerns about the use of NPD data by private and for-profit organisations.

Our argument for using NPD data - PiC is only able to bring about positive social change through its use of the NPD. Our argument is that when proper safeguards are applied, private organisations like us can do a great deal of social good – fulfilling the intention of allowing access to the NPD.

A case study: what PiC does with NPD - To frame our argument for using the NPD, we will share a brief case study of how we use NPD data.

The case study will introduce our socio-economic background ranking tool currently used by several leading companies and professional organisations to monitor their social economic background diversity. We will also explain the efforts we make to ensure anonymity.

Conclusion: use caution, but don't go too far

The talk will conclude by reiterating the position and interest PiC has in this debate, especially imploring those in attendance not to 'throw the baby out with the bathwater' when it comes to limiting access to the NPD.

Intention of the talk - We hope the talk will open up and better inform the debate on access to the NPD, both in the User Group meeting and more broadly.

**Dave Thomson**  
**Chief statistician**  
**Education Datalab**  
**dave.thomson@fft.org.uk**

*Linking NPD with Longitudinal Educational Outcomes (LEO)*

The Longitudinal Educational Outcomes (LEO) dataset linked to NPD can be used to examine the longer term effects of education policy and choices, such as the impact of the Academies Programme and the impact of different qualification choices at 16 and 18. In this presentation I will discuss the coverage and quality of the LEO data for the first 8 cohorts of 16 year olds in NPD, i.e. those who turned 16 in the 2001/2 to 2008/9 academic years. As an example, I show the longer-term outcomes of those who took an advanced GNVQ, a popular alternative to GCSE in the early part of the century.

**Emily Emmot**  
**Research Associate**  
**Thomas Coram Research Unit**  
**UCL Institute of Education**  
**emily.emmott@ucl.ac.uk**

*Introducing the Children in Need Census: Things you need to know to use and understand the data*

In this talk, we introduce the Children in Need Census (CIN) - a case-based administrative dataset within the NPD, holding information on all children referred to children's social care services in England. Between October 2008 and March 2016, CIN contained information on ~2.76 million individuals and ~4.67 million social care case records.

CIN is potentially an invaluable data source holding information on potentially vulnerable children with additional support needs. However, handling and understanding CIN is a challenge in itself, requiring knowledge of the cohort and the social care system (Who are 'Children in Need?'), understanding of the available information (What exactly is in the CIN?), as well as the quirks and issues of the dataset.

Here we present an overview of the CIN for researchers, to facilitate future use of the dataset. We introduce the cohort, and review the data collection process, data structure and data content. We pay particular attention to data quality and validity, combining qualitative interview findings from three Local Authorities as case studies. We highlight differences in data recording practices between areas, as well as differences in the meanings of key terms such as “referrals” and “health services.” We discuss the implications of our findings in terms of using and understanding the CIN, and provide tips on ways to identify recording differences between Local Authorities.

**Marta Lapsley**  
**Data Scientist - Model Development**  
**Intelligence Unit**  
**Greater London Authority**  
**marta.lapsley@london.gov.uk**

*Supporting school place planning in London local government*

The Greater London Authority uses extracts from the NPD for a range of work to support the Mayor’s programme of initiatives aimed at driving up standards in education and ensuring there are suitable places for all children in the city. This talk covers four current use cases:

- London Schools Atlas
  - A publicly accessible web tool for anyone engaged in meeting the challenge of providing good school places for all, including local authorities, free school groups, academy chains, policy makers and commentators.
  - The talk will include a demo of features including visualisation of current patterns of demand for school places, primary to secondary school transitions, and overlays with other geographic data such as the Income Deprivation Affecting Children Index.
- School Rolls Projections
  - London’s dense population and extensive transport network mean that pupils often do not attend their nearest school. Using the home LSOA field from the NPD, we are able to produce projections for the number of school children in London which explicitly accounts for these mobility patterns.
  - We produce a set of publicly available roll projections which give policy makers a pan-London view, and also bespoke projections for London Boroughs as part of our School Rolls Projections Service. The Boroughs use these projections to inform school planning decisions and their SCAP returns to the DfE.
  - The talk will cover a high level projection methodology view, the benefits to London Boroughs of using the NPD, a view of our work flow for producing projections (using R, R shiny, and Excel), and ways in which we would like to develop the model.
- Parental preference analysis
  - We have started to carry out exploratory analysis of school parental preference data to gain a better understanding of school place demand.
  - The talk will discuss initial findings and what other aspects of the data we intend to investigate

- SEND trends/projections in London
  - We are currently working on analysis of London wide demand for SEND places, and borough level projections to aid boroughs in their planning for SEND provision.
  - The talk will include a brief discussion of preliminary results.

**Steve Strand**  
**Professor of Education**  
**Department for Education**  
**University of Oxford**  
**steve.strand@education.ox.ac.uk**

*Ethnic disproportionality in the identification of Social, Emotional and Mental Health (SEMH) Needs: A national longitudinal cohort age 4-11*

The over-representation of Black pupils in special education is a long standing concern which continues in recent analyses of national data in the US (e.g. Donovan & Cross, 2002) and in England (Strand & Lindsay, 2009; 2012). This paper reports on the results of a ESRC-DFE funded project investigating ethnic disproportionality in the identification of Social, Emotional, and Mental Health (SEMH) needs, a category of SEN, amongst primary school pupils in England.

A frequent explanation for the over-representation of Black students is inappropriate interpretation of ethnic and cultural differences including teacher racism, low expectations and a failure of schools to provide quality instruction or effective classroom management (e.g. Artiles et al, 2010; Waitoller et al, 2010). However an alternative hypothesis is that disproportionality reflects the fact that Black students are more at risk of SEN because of the greater socio-economic disadvantage (SED) they experience (Strand & Lindsay, 2009). Some recent US Studies have claimed that Black and Latino students are actually under-represented for SEN at age 9 after control for educational achievement and teacher's ratings of students behaviour at Kindergarten entry (Hibel et al, 2010; Morgan et al, 2015).

To address this question we analyse the England NPD focussing on the 562,274 pupils starting Reception class at age 4+ in September 2008, and matching in their records from the Annual School Census every January between 2008 and 2015 (Reception-Y6). This is the first study, to our knowledge, to explore ethnic disproportionality in the emergence of SEMH identification in a longitudinal study over ages 4-11 with national pupil level data. We utilise Cox's regression (sometimes also called event history analysis or logit hazard modelling) to identify how the likelihood of SEMH identification cumulates over time as children age.

We find that even after controls for pupils' age, gender, entitlement to a FSM, neighbourhood deprivation, attainment and social-adjustment at age 4, and for school composition variables, Black Caribbean (HR=1.42) and Mixed White & Black Caribbean (HR=1.46) pupils remain more likely to be identified with SEMH than comparable White British pupils. However this does not apply to the numerically larger Black African group, who are not over-represented in raw data (HR=1.12) and are significantly under-represented in the adjusted model (HR=0.65). The data militate against any simple Black/White interpretation and suggest a range of factors are implicated in disproportionate identification.

**Yi Liu**  
**Senior Lecturer**  
**Postdoctoral Research Fellow**  
**University of Exeter**  
**y.liu3@exeter.ac.uk**

*The effects of the academisation of English schools on educational trajectories of children with special educational needs*

In this study we explore the causal effects of the academisation of English schools on educational trajectories of pupils with special educational needs (SEN) using administrative data from the National Pupil Database. We examine the implications of schools' conversion to sponsored or converter academies on inclusion and educational trajectories of pupils with SEN. In particular, we are interested in whether converting a secondary school to an academy leads to fewer admissions of children with SEN in Year 7, change in the probability of removing the SEN status at the point of transition from primary to secondary school and change in the probability for pupils with SEN to move to another school.

Our modelling strategy is the following.

1. We apply the school fixed effects design to compare the proportions of the SEN children schools admit before and after academisation.
2. We model the effect of academisation on the probability of reclassification of pupils with SEN (from SEN to No SEN and vice versa) at Year 7 intake.
3. Using an instrumental variable design to model the assignment of the academisation "treatment", we model the probability of removing pupils' SEN status following academisation, for pupils that were enrolled in the "predecessor" state of schools before their academisation.

We find that converting schools to sponsored academies (but not converter academies) on average had negative effects on the inclusion of pupils with SEN, increased the probability of removing the SEN status during the transition from primary to secondary school, but increased the probability of staying in the school for already enrolled pupils with SEN. The findings are preliminary and subject to robustness checks.

In this talk we will discuss:

- Our experiences of cleaning and using the NPD;
- Our research design and modelling strategies;
- Our findings and the implications of the academisation and the pressure to raise academic standards on educational inclusiveness

**Ben Styles**  
**Research Director - Statistician at NFER**  
**National Foundation for Educational Research**  
**b.styles@nfer.ac.uk**

*Using the NPD to optimise education RCT design – the Literacy Octopus*

We use the NPD in a variety of ways to help in the design and execution of randomised controlled trials in education. These include baseline testing, randomisation stratification, sampling within schools, follow-up testing, sub-group analysis, missing data analysis, verification of trial results and longitudinal follow-up. The Literacy Octopus project, funded by the Education Endowment Foundation, consisted of two large multi-armed trials and was only possible through using NPD data. Here we focus on the trials from a methodological perspective, illustrating why trials that use an NPD variable for their primary outcome measure are usually the most robust possible in English schools. The trials explored whether sharing evidence-based resources with teachers in various ways made a difference to teaching and learning. One was a large-scale trial of simple passive approaches to sharing evidence-based resources, namely email and post, involving 12,500 primary schools. The evidence-based literacy support trial, which included both passive and more active approaches such as teachers' attending conferences and CPD sessions, involved 823 primary schools.



## POSTERS

**Matthey Jay**  
**PhD Student in Epidemiology and Statistical Science**  
**UCL GOS Institute of Child Health**  
**matthew.jay.15@ucl.ac.uk**

*Evaluating the Children Act: a legal epidemiological investigation of the educational outcomes of looked after children and children in need*

Protection of children from harm and the promotion of their welfare in England is achieved through a variety of statutory mechanisms. Chief among these is the Children Act 1989, which provided a unified code for resolving disputes about children's upbringing. The Act invests powers and duties in local authorities (LAs) with respect to assessing children and providing services for children in need (CiN); instigating child protection investigations and child protection plans; and receiving children into care either through a court order or with consent of the parents, by which they become children look after (CLA). The welfare of the child is the paramount consideration. To date, limited population-level research has been conducted into either how the system operates or the long-term outcomes of the children it affects. This is especially so regarding their education. The research question of my PhD are:

1. For children who become CLA or CiN during primary school, what is the nature of the schools that they attend for their secondary education in terms of school type and quality and what factors determine this?
2. How well do CLA and CiN perform in their GCSE exams and what factors predict this?

My presentation will outline this legal background, provide some descriptive data of my cohort under study and further plans for analysis.

**Margaret Antony**  
**PhD Student in Economics**  
**Royal Holloway, University of London**  
**margaret.antony.2012@live.rhul.ac.uk**

*School restructuring & efficiency: analysing the impact of the academies act 2010 on primary education in England*

Education policy reform in the UK, in recent years, has witnessed a landmark change with the implementation of the Academies Act 2010. The Act paves the way for greater school autonomy in the primary school sector of the country. Under the new reform, the academy programme which began in the secondary school sector has been extended to include maintained primary as well as special schools. Of the 16,766 primary schools in England, 2,440 now have academy status. This de facto variation in institutional arrangements provides a rich context for analysing the effects of school restructuring and forms the focus of the current paper. Greater school choice and autonomy, it is envisaged would raise standards and improve performance and efficiency in state schooling. Whether this is evident in practice is, however, a contested issue. The current research sets out to examine the aspect of technical efficiency, in the context of school restructuring. Efficiency, here, refers to how well a school utilises the resources at its disposal in maximising a common educational outcome – KS2 SATs results. Applying a novel approach in policy evaluation studies, called the synthetic control method, the

study attempts to estimate the 'academy' effect on efficiency. Pre and post-reform efficiency trends of academy schools (treatment group) are compared to that of maintained schools in the sample (control group). The identification strategy involves defining a 'synthetic' control group for each treatment unit as a weighted average of units (schools) that do not receive treatment, but similar in pre-treatment characteristics and outcome trajectory. The outcome of interest – school-level efficiency scores, are obtained using the stochastic production frontier methodology in efficiency measurement. The analysis is undertaken separately for the pre and post-treatment period, respectively. Efficiency-outcomes for the control units pre-treatment are then weighted to determine the counterfactual outcome for each treated unit, post-treatment. The study uses data on school inputs and test-score output for the census of primary schools in England. The primary source of data is the National Pupil Database (NPD) that compiles School Performance Data for the two key stages of primary schooling (KS1 and KS2) with School Census data on school and pupil characteristics and data on school finance from the Consistent Financial Reports covering all primary schools in the country.

The study has much academic and policy relevance in analysing the effectiveness of schools and the effectiveness of education policy. Given the paucity of studies on the 'efficiency' aspect of school performance, the current study makes a useful contribution to the 'economics' of schools literature.

**John Freeman**  
**Chair**  
**National Consortium for Examination Results**  
**johnfreemanncer@yahoo.co.uk**

*The NCER Children Looked After Project*

NCER has developed an online pupil-level tool for use by Virtual Headteachers to set education targets and to allocate Pupil Premium Plus funding for individual children in care, based on fair comparisons across the LA and nationally, using NPD and SSDA 903 data. The tool is based on the Rees Centre research on the care factors that impact on educational outcomes. The project was initially supported financially by the DfE, the Association of Directors of Children's Services and is now part of the NCER core offer.