

Transcript: Exploring Educational Outcomes through National Datasets



[0:00:00]

Jools Kasmire: Hello, everyone, I'm Dr. Jools Kasmire from the UK Data Service Computational Social Science Training Team, based in Manchester. Today I'm going to talk to you about the UK Data Service and how it can contribute to an exploration of educational outcomes. I'm going to start by giving you a quick summary of what the UKDS is. Then I'm going to give you an overview of the data we hold, and finally I'll show you how to get started finding and accessing our data.

The UKDS is a comprehensive resource funded by the Economic and Social Research Council, one of many funded by UK Research and Innovation. As such, we provide access to the largest collection of social, economic and population data in the UK. Data providers deposit their data with us, and then we make this available to users. As well as providing access to all of the data, the UK Data Service provides support, training and guidance to help researchers find, access and use this data. So who's it for? Well, we like to say it's for everyone. This includes, but is not limited to, academic researchers and students; government analysts; charities and voluntary sector foundations; business consultants; independent research centres; think tanks and many more options.

And our data comes from many different sources. A lot of official agencies, mainly the central government; international statistical time series; research institutions; individual academics who make their research available as part of their research grants; market research agencies; public records and historical sources, lots of things. And we hold all different types of data. Mostly survey data, both cross-sectional and longitudinal. We also hold aggregate data, international macro data, census data and qualitative and mixed-methods data, as well as a few things that don't fit in any of these categories very well.

So now, I'll just quickly cover how you can find and access all of these different sources of data from the UK Data Service. If you don't know how to get started, you can click on the Find Data tab at the top, which directs you to a few common options, as well as to some links to tutorials, to learn more. One of the most popular options is our catalogue search tool. The catalogue search tool allows you to search the data sets available from the UK Date Service. You can enter relevant search terms, search for particular data sets through name or study number. You can also use the filters down the left-hand side to further refine your search.

But maybe you also want to browse data. This is available as well, the theme being the first browsing option presented to you. One of the top themes is education. There are more themes if you scroll down, but if you keep scrolling down you will also find you can browse by data type. Keep scrolling down even more to browse by teaching data sets, as well as a few other categories that we have grouped together under 'general'. You can also search for particular variables within data sets, in our variable and question bank. For example, if we were to type in education, we could see which data sets contain variables on this topic. Just to note on the variable and question bank, though, this doesn't contain all of the data sets we currently hold, so it's worth searching the catalogue for later versions of data sets that you find through the variable and question bank, to double check.

The options on the left-hand side here, that you can expand with these little plus signs, these are specific to the variable and question bank. Of course, you could also choose to search instead in the quality bank. This is especially useful if you are interested in qualitative data, as it allows you to search key terms within different data types of qualitative resources. Note that here too, the options and menus on the left are unique to the quality bank. Finally, we have the HASSET thesaurus. HASSET stands for Humanities and Social Science Electronic Thesaurus, for those to whom that may not be obvious. This search tool contains key social science terms and related concepts.

[00:04:48]

For example, here, Conservatories is categorised under Educational Institutions, which is under Educational Systems and Institutions, which is under Education. This nesting structure is clear when you look at the tool via the Hierarchy tab. Alternatively, you can look at the thesaurus through the Alphabetical tab, which is pretty self-explanatory. The thesaurus is a helpful way to show you which other terms might be useful to search in the catalogue, either because they fall higher or lower in the hierarchy, or because they are listed under the entry terms and related concepts within an article.

Just a quick point here on different data access levels. Starting at the most restricted types, we have controlled data, which is also called secure access data. This data can only be accessed through a safe room, a secure lab within the UK or the SafePod network. Next, we have safeguarded data, which is available through end user license. Access to this kind of data requires registration. Some of this data is also special license data, which has additional requirements. There is also re-shared data, which is self-deposited by data creators or owners. Access to and controls on this kind of data can vary. Finally, the least restricted type is open data. This is data that can be accessed by anyone, even without the need to register.

Importantly, most of the data in our collections can be accessed free of charge. You can simply register using your institutional log-in. If you don't have an institutional log-in, you can apply to get a UK Data Archive username. Thanks, everyone, for listening, and let me know if you have any questions.

Francesca Borgonovi: I am Francesca Borgonovi, and I lead the Skills Analysis Team in the OECD Centre for Skills, and I am also an honorary professor at University College London in the Social Research Institute. I will talk about gender gaps in education through the lens of international large-scale assessments. And so the way in which I will structure my talk is to provide first an overview of international large-scale assessments, what they are and some of the different assessments that are available. And then I will detail one of the most important assessments that the OECD leads on, the Programme for International Student Assessment. And then I will provide you with some examples of my own research and detail how

international large-scale assessments can be used to examine gender gaps in education.

Now, international large-scaled assessments in education are system-level instruments, so they are not assessments designed to give information about the achievement of individual students, but are rather designed to allow the measurement and allow comparisons at the system level, and how well different systems are equipping young people with knowledge and ability. They are typically low-stake, so there are no consequences for individual test-takers. They include actual assessments, or tests, as well as questionnaires that are used to have an understanding of test-takers - who they are, socio-economic background et cetera. They are called large-scale because they involve large representative samples from many different countries in different continents.

There are typically repeated cross-sections that allow for monitoring over time, as well as comparisons between countries. And they mostly focus on school-age populations, although in particular the OECD has led efforts on assessments of adults. Now, one important difference between different assessments is what they intend to measure, so before starting analysing data it is important to have a look at the assessment frameworks of different assessments. Different assessments I think can be, in a sense, divided between those that have an emphasis on the curricular content and coverage, and so the test questions are developed in a way that reflects curricular content, whether that is in mathematics or science, for example. And a different set of assessments, on the other hand, emphasises problem-solving abilities and transversal skills.

And because of this different focus, the target population of the two types of assessments is quite different. The assessments that focus on curricular content are grade-based because you want to see if, by a certain grade, young people have mastered the material that is part of a particular curriculum, for example.

[00:10:02]

Whereas problem-based assessments are mostly age-based. Now, the OECD and the IEA are the twin institutions that are responsible for the majority of international large-scale assessments that are across continents. There are also a number of initiatives that are regional in nature, but I will not cover them in my presentation. The OECD is responsible for the Programme for International Student Assessment, PISA; the areas covered are reading, maths and science, these are the core domains, as well as innovative domains, and the focus is 15-year-old students. PEAC, the survey of adult skills, focuses on literacy, numeracy, problem-solving, and the focus is the population of 16- to 65-year-olds, so it goes to older age groups but also focuses on young people who are not in school anymore.

The International Early Learning and Child Well-being Study focuses on social and emotional skills and it covers 5-year-olds and the Study of Social and Emotional Skills covers social and emotional skills and it focuses on 15-year-old students, although there is an option for 10-year-olds. The IEA's study, TIMS, Trends in International Mathematics and Science study, focuses on mathematics and science. It has a grade-based target population of fourth and eighth graders, although there is the possibility for countries to also adopt an assessment of the last year of secondary school. PIRLS focuses on reading literacy among fourth-graders; ICCS is about civic and citizenship competencies, and again it is targeted at eighth graders, and ICLS focuses on digital competencies and it focuses on eighth grade students.

Now, international large-scale assessments have been used in international batch marking and global monitoring initiatives such as the sustainable development goals; in particular SDG 4 on quality education focuses on this aspect of quality. So the evolution... It is not the only one, but an important evolution from the millennium development goals is to not only monitor the participation in education but also the outcomes of education, whether young people who are taking part in education are developing skills, are learning something. And so in this context, international large-scale assessments have

played an important role to monitor trends over time for the same country, but also to monitor disparities in education provision for different groups.

Now, because different studies focus on specific world regions, not every country takes part in every single study. There have been attempts to link across studies to provide global mapping of learning outcomes and human capital, using linkages across the different studies. And here is an example of an initiative that has tried to do that.

Now, moving to the Programme for International Student Assessment, PISA, this is a low-stakes assessment, it is designed to take around two hours to complete, so it is a timed assessment, and the test is followed by a questionnaire. The aim is to provide system-level monitoring so there are no grades or estimates for individual students. The target population is 15-year-olds, in fact, it's students between the ages of 15 years and 3 months and 16 years and 2 months at the time of the assessment window, who attend at least lower secondary school, and it is problem-based. It has been implemented every three years since 2000, although the 2121 edition was postponed because of COVID, and it has been implemented in the latest edition by over 80 countries and economies worldwide.

The core assessment domains in every cycle are science, mathematics and reading and the 2022 assessment focused as a main domain on mathematics, as well as innovative domains. There have been assessments of problem solving, collaborative problem solving, financial literacy, global competence and creativity. The sample size is a minimum of 4500 students per country, attending at least 150 schools. There is flexibility - countries are allowed to do over-sampling or regional sampling or add grade-based components for their own purposes, and the study is conducted in the language of instruction, and this has important consequences for monitoring the skills and abilities of particularly very recent migrants in a country.

[00:15:49]

The study was conducted using pencil and paper instruments until 2012, while it has been administered using computers since 2015. And this has an influence on the test design and the metrics designed and used in the assessment, allowing for a more complex design. But also on the items themselves, because with computers it is possible to design scenarios or interactive items and it is possible to adopt an adaptive element to the test. On top of the test, there are reach the ground questionnaires that are administered to students as well as school principals and parents. And there is also the possibility for countries to implement ad hoc student questionnaires, for example on ICT or educational careers as well as additional assessment instruments.

Now here is a link to the PISA databases and analysis tools that are available on the OECD website. You can search for the different years, and there is the possibility to explore data interactively on the PISA main page. All the questionnaires as well as national versions are available on the PISA website, as well as the code book and figures and tables produced by the OECD in the international reports. And finally it is possible to download all the databases, either in SAS or STS, and what you can also find are the codes that were used to compute the main indicators that are used by the OECD. And importantly, because indices to describe the economic, social and cultural status of students need to be comparable over time as well as across countries, you can see the time trend data here, in case you are interested in doing that.

Now in terms of some of the analysis, I wanted to show you some results of work that I have done, moving away from gender gap singular to gender gaps plural, and how these vary, depending on what, where and when gender gaps are assessed. So in terms of what, even if we do not look at different assessments and we focus just on PISA, things look quite different depending on the assessment domain that is considered. Now, the PISA scales are standardised to have a mean of 500 and a standard deviation of 100 in the base year. And a different gender gap in favour of boys, for example in mathematics, of 9 points, is roughly equivalent to 9% of the standard deviation. In terms of reading, the

gender gap is in favour of girls and it's much larger, 24 points, and in collaborative problem-solving, in 2015, this was 29 points.

Now gender gaps vary depending on the performance distribution, so what you can see is that the gender gap in mathematics is much larger at the 90th percentile, it is 22 score points on average. And the gender gap in reading in favour of girls is, on the other hand, much larger at the bottom of the performance distribution rates, 34 points. In terms of when, one way in which we can consider temporal dimensions is the age at which assessments occur. So when comparing the gender gap in literacy and numeracy on synthetic cohorts that took part in different large-scale assessments at different ages. So aged around 9, 10 for TIMS, PIRLS; aged 15, 16 for PISA and then 26, 27 for the OECD survey on adult skills. What we do notice is that the gender gap differs for different age groups of the same cohort. Whereas for literacy it takes an inverted U shape, in numeracy it grows over time, in favour of men.

And if you want to see what happens when you look at the evolution of numeracy gender gaps across the distribution, this grows larger, particularly at the 90th percentile. Another way to consider the 'when' question is to consider the test, the PISA test, as being a long test. And it is possible to see how the performance of girls and boys changes over time in this two-hour test. So here what you can see is the gender difference in the decline in the proportion of correct answers that boys and girls provided in the reading test in PISA 2009. And what you can see is that there is quite a bit of difference in performance across countries, but in general the performance drop is higher for boys than for girls. And this is important, because in another study in which we were able to look at longitudinal follow-ups of cohorts that took part in PISA in 2000 and 2003, in Australia, Denmark and Switzerland, what we saw is that on top of controlling for performance, performance decline predicted the likelihood that young people would complete university.

[00:22:45]

Now, in terms of where, one interesting fact is that the gender gap in reading is extremely wide when young people are tested in the context of PISA, but it's much, much smaller when young people are tested in the survey of adult skills. Now here are results from a study that I did comparing the gender gap in literacy in PISA and PEAC in 2012, so it is the same cohort. I only focused on young respondents in PEAC, and what you see is that the gender gap is very much present and large in PISA but it is much, much smaller in PEAC. And even when you try to standardise the test for different factors you cannot reconcile the difference. Now, an important difference between PISA and PEAC is that PISA is implemented in the context of schools while PEAC is implemented in people's houses. And there is a possibility that the setting may matter for gender gaps, and that in particular boys may perform and be motivated to a different extent when they take assessments in a school setting and take assessments in a household setting.

And finally I think I wanted just to bring about a couple of issues related to the new measurement possibilities that come from the fact that the assessments are now conducted on computers, in particular the OECD assessments allow you to analyse process data; log files that contain traces of communications between the test taker and the computer interface; and they are being increasingly because of methodological interest and substantive interest in terms of how long it takes for individuals to solve different questions, and whether the problem-solving process follows some logic or not. Now, at the moment, a lot of the test items haven't been specifically developed using a theoretical model of human cognition, and so they are opportunistic, the way in which information on the text-taking process is not necessarily responding to some ex-ante willingness to develop indicators.

There are some exceptions, in particular the PISA 2012 problem solving assessment, for which there were strategies designed to see students try to vary one thing at a time to provide answers. And I think log files are valuable, in which the set of information can be collected in a cheap and scalable way. And they enable to observe, and I would say quote, unquote, test takers' behaviours. And

I would say at the same time that not everything we would like to observe is recorded, and the interpretation of log files is invariably item-specific. And this poses some challenges because interpreting information is impossible without accessing the full item description and what happens in the files.

And just to give you an example that log files do not necessarily provide all the information about what people are doing, here's information from the study in which we looked at enmeshed information from the log file with information from what people were doing using glasses that recorded information from test takers' behaviours. And so a long delay in a log when somebody is trying to respond to a test question could signal the fact that a person is engaging, but it could also signal that they are using tools outside of the computer platform to solve the item; for example, they may be using their hands or they may be using an external calculator, and in this case it is difficult to interpret timing information as providing information on engagement, motivation or rather, for example, what people are doing outside of the computer platform.

And with this I would like to thank you for your attention and I will close. Thank you.

Claire Crawford: My name is Claire Crawford. I am an associate professor of economics at the UCL Centre for Education Policy and Equalizing Opportunities, and I am here to talk to you today about some work that we've been doing using the longitudinal education outcomes data, on which I'm leading a grant in partnership with the Department for Education to support the development and enhancement of that data.

[00:28:44]

So just to give you a bit of background to our project, we know that there's a vast literature highlighting really substantial gaps in education outcomes between individuals from different socio-economic backgrounds. A slightly more limited literature on labour market gaps. But in many cases in these pieces of analysis, the groups of interest are quite coarse, especially when they are determined

using administrative data. So, for example, in a UK context we have an indicator of low family income, which suggests whether a student is eligible for free school meals or not. In that case the group is just split into two, one who is eligible, one who is not. And there is obviously potential for substantial variation in outcomes within those two relatively large groups. So now we have new administrative data which means we can actually track individuals throughout their education careers and actually now into the labour market. That means we can capture eligibility for free school meals throughout a variety of points in time throughout an individual's school career, and that allows us to get a slightly richer measure of socio-economic background for those individuals. So what we're doing in this study is exploring how education and early labour market trajectories vary for those who are eligible for free school meals for a different number of years.

And we are doing that using some relatively recent administrative data which has been collated in England, called the longitudinal educational outcomes, or LEO, data. So for those who are not aware of this data, it essentially combines the education and labour market records for anyone who has interacted with either the school or further education system in England since 2001/2. So that's mostly going to be individuals who have been through the school system since then, born since 1985/6. So on the education side it pulls together records from the school system, from a database called the National Pupil Database, that contains a census of all individuals attending state schools in England, and that's where we derive our free school meal eligibility from, and it also includes a variety of measures of attainment which are taken by students across the schooling system. That's linked to further education records from what's called the individualised learner records, and also higher education records as well. And more recently we have added to the data university applications information, so not just where someone ends up at university, but also where they applied.

So combined with all of these education records, we have tax and benefit records. On the tax side we have information on both employees and the self-employed from the different tax systems that collect that information. For employees we observe different employment spells, when they were working for different

companies, as well as how much they earn in any given tax year. We have also recently added to the data information on the industry in which the individual's firm is based, but one downside to this data is that we don't have access to information on hours of work, so that's obviously going to be an important limitation when we think about how much someone is earning. We can combine that with self-assessment data, that means information from individuals who are not employees but might be self-employed, so we have a measure of annual income or profits for those individuals. And then on top of that we can supplement information on out-of-work benefit spells, so we know when those claims started and finished as well as the type of benefit that individuals were claiming, and we have both of those sets of records across pretty much the whole stretch of the data to which we have access.

So in our study we are going to be focusing on one particular cohort, which is those who took their GCSEs, the national qualifications taken by most students at the age of 16, in 2011/12. And we chose that because it gives us a good stretch of information, both beforehand and afterwards. So in terms of backward-looking information, we are able to track an individual's eligibility for free school meals, pretty much through their schooling from primary through to secondary, so from years one to eleven. But then we can also follow them forward into the labour market, up to around age 23, which brings us up to the 2018/19 tax year, i.e. before the pandemic.

So in the analysis I am going to present to you today, we focus on individuals with a complete free school meal history, so that means we just focus on individuals in the state sector, we don't capture individuals who might have exited the system in order to take up a place in a private school for any period of time, for example, and that gives us around 450,000 individuals in total.

[00:33:55]

So what we then do is split the sample according to the number of years of free school meal eligibility. Of course, there are still going to be individuals who were never eligible at any point during their school career, and what we do for those individuals is use information on the neighbourhoods in which they live, very local neighbourhoods, and we draw together information about the characteristics of the populations who live in those neighbourhoods, in terms of different measures of socio-economic status, in order to provide a proxy for their own individual measures. So for those individuals we split them into four groups, quartiles, on the basis of how disadvantaged or advantaged their very local neighbourhood is.

So just to give you a flavour of the variability that we have in the data, this table is showing you the number of years for which an individual is eligible for free school meals, split according to primary school in the rows and secondary school in the columns. So you can see here through this figure that I have highlighted in red that around 70% of individuals in our sample were never eligible for free school meals at all during their school career, and we can contrast that with around 5% of individuals who were eligible throughout almost the entirety of their school career. But you can see that there is a mixture of individuals spread across the remaining cells of the table as well. So there's quite a lot of movement in and out of the system, and you can see here that for individuals who were eligible for just one or two years in primary school, about 5% were in that group and then were not eligible at all during secondary school. By contrast, if you focus just on individuals who had a similar number of years' entitlement, but in secondary rather than primary school, you can see that figure is a bit lower and that is typical throughout the cells in the table. People are more likely to be eligible for free school meals in primary rather than secondary school. And in future work, we are going to explore how much that might matter for later outcomes.

What I'm going to be talking to you about here today is individual trajectories, so what we're really interested in doing is characterising how individuals are moving through post-compulsory education and into the labour market. And the way in which we do that is to classify the activities that individuals are undertaking in each six-monthly period between the ages of about 17 and 23. And the way in

which we classify those is first in understanding whether they're undertaking any educational qualifications during those periods, and we can classify individuals separately according to both the level of education and the type of education. So whether it's below level 2, 3 or 4, and for level 2 and 3 we can split that into academic and vocational education. So for those who are less familiar with the English system, level 2 is roughly equivalent to a set of GCSE passes, and level 3 is roughly equivalent to a set of A level passes, and then level 4 and above is higher education.

And a really nice feature of the LEO data is that we are also able to look not only at whether people are in education at a given point in time, but also whether they are simultaneously in work. So for each of those types of education we can classify people according to whether they are or are not also working in the same six-monthly period. And then we've additionally got a set of other states which are just being in work, just being on out-of-work benefits and not being observed in our data. And in total that gives us about 15 states which we're measuring across each of 14 periods. So we will be classifying trajectories using that information.

And then finally I will just talk briefly at the end of the presentation about some results we've got using daily earnings. So I mentioned earlier that we had access to annual earnings and employment spells, but not hours, but obviously using that information we can also calculate a measure of daily earnings, and so we've made use of that in some of our analysis. Okay, so diving straight into some results I am going to show you four of these types of pictures. So what we're doing here is, for each six-monthly period between 17 and 23, we're summarising the proportion of the cohort who are in each of those 15 states that I mentioned.

[00:38:50]

And we're doing that separately here, and I will show you another slide which does a similar thing for different groups, but here I'm focusing on those who are in the no free school meal eligible part of the population, the lowest quartile of that group on the left-hand side, the highest quartile of that group on the right-

hand side. So you can think of this as the most and least disadvantaged amongst that group of individuals. So in each case we've grouped activities in different colours, so the different shades of each colour are telling you, in terms of the education outcomes, whether someone is working or not working at the same time.

So if we take the right-hand figure to start with, we can see that this is really a traditional route through an academic set of qualifications. So the kind of sizeable brownish figures in the middle on the left-hand side are people who are taking level 3 academic qualifications, mostly A levels, and in each case the darker colours are showing people who are just doing that qualification, the lighter shading is showing those who are doing that qualification but also working alongside it. So most individuals in this group are following that track for a couple of years, and then moving into university, so that's the level 4 and above, very sizeable orange bars at the top, again split into whether individuals are working or not. And then once they've exited higher education, many seem to be moving into employment, the grey shaded areas.

So we can contrast this with the much more varied picture, particularly in terms of the earlier periods that we're seeing on the left-hand figure. So here we have still a reasonably sizeable proportion of the population following that traditional, academic route, but we've also got many more individuals going through vocational education routes, the darker browns and lighter blues at the top. And similarly moving into employment once they've completed those qualifications. So what we can see on this figure is exactly the same types of diagrams, but here instead focusing on those who are eligible for just one or two years of free school meals during their education careers, and then for almost the entirety of their education careers on the right-hand side.

So I'm not going to flip backwards and forwards so you can check it, but the figure for those eligible for one or two years of free school meals, which was the first group just inside our more disadvantaged group, is not a million miles away from the figure I showed you for the lowest of our quartiles based on the neighbourhood measures. There's not a dramatic difference between individuals,

just as you cross over that boundary for being eligible for free school meals. On the right-hand side, though, you can see that there are some quite big differences compared to the left-hand figure. So I guess the most noticeable of these is the fact that we have a really substantial proportion of the cohort who are on out-of-work benefits, the kind of white shading towards the bottom of the diagram, and a commensurately much lower proportion of the cohort who are in work. The pattern in terms of education is not quite so striking; certainly, fewer going through the academic route from A levels into university, but still a large number going through the vocational qualifications, but also quite a lot doing below level 2 qualifications in yellow.

Now of course the challenge with those kinds of figures is that it is not necessarily showing us what any given individual is doing as they move through the education system and into the labour market, it's just saying, in summary across this group, what are the activities that people are undertaking at different points in time? But what we want to move towards doing is summarising the actual trajectories that people are following. So classifying people following what I call the more traditional route of A levels to degree and into work, and differentiating that from the different patterns in terms of moving through vocational qualifications or the other routes we see. So what we're doing in order to elicit this is some sequence and cluster analysis, so this is a methodological approach to try and summarise the similarity of different pathways and group them together in a kind of way that makes it more tractable for analysis.

So this is a first pass attempt at that, this is all work in progress, we are currently developing this methodology further, but this gives you a flavour of the types of things we are doing. So this is classifying each of our socio-economic groups according to the proportion that are in each of seven clusters, and the seven clusters we have named according to our perception of what the routes look like.

[00:44:20]

So you can see that we've moved from no activity and benefits dependency for clusters one and two, you can see that the benefits dependency, as was suggested by the previous figures I showed you, is very strongly socially graded. And quite a large proportion of the population are falling into that group when we are looking at the most disadvantaged. And the towards the other end, you see clusters six and seven which are that traditional route through the academic system, split according to whether someone is working whilst they are studying, in green, or not, in dark blue. And again you can see those are socially graded but in the opposite direction, with those in the highest socio-economic status group being the most likely to follow those routes, and actually even more strongly graded for those who are not working at the same time, as you might expect.

So what we're ultimately going to end up doing with those clusters is to understand to what extent differences in pathways can help to explain these very large differences in earnings by socio-economic background. So here we are using information taken at age 25 for this cohort, and it's showing you how much people are earning relative to those in our SES quartile one, so those who are just not eligible for free school meals at all during their school careers. And you can see that there are really substantial differences across the distribution for men on the left, women on the right, so in terms of the higher socio-economic status individuals, for example women in that top group are really earning a lot more than those in SES quartile one. Men still earning more but a relatively smaller effect size. And then you can see, in terms of the penalty associated with being eligible for free school meals for a larger number of years, roughly similar between men and women, and approaching 20% by time you get to being eligible for your entire school career. So really very significant differences in earnings across the socio-economic distribution. And so what we will be doing next is understanding the extent to which these different trajectories can help explain these patterns.

So just to wrap up, what I've shown you today hopefully gives you a flavour of the power of some of the LEO data that we've been able to use. It's demonstrated a very strong monotonic relationship between years of free school meal eligibility

and later outcomes, and that leads us to think that there is a significant gap within the gap, as people have referred to it. So the idea that if you just look at the difference between free school meal eligibility and not, there's already a large difference, but actually if you further differentiate those two groups, you see the scale of the differences increase as you are able to differentiate people more and more.

So in terms of next steps, I've mentioned these as we have gone through, but we are going to be focusing more attention on the idea of whether eligibility for free school meals earlier versus later, for example at different ages, whether that matters, or similarly if you hold the number of periods constant but you have them all grouped together at once, as opposed to moving in and out, does that matter in terms of trajectories and later wages? And then, as I mentioned, the next part of what we are going to be doing is a mediation analysis, which is going to try and help us understand how much of these very large differences and wages we can explain using the kind of trajectories that we are creating. So hopefully that gives you a good sense of the data and some examples of the types of analysis that can be done with it, and now very happy to answer any questions you may have. Thank you.

[End of Transcript]