The University Of Sheffield.

Department Of Economics.

Sheffield Economic Research Paper Series.

Estimating the impact of health on NEET status

Daniel Gladwell, Gurleen Popli and Aki Tsuchiya

ISSN 1749-8368

SERPS no. 2015016

June 2015

Inst EAD. | Institute for Economic Analysis of Decision-making.

www.sheffield.ac.uk/economics instead.group.shef.ac.uk

Estimating the impact of health on NEET status

Daniel Gladwell [a]

Gurleen Popli [b]

Aki Tsuchiya [c]

June 2015

Abstract

This paper uses a dynamic Structural Equation Model of ability formation to investigate the determinants of NEET (not in education, employment or training) status in adolescents, with special focus on health. The model addresses the issue of measurement error in estimating ability and mental health; and explores the determinants of ability and NEET status through time. The analysis finds that ability remains the key predictor of NEET status; and while general health plays an important role in the formation of ability for both girls and boys, the impact of mental health differs between the sexes.

JEL Classification: I10, I21, J21

Keywords: Adolescence, NEET, Health, Structural Equation Modelling

[a] School of Health and Related Research, University of Sheffield, UK. (d.gladwell@shef.ac.uk)

[b] Department of Economics, University of Sheffield, UK. (g.popli@shef.ac.uk)

[c] School of Health and Related Research & Department of Economics, University of Sheffield, UK. (a.tsuchiya@shef.ac.uk)

Acknowledgements: We are grateful to the Department of Education for the use of the Longitudinal Survey of Young People England (LSYPE), and to the UK Data Archive and Economic and Social Data Service for making them available. Daniel Gladwell is funded by the University of Sheffield Faculty Scholarship. We would like to thank David Fairris, Steve McIntosh, and Jennifer Roberts for their comments on an earlier draft of the paper. All responsibility for the analysis and interpretation of these data lies with the authors.

1. Introduction

NEET is a term used to refer to young people who are *n*ot in *e*ducation, *e*mployment or *t*raining. There are numerous choices available to young adults when their compulsory education ends. Some go on to further education, others go into employment, and still others into training. There are, however, a substantial number who do none of the above and are classified as NEETs: this would include all those who are unemployed and looking for jobs, and those who are inactive (also referred to as discouraged workers). In the UK the percentage of 15-19 year olds who were classified as NEET in 2014 was 9.5%, the equivalent figure for OECD which was 7.2% (OECD, 2014). [1]

There are long term consequences of being NEET, both for the individual and society. For the individuals evidence suggests that there is a degree of correlation between the status immediately after compulsory education and a range of longer term outcomes. Those who leave full time education early are unlikely to return to it (Dickerson and Jones, 2004); and the resulting lower educational attainment is associated with both lower pecuniary outcomes such as lifetime wealth and consumption (Card, 1999 and 2001), and lower non-pecuniary outcomes regarding adult health, marriage and parenting style (Oreopoulos and Salvanes, 2014). Further, youths who face spells of unemployment and inactivity immediately after the end of their compulsory education have lower attachment to the labour market in the long term (Bell and Blanchflower, 2011), and lower earnings later in life (Gregg and Tominey, 2005; Mroz and Savage, 2006). There are associated societal costs: NEETs are more likely to claim benefits and attach themselves to the informal economy; and loss of earnings of individuals results in loss of tax revenues and increased welfare costs (OECD, 2013). In a UK based study Coles et al. (2010) estimate the public finance cost of NEET, based on 2008 figures, to be £12 billion: largely as a result of accumulation of benefits and lost tax revenues, but also the (relatively small) costs for the health and criminal justice systems.

While NEET status is observed after the completion of compulsory schooling the process which leads to this outcome is understood as a culmination of factors arising and decisions made over the years up to school leaving age. In this paper we try to understand the factors and pathways that lead a young person to drop out of education and labour market and be classified as NEET. We use the data from the Longitudinal Survey of Young People in England (LSYPE [2]) and look at a group of adolescents who turned 17/18 years old in 2008. We use a dynamic model of ability formation (Cunha and Heckman, 2007), where we use the available information on: the educational progress and attainment of these young people throughout their secondary education (starting age 13/14); their socioeconomic background; their own and their parents' aspirations; and their health. The final outcome we are interested in is the probability of being NEET at the end of compulsory education (age

[1] OECD definition of NEET covers 15-29 year olds; the UK definition of NEET covers 16-24 year olds. The school leaving age in the UK is 16 (which is higher than the OECD median of 14 years), and almost 100% of the 15-16 year olds in the UK are in education.

[2] For details on the LSYPE see: https://www.education.gov.uk/ilsype/workspaces/public/wiki/Welcome/LSYPE

16) and persistence of this state a year-on at age 17/18. In our analysis we pay particular attention to both the direct and the indirect (via past ability) impact that health has on individuals' likelihood of being NEET.

In all the studies that we are aware of NEET status is modelled as a static concept, using one of the following two approaches. (1) NEET status at any given time is modelled as a binary variable (for example, Duckworth and Schoon (2012), Cornaglia et al. (2015), Mendolia and Walker (2014)) such that the young person is categorised as either NEET or not, where not being NEET could mean any of the following: in full time education, in part time education, in training, or in work. (2) NEET status is modelled as a categorical variable (for example, Crawford et al. (2011) and Dickerson and Jones (2004)), such that the young person is explicitly categorised as being in education (full time and/or part time), in training, in a job, or NEET.

In our analysis we model NEET as a dynamic concept, but at any given time we consider NEET as a binary variable. Incorporating everything within a dynamic model means we are able to look at the relative importance of the different determinants of NEET status and the stage of a young person's life when these factors have the biggest impact. Ours is the first study, as far as we are aware, that looks at the impact of both physical and mental health on NEET status within the same framework. We use a Structural Equation Model (SEM) to understand the process that leads to the outcome of a young person being NEET. The methodology allows us to address the issue of measurement error in estimating ability and mental health. It also allows inferences (both direct and indirect) to be made about an individual's ability at a singular point in time while enabling an exploration of the determinants of ability and NEET status through time.

The remainder of this paper is organised as follows. In the next section, we briefly review the relevant literature on the determinants of NEET status. Section 3 presents the empirical specification that we use to estimate the dynamic model of ability formation and estimate the probability (and persistence) of NEET status. Section 4 describes the data and the different variables we use in our analysis. Section 5 presents our main results, and section 6 draws some conclusions.

2. Background: determinants of NEET status

The factors determining NEET status often discussed in the literature are: earlier academic attainment of the young person; parental socioeconomic status; aspirations and attitudes, both of the parents and the young person; neighbourhoods; the macroeconomic conditions; and health of the young person.

Dickerson and Jones (2004) in their work, using the Youth Cohort Survey [3] from the UK, find considerable persistence in the state/activity over the two years following the end of compulsory education. Specifically they find: of those who remain in full time education at age 16, one in two are

[3] For details of Youth Cohort Survey see: http://discover.ukdataservice.ac.uk/series/?sn=2000061

still likely to be in full education two years on; whereas for those who leave full time education at age 16, only one in 20 are likely to return to full time education two years on. They further find that prior educational achievement, controlling for other factors, is the most significant determinant of the decision to stay in education.

Crawford et al. (2011) use data from two different cohort studies in the UK (LSYPE, and the British Household Panel Survey (BHPS) [4]) and the Labour Force Survey (LFS) [5] to look at the choices made by 16/17 and 17/18 year olds. Those who stay on in education after age 16 have the highest prior educational attainments (KS2 and GCSE scores) [6]. While there does not seem to be a difference on average in the prior educational attainments of those who pursue jobs (with or without training) and those who are NEET at 16/17, those who are still NEET two years on at age 18/19 had the lowest KS2 and GCSE scores. Their findings suggest that there is a degree of persistence in NEET status: 50% of those who are NEET in the year immediately after the end of the compulsory education are NEET a year on. Those who come from socioeconomically advantaged families are the least likely to be NEETs at age 16/17.

Ryan (2001) carries out a cross national comparison of school-to-work transition (defined as the period between the end of compulsory schooling and the attainment of full-time, stable employment). He looks at seven countries: France, Germany, Japan, the Netherlands, Sweden, the UK and the US. While socioeconomic disadvantage and low educational attainment remain the key driving forces for youth inactivity and unemployment across all the seven countries, it is most acute in the UK and the US. Continental Europe sees weaker correlations, relatively, attributed to successful vocational education and apprenticeship programs (especially in Germany).

Social housing and poor neighbourhoods have been long associated with social immobility, worklessness and welfare dependency (Robinson, 2012; Stroud, 2010). Feinstein et al. (2008) look at the link between social housing and disadvantage in the UK since the Second World War. For the British Cohort Study (BCS, born in 1970) [7], they find that individuals who live in social housing are four (eleven) times more likely to be NEETs at age 18 (30) years than the rest of the cohort, after controlling for parental socioeconomic status and the individual's prior achievements.

Duckworth and Schoon (2012) use data from two UK cohort studies (LSYPE and BCS) and consider the 'protective factors' that result in some young people 'beating the odds' i.e. avoid being NEET despite unfavourable backgrounds. Their results show that prior attainment, educational aspirations and engagement with school can reduce the cumulative risk faced by a young person with

[4] For details of BHPS see: https://www.iser.essex.ac.uk/bhps.
[5] For details of LFS see the 'The LFS Users Guide Volume 1: Background and Methodology' (http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/labour-market-statistics/index.html)
[6] KS2 (Key Stage 2) exams are the national exams held at the end of primary school (age 11) in the UK; and GCSE (General Certificate of Secondary Education) is awarded after the young people take the national Key Stage 4 (KS4) exams at age 16.
[7] For details of BCS see: http://www.cls.ioe.ac.uk

multiple risk factors (low parental socioeconomic status, lone parents, social housing and workless households). [8] Yates et al. (2011) use the BCS data to discuss the role of aspirations in determining NEET status. Key findings from their work suggest that young people who have uncertain or misaligned aspirations [9] are more likely to be NEETs, especially young men from a low socioeconomic background.

Mendolia and Walker (2014) use data from the LSYPE to investigate the relationship between personality traits and the likelihood of being NEET. They use a range of regression techniques to test the robustness of the relationship between traits including the individual's ability to persevere with long term goals and the extent to which an individual believes that they can affect and control events. Across the range of techniques used they find a significant positive relationship between these traits and the individuals not being NEET.

Impact of macroeconomic conditions, especially unemployment levels, has also been considered extensively in the literature. Petrongolo et al. (2002) use micro level data from Spain and show that higher local youth unemployment rates do not determine the demand for education beyond age 16; the key determinants of post-secondary education are parental socioeconomic status. Meschi et al. (2011) use the LSYPE and consider the impact of local labour market conditions (unemployment rates and wage rates) on the choices that 16/17 year olds make. Their findings confirm the importance of past achievements of pupils, parental socioeconomic status, and aspirations as the key determinants of continued participation in schooling for 16 year olds; they also find that while young males choose to continue education in response to higher local unemployment rates, there is no significant response for young females.

2.1 Health and NEET status

There is a large literature looking at the relationship between health and educational outcomes (Perri, 1984; Currie and Hyson, 1999). While these studies do not address the outcome of NEET as such, given that prior (academic) ability is a predictor of NEET status and health has an impact on the acquisition of these abilities, then health will have an impact on NEET status indirectly.

Ding et al. (2009) use data from the Georgetown Adolescent Tobacco Research study to identify the impact of ADHD, depression and obesity on GPA scores. They find all three health conditions are significantly correlated with lower test scores for both girls and boys. The negative

[8] Educational aspirations are represented by a question asked at age 14, both of the young person and their parents, whether or not the young person would like to continue in post-compulsory education. School engagement is captured by a score generated from a series of attitudinal questions (five in BCS and 12 in LSYPE) asked of the young people at age 14.

[9] At age 16 young people were asked about what they would like to do with their lives, they could choose from a list of jobs/careers/professions varying in the degree of training and qualifications. One of the response the young people could give was 'can't decide', all those who choose this option (about 7% of the sample) are classified by the authors as having 'uncertain aspirations'. If the choice of the young person did not match their academic expectations then they were classified as having 'misaligned aspirations'.

impact of health problems and educational outcomes is supported by research undertaken by Rees and Sabia (2009). The study uses data from the National Longitudinal Study of Adolescent Health [10] to estimate the relationship between migraine headaches and educational outcomes including the individuals' high school grade point average test score. The study finds migraines have a significant negative impact on academic test scores.

Recent studies have identified an increase in the prevalence of mental health difficulties amongst adolescents (Collishaw et al., 2004; West and Sweeting, 2003). The relationship between mental health and educational attainment is bi-directional, where mental illness can potentially lead to poorer educational attainment; and poor educational attainment can potentially increase the likelihood of mental illness (Roeser et al., 1998). In an attempt to address this endogeniety Fletcher (2008) looks at the link between depression in adolescences and later educational outcomes, using the American National Longitudinal Study of Adolescent Health (Add Health), which is following children who were in 7 to 12 grade in 1994-95 [11]. They find a negative association between earlier depression and later educational outcomes only for females. Their key findings, controlling for past educational attainment and socioeconomic variables (parental education, income, ethnicity, etc.) are: females with depressive symptoms have a higher probability of dropping out of high school; conditional on graduating they are less likely to enrol in college; and if they do pursue further education they are less likely to enrol in a four-year college.

Eisenberg et al. (2009) use a randomly selected sample of students enrolled in a 'large, public, academically competitive university' in the US. The mental health of students is first assessed in 2005 and then two years later in 2007; their academic progress is monitored throughout the two years. The key findings of the study suggest that depression (contemporary and lagged) is linked with lower academic progress (as measured by GPA) and higher dropout rate, after controlling for gender, race/ethnicity, age, past achievement (as measured by the results of the entrance exam to the university) and family background (as measured by the financial situation when growing-up).

One recent study which looks at the direct impact of depression on NEET status is Cornaglia et al. (2015). They use the LSYPE and, controlling for past achievements, socioeconomic status of families, and aspirations of both the young person and their parents, find a negative association between past incidence of depression and educational outcomes, as measured by GCSE; and a positive association between past incidence of depression and probability of being NEET. [12] The results are stronger for girls.

[10] For details of Add Health see: http://www.cpc.unc.edu/projects/addhealth

[11] Depression is measured in wave 1 (1994-1995) and wave 2 (1996) when the average age of the sample is 17 and 18 years respectively. The educational outcomes are measured in wave 3 (2001-02) when the average age in the sample is 22 years. .

[12] In both cases depression is from a period before the period in which the educational outcome and NEET status is measured.

To summarise, the literature shows that NEET status is determined (amongst other things) by educational attainment, which in turn is determined by health. However, none so far have modelled the relationship across NEET status, latent cognitive ability and health simultaneously, allowing for direct and indirect effects dynamically. This paper fills this gap by building a latent dynamic model of cognitive ability formation focusing on health to explain their causal impact on NEET status.

3. Estimation Method

To understand the dynamics of cognitive ability formation up to the end of compulsory education we use the *value added model of ability formation* given by Todd and Wolpin (2007), whereby an adolescent's current ability is a function of their prior ability and a host of exogenous variables which impact on the acquisition of ability. Further, like Cunha and Heckman (2007) we assume the ability of the young person to be latent. At the end of compulsory education, the stock of cognitive ability is then used to explain the young person's post-compulsory-education outcomes, in our case NEET status, and its persistence at the following year.

For our analysis we divide the young person's life from age 11/12 to 17/18 into four time periods; $t = 1, \dots, T$, with $T = 4$, not necessarily covering a year each; at the end of each period we have some measures for the outcome of interest. The period prior to this, from birth to the end of primary school (ages 0 to 10/11 years) is denoted as $t = 0$, and gives the initial conditions. At the end of $t = 0$ we have some measures (KS2 results taken at age 11) of baseline latent cognitive ability (θ_0) of the child.

Time periods $t = 1$ and 2 cover compulsory secondary education, ages 10/11 to 16 years. For period $t = 1$, latent cognitive ability of the young person is measured at age 13/14 years by the Key Stage 3 results (KS3), the first point of assessment available for secondary school; and for period $t = 2$, the latent cognitive ability of the young person is measured at age 15/16 years, by the KS4 results (KS4-GCSE). Time periods $t = 3$ and 4 cover post-compulsory education (16 to 17/18 years): for these periods the outcome we are interested in is NEET status, in the academic year immediately after the end of compulsory education; and in the subsequent year. NEET status is observable.

We estimate the model using SEM which has two components: a structural model for the dynamic pathway of interest from cognitive ability to labour market outcomes; and a measurement model to estimate the latent factor (Cunha et al., 2010).

3.1. Structural model

Let θ_t be the stock of latent cognitive skill (ability) of the child at time t. Ability at time t, θ_t, depends on: past ability, θ_{t-1}; and exogenous covariates X_t. The dynamics for the first two time periods are given as,

$$\theta_t = \gamma_{1t}\theta_{t-1} + \gamma_{2t} X_t + \eta_t^{\theta} \quad \text{for } t = 1, \text{and } 2 \tag{1}$$

where γ_{1t} and γ_{2t} are vectors of time-varying parameters to be estimated; and η_t^{θ} is the error term, assumed to be independent across individuals and over time.

For $t = 3$ and $t = 4$ the outcome of interest is NEET status at time t, Y_t; the dynamics for $t = 3$ and t = 4 are given as:

$$Y_t = \beta_{1t}Y_{t-1} + \beta_{2t}\theta_{t-1} + \beta_{3t} X_t + \eta_t \quad \text{for } t = 3, \text{and } 4 \tag{2}$$

where β_{jt}, for $j = 1, \dots 3$, are vectors of time-varying parameters to be estimated; and η_t is the error term, assumed to be independent across individuals and over time. For $t = 3$, Y_t depends on past ability, θ_{t-1}, and exogenous covariates X_t; $\beta_{13} = 0$, as by the very definition of NEET there is no observation for Y_{t-1} for $t = 3$. Thus for $t = 3$, we get:

$$Y_3 = \beta_{23}\theta_2 + \beta_{33} X_3 + \eta_3 \tag{2a}$$

For $t = 4$, Y_t depends on past NEET status, Y_{t-1}; past ability, θ_{t-1}; and some exogenous covariates X_t. In our empirical application, the LSYPE does not have any measures for the latent ability at $t = 3$, θ_3. We therefore make an assumption that the latent ability does not change significantly between $t = 3$ and $t = 4$, and that θ_2 can be used as a good proxy for θ_3; which means instead of Y_4 as a function of θ_3, we have Y_4 as a function of θ_2. Thus for $t = 4$, we get:

$$Y_4 = \beta_{14}Y_3 + \beta_{24}\theta_2 + \beta_{34} X_4 + \eta_4 \tag{2b}$$

Covariates in vector X_t vary over time; we allow for the covariates to be both latent and observed (we discuss the covariates in detail in section 4 below). A diagrammatic representation of the structural model is given by a path diagram (Figure 1). The single headed arrows in the path diagram represent the hypothesised direct effect of one variable on another. For example, the arrow from ability in period 1 (θ_1) to ability in period 2 (θ_2) indicates that we expect θ_1 to impact upon θ_2.

3.2. Measurement Model

As cognitive ability is taken to be latent, we have a measurement model for it.

$$Z_{j,t} = \mu_{j,t} + \alpha_{j,t}\theta_t + \varepsilon_{j,t} \quad t = 0, 1, 2 \tag{3}$$

Where $Z_{j,t}$ for $j = 1, \dots, m_t$ are the measures available for the latent variables at time t. m_t are the number of measures available and, in order to enable identification, $m_t \geq 2$. $\alpha_{j,t}$ are the factor loadings, which can be interpreted as the amount of information that the measures ($Z_{j,t}$) contain about

the latent variable (θ_t). $\varepsilon_{j,t}$ are the measurement errors, these capture the difference between the observed measures and the unobserved latent variables.

For NEET status, we observe the discrete outcome (interpreted as the young person's choice), which we code as a binary variable. Random utility theory models the observed outcome variable as:

$$Y_t = 1 \text{ if } u_Y > u_A$$
$$Y_t = 0 \text{ otherwise}$$

where u_Y is the utility from being NEET and u_A is the utility from the alternative, where the alternative can be any of the following: education (full time and/or part time), training, job. Without loss of generality we can assume $u_A = 0$. The decision of the individual, in time period $t = 3$, and 4 is modelled as:

$$\begin{aligned} P(Y_t = 1) &= P\,(u_Y > 0 | \beta_{1t} Y_{t-1} + \beta_{2t}\theta_{t-1} + \beta_{3t} X_t) \\ &= P\,(\eta_t > -\beta_{1t} Y_{t-1} - \beta_{2t}\theta_{t-1} - \beta_{3t} X_t) \\ &= 1 - F(-\beta_{1t} Y_{t-1} - \beta_{2t}\theta_{t-1} - \beta_{3t} X_t) \end{aligned} \quad (4)$$

where $F(.)$ is the cumulative distribution function for the error η_t. We treat η_t as a normal distribution, and therefore estimate a probit model.

3.3. Identification

The factor loadings in equation (3) can be identified only up to a scale. It is therefore necessary to normalise them. The normalisation we use here is: $\alpha_{1,t} = 1$. In addition, for the latent variables we cannot separately identify both their mean, $E(\theta)$, and their intercept $\mu_{j,t}$. Therefore, we assume $E(\theta) = 0$ and identify $\mu_{j,t}$. Further, to be able to identify all of the parameters of interest in equations (1) to (4) we need to make the following assumptions:

Assumption 1: $\varepsilon_{j,t}$ has a mean of zero and is independent across individuals and over time for $t = 1, \ldots T$, and $j = 1, \ldots, m_t$.

Assumption 2: $\varepsilon_{j,t}$ has a mean of zero and is independent of the latent variables for $t = 1, \ldots T$, and $j = 1, \ldots, m_t$.

For the empirical analysis, to aid computation, we further assume that η_t^{θ} and $\varepsilon_{j,t}$ have a normal distribution, though this is not needed for identification.

3.4. Estimation and diagnostic statistics

One of the advantages of estimating a dynamic SEM is that we can look at both the direct and the indirect effects of one variable upon another. For example, from equation (2b) we have an estimate of

the direct effect of ability at the end of compulsory education, $t = 2$, θ_2, on NEET status at time $t = 4$, Y_4, given by the coefficient β_{24}. But we know that θ_2 also has an indirect impact on Y_4 via Y_3, given by β_{23} (equation (2a)). The total effect of θ_2 on Y_4 is given by: $\beta_{24} + \beta_{14} * \beta_{23}$ (the direct effect + the indirect effect, respectively). This requires us to estimate $\beta_{14} * \beta_{23}$ and its statistical significance. (For details on estimation see Muthen, 2011).

Estimation is undertaken using the weighted least squares (WLS) estimator in MPlus v7.3 (Muthen and Muthen, 2010). Equations (1) to (4) are estimated simultaneously. The asymptotically distribution-free WLS was chosen instead of the maximum likelihood (ML) approach because the ML approach requires the indicator variables used in equations (2)-(4) to be continuous and multivariate normal. In our application many of these indicators are either dichotomous or ordinal variables (see next section).

Since previous studies have indicated the predictors of remaining in education can vary between the genders (Fletcher, 2008 and 2010), the models are estimated separately for females and males. Further, as the majority of existing research uses probit regressions to model NEET status, we also estimate two probit models, at $t = 3$ and $t = 4$, to compare with our SEM models.

We use two diagnostic statistics to determine the goodness of model fit. Firstly, the Comparative Fit Index (CFI) which accounts for the discrepancy between the data and the hypothesized model while adjusting for sample size (Bentler, 1990). Values from the CFI vary between 0 and 1 with higher values indicating better model fit. Secondly, we use the root mean square error of approximation (RMSEA) which shows the amount of variance that is not explained by the model. Values for the RMSEA again vary between 0 and 1, but for the RMSEA lower values are indicative of a better fitting model (Steiger and Lind, 1980). We take the guidance given by Hu and Bentler (1999) that for a model to be considered to fit the data adequately it should have RMSEA < 0.06 and CFI > 0.90. In addition to using these diagnostic statistics the validity of the model will be judged on the basis of the individual parameter estimates produced. Specifically, we will consider whether they have face validity given expectations based on both wider economic theory and the findings of previous empirical studies in this area.

4. Data and measurement

4.1 Overview of the dataset

The analysis is undertaken using data from the first five waves of the LSYPE. The study follows a cohort of approximately 15,500 young people in English secondary schools. In the first wave, in 2004, participants were aged 13/14. The survey was conducted annually, and by wave 5 the individuals were aged 17/18 (Department for Education, 2013). Until the LSYPE, the only nationally representative cohort studies undertaken in relatively recent times were the British Cohort Study, which followed

individuals born in 1970, and the Millennium Cohort Study [13] whose subjects are only now starting adolescence. The LSYPE is the first national survey for many years to follow a group of English adolescents through much of their secondary education and into early adulthood (Chowdry et al., 2009 and 2010) [14]. The main aim of the study was to provide evidence on the factors central to individuals' educational progress and attainment (Department for Education, 2013).

For the first five waves the dataset contains responses from individual face to face interviews with both the young people concerned and their parents or guardians. We link the five waves of LSYPE to the four time periods and the initial conditions of our model (Figure 2 provides a visual representation of how data from the different waves of the LSYPE are linked to the five time periods). The information for initial conditions and time period 1 come from wave 1; information for time period 2 comes from waves 2 and 3; time period 3 corresponds to wave 4; and time period 4 corresponds to wave 5 of LSYPE. Additionally, for all children educated in the state sector and around a third of those educated in the independent/private sector, LSYPE is linked to the National Pupil Database (NPD), an administrative database which contains information on national examination results (KS2, KS3, KS4-GCSE) (Department for Education, 2011). We use these national examination results as measures for cognitive ability.

Over the five waves of interviews a number of individuals drop out from the study. Our analysis is undertaken on the individuals who are present in all of the five waves with the three national exam results from the NPD. Additionally, individuals are dropped if their records contain missing data for the observed indicators included in the structural model. Table 1 illustrates the effect of attrition and non-response on the availability of individuals suitable for analysis. Individuals are only included in the analysis if their records contain the required responses across all of the five waves – 5,194 individuals (2696 girls and 2498 boys) are therefore included in the analysis. The LSYPE used a stratified sampling approach; in our analysis we use the weights from wave 5 (the final wave in our analysis), as these weights take into account both the sample design and non-response bias. See the report by Anders (2012) for further details. [15]

4.2. Variables incorporated in the dynamic model

A list of all the variables along with descriptions is provided in Appendix, Table A1. The outcome variables and the covariates in each time period are shown in Figure 2. Three measurement models for

[13] For details of Millennium Cohort Study see: http://www.cls.ioe.ac.uk/mcs

[14] The Avon Longitudinal Study of Parents and Children (University of Bristol, 2013) contains the responses of young people of a similar age. However, it does not contain respondents from the whole of England; it only contains the responses of young people and their parents who live in Bristol and the surrounding area.

[15] While the longitudinal nature of the data allows for a dynamic analysis, this also imposes an important limitation. It was only possible to include individuals from the LSYPE if they responded to a number of questions across five waves of data collection and if their examination results were available from the LSYPE-NPD link. This necessitated the exclusion of approximately two thirds of the original sample. This is a common limitation of analyses based on longitudinal data. The weights used in the analysis should address some of these issues.

cognitive ability are estimated. Each attempts to model the individuals' ability at a different time period. The first model for baseline ability (θ_0) incorporates indicators from the national KS2 exams, which were undertaken when the individuals are aged 10/11, three years before being interviewed for the LSYPE. The second measurement model for θ_1 is estimated using test scores from the KS3 exams, taken when the individuals are aged 13/14 (LSYPE wave 1). The third measurement model for θ_2 is estimated with indicators based on the test scores in KS4 (GCSEs), undertaken at the end of compulsory education when the individuals are aged 15/16 (LSYPE wave 3).

From wave 4 of the LSYPE (when individuals are aged 16+) the respondents are no longer in compulsory education; the LSYPE therefore contains data on the education or labour market status of the young people: whether they are in full time education, in a job with training or without training, in training, or NEET. A binary variable (Y_t) in wave 4 ($t = 3$) and wave 5 ($t = 4$) is created from this information: the variable takes value 1 if the young person is NEET and 0 otherwise.

Mental health was measured using a latent variable approach in a manner similar to that adopted for cognitive ability. The 12-item General Health Questionnaire (GHQ-12) (Goldberg and Williams, 1988) was included in the survey when the individuals were age 14/15 (LSYPE wave 2) and also when they were aged 16/17 (LSYPE wave 4). In line with the study by Hankins (2008) confirmatory factor analyses supported the approach of modelling mental health as a single latent variable while explicitly accounting for measurement error that likely results from response bias on the negatively phrased items of the GHQ-12. Additionally, confirmatory factor analyses supported the interpretation of the GHQ-12 items as binary rather than likert variables. This interpretation of the items is consistent with the findings of Goldberg et al. (1997). They identified that for the GHQ-12 the binary method of interpretation performed better with respect to sensitivity and specificity than the likert approach when used in centres across a range of countries.

The variables available for capturing the general health of individuals vary across the waves of the survey. Three questions from when the individuals are aged 13/14 (wave 1, $t = 1$) inform variables used in the analysis. In the first question the young person's parent is asked whether their child has "any long-standing illness, disability or infirmity". The second and third questions are only asked if the parent responds yes to the first. The second question asks if the problem makes it harder for their child "to attend school or college regularly"; the third asks if the problem affects their child's ability to "do his/her school work". The responses to these questions are used to create two dummy variables (base is no health problem): health problem not affecting school work or attendance; health problem that affects school work or attendance. In year 11 (wave 3, $t = 2$) the individuals were asked "In the last 12 months would you say your health has been very good, fairly good, not very good or not good at all?". In the analysis these responses are coded from zero (for the response "very good") to 3 (for the response "not good at all"). Finally in the first year post-compulsory education (wave 4) the initial

screening question in wave 1 is asked to the young person, while the subsequent question focuses on whether the health problems "limit your daily activities in any way compared to people of your age?" rather than focusing on school work or school attendance. The responses to these questions are used to create two dummy variables (base is no health problem): health problem not affecting daily activities; health problem that affects daily activities.

Variables relating to household socioeconomic status [16] are included from when the young people are aged 13/14 (wave 1, $t = 1$), 15/16 (wave 3, $t = 2$), 16/17 (wave 4, $t = 3$) and 17/18 (wave 5, $t = 4$). These variables are based on the National Statistics socio-economic classification (NS-SEC) of the household reference person. The household reference person is the person who owns or rents the property the young person lives in. If the property is jointly owned or rented then it is the parent with the highest income (Department for Education, 2013). In the wave of interviews undertaken the second year after compulsory education (wave 5) the household reference person is not explicitly listed. Therefore the variable identifying the household socioeconomic status of the family in this period takes the value of the worker in the household with the highest occupational classification.

The aspiration of the child is incorporated in the analysis by including their response to the question "How likely do you think it is that you will ever apply to go to university to do a degree?". Their responses to this question when they are aged 14/15 (wave 2, $t = 2$) and 15/16 (wave 3, $t = 2$) are included in the analysis. Two variables relating to the parental aspirations for the child are also included. In the first question the parent is asked what they think their child will do "when he/she reaches 16 and can leave school"; in the second they are asked what they themselves would like their child to do when they reach this same stage. The parent's responses to these questions when their child is aged 14/15 and 15/16 are included in the analysis as predictions and preferences. The variables relating to the child's and parent's aspirations are likert variables – they are described in greater detail in Appendix, Table A1.

Where variables are incorporated in the structural model across multiple time periods, we take the general approach to include the control variable that is from the period contemporaneous to the outcome variable. This general approach is altered for the control variables relating to the child's mental health and the aspirations of the child and parent. Arguably these variables at time t could be affected by the young person's outcome of interest: test results or their education-labour market choice at time t. In order to reduce the risk of reverse causality these control variables are taken from the time period prior to the outcome variable.

[16] Ideally we would have liked to use household income. However, household income is a variable with particularly high number of missing observations, and if this variable were included in the analysis only 2,853 individuals would be available for inclusion. In order to avoid omitting such a high proportion of individuals alternative variables which are highly correlated with household income, such as parental education levels and family socioeconomic occupational class, are incorporated into the analysis as controls.

Other control variables incorporated into the analysis include: the mother's education; the parental rating of the child's school and of the teachers; the child's birth weight; the number of weeks early the child was born; the child's ethnicity; and the month of the year in which the child is born. Apart from the young person's ethnicity these variables are only included as initial conditions. Ethnicity is included again in the regression which explores the predictors of NEET status in the first year after compulsory education; this is because ethnicity may have an independent effect on their choice over education-labour market if racial discrimination is present within the labour market, or perceived to be so by the young person.

Table 2 presents the weighted summary statistics for the sample. It is perhaps worth noting that there appears to be a large degree of similarity between the sample of girls and boys with respect to control variables such as household socioeconomic status, mother's highest qualification and the parent's rating of the child's school and teachers. As a group, however, the girls' mean points in their GCSE exams are higher than the boys in the sample; similarly a larger proportion of girls achieve C or higher in their GCSE English exam. At the end of compulsory education, 4% of the girls and 6% of the boys in our sample are NEETs. A year later this number increases to 6% for girls and 10% for boys.

5. Results

5.1. Probit models for NEET status

The literature most commonly attempts to identify the predictors of NEET status in the first year post-compulsory education; following which Table 3 gives the estimated coefficients from the static probit specification when the dependent variable is NEET status in $t = 3$, the first year after compulsory education. Table 4 shows the probit results when the dependent variable is NEET status in $t = 4$. In both probit models we control for the socioeconomic background of the person, their past ability and aspirations. In the probit for $t = 4$ we also control for their NEET status in the previous period. In both Table 3 and Table 4 we report standardized coefficients for the continuous control variables.

The key predictors of NEET status in the period immediately after compulsory education, for both girls and boys are: past ability, their own aspirations and their parent's expectations. Higher ability significantly lowers the probability of being NEET; and the young people who state they are likely to apply to go to university are significantly less likely to be NEET. In contrast, for both girls and boys, the variables relating to general health fail to achieve significance at thresholds of 5% or below. At the cohort level the impact of mental health difficulties differs between the genders. For girls past mental health difficulties are significantly linked with an increased risk of being NEET; while for boys there is no significant correlation.

Regarding NEET status a year after the end of compulsory education (Table 4), in line with the literature, past ability remains a significant predicator of NEET status at age 17/18. We also find evidence of persistence in NEET status: controlling for past ability, a young person who is NEET at age 16/17 is more likely to be NEET a year on at age 17/18. Young people (both girls and boys) from deprived areas are more likely to be NEET at age 17/18; although a lower household SES is a predicator of NEET status only for girls. Similarly to the results for the previous time period, for both girls and boys, the variables relating to general health fail to achieve significance. Past mental health difficulties is again significantly correlated with NEET status for girls but not for boys.

For the probit models the diagnostic statistics indicate a good fit when fitted to the data for girls across both time periods. When fitted to the data for boys the CFI and RMSEA meet the suggested thresholds for the $t = 3$ probit (Table 3) however for the $t = 4$ probit (Table 4) the CFI statistic of 0.89 is slightly below the recommended threshold of 0.90.

5.2. Dynamic SEM

We next present the results (Table 5) from our dynamic model, starting from the results relating to ability formation in periods $t = 1$ and $t = 2$. Past ability is significant in both equations for both girls and boys; we therefore have evidence supporting the "self-productivity" [17] of this skill (Heckman and Masterov, 2007). Illness that affects schooling ($t = 1$) and poor self-reported health ($t = 2$) are significantly correlated with lower levels of ability in the contemporaneous time period. For girls past mental health would also appear to be an important predictor of ability ($t = 2$); for boys past mental health is not significant. Lower socioeconomic status ($t = 1, 2$), a deprived neighbourhood ($t = 2$) and non-white ethnicity ($t = 1$) all appear to have a negative impact on ability formation. On the other hand, the prior aspirations of the young person and the predictions of their parent ($t = 2$) are positively correlated with ability.

For NEET status in the period immediately after compulsory education ($t = 3$), it can be seen that prior ability has a negative impact on the probability of being NEET. The aspiration of the young person to remain in education in the previous period is correlated with a significant reduction in the probability that they will be NEET. For girls, neither the parent's predictions or preferences are correlated with a significant reduction in the probability that the young person will be NEET. For boys, the predictions are, although the preferences are not . For both girls and boys general health in the same time period has no impact on NEET status.

For both girls and boys, similarly to the results from the probit model, NEET status in time period 4 ($t = 4$) is significantly correlated with NEET status in period 3 ($t = 3$) indicating a degree of persistence. However, the coefficient is much lower than in the corresponding probit specification; in

[17] Self-productivity exists when higher ability at time $t - 1$ is associated with higher ability at time t.

the static probit model (Table 4) the coefficient is 0.69 for girls and 0.85 for boys, compared with the coefficient of 0.34 for girls and 0.43 for boys in the dynamic model (Table 5).

In the dynamic model, for girls, after controlling for NEET status at age 16/17, past ability maintains a negative impact on NEET status, while for boys, past ability is not a significant predictor of NEET status at age 17/18. For girls past mental health predicts NEET status; for boys the correlation fails to achieve significance at thresholds of 5% or below.

The model fit diagnostic statistics indicate that the SEM fits the data from both genders adequately; CFI is well above the recommended level of 0.90 and RMSEA is below 0.05, as recommended. Additionally the individual parameter estimates reported appear to have face validity.

5.3. Direct and indirect effects

Table 6 and Table 7 present the direct and indirect effects (as explained in section 3.4) of a number of variables on NEET status in time period 3 and 4 respectively. While the indirect effects of all the variables in the model on NEET status can be calculated we focus here only on the indirect effects of health and ability on NEET status.

Ability at age 16 (as measured by the GCSE exams) has a direct impact on NEET status immediately after compulsory education at age 16/17. However, ability at age 16 depends on past ability (as measured by KS2 and KS3), and thus past ability also has an indirect impact on NEET status. We find evidence of this indirect effect of past ability, right from the end of the primary education, being significant (Table 6, first three rows).

Table 7 illustrates that for girls ability at age 16 has significant direct and indirect effect on the NEET status at age 17/18, even after controlling for the past NEET status at age 16/17. For boys, controlling for past NEET status at age 16/17, past ability (measured at age 16, $t = 2$) has no direct impact on the continuation of NEET status at age 17/18; however, past ability still has a significant indirect effect on NEET status (Table 7, first three rows).

As might be expected, given the strong correlation between both prior ability and future NEET status for both girls and boys difficulties with general health in period 1 and 2 are indirectly but significantly correlated with an increased probability of being NEET in period 3 and period 4. For girls mental health difficulties in period 2 are also significantly correlated with an increased probability of being NEET in period 3 and period 4; for boys the correlation is insignificant.

6. Concluding discussion

Ensuring young people start their adult lives in education, employment or training benefits both the individuals concerned and society as a whole. The longitudinal nature of the data (LSYPE) and the SEM approach used in this paper allows for a dynamic analysis of the predictors of NEET status. To

our knowledge this is a first study which puts together various determinants of NEET status, notably cognitive ability and health (both general and mental health) within a single framework, and addresses the issue of measurement error in both ability and mental health by using a latent factor model. In contrast to previous studies using probit analysis, the incorporation of a dynamic analysis allows us not only to investigate the relative importance of the different determinants of NEET status, but it additionally allows both an analysis of the period in which these different determinants have their greatest impact and the pathway through which their cumulative impact is realised.

The analysis in this paper indicates that academic ability plays a substantial role in protecting or exposing individuals to the risk of being NEET. We are able to further our understanding of how factors in earlier adolescence influence individuals' formation of cognitive and therefore indirectly affect their probability of being NEET in early adulthood. Consistent with the literature we identify that the mother's education, the individual's ethnicity and the early aspirations of both the young person and their parent are key predictors of an individual's NEET status in early adulthood. Our findings suggest that these factors influence future NEET status through the pathway of cognitive ability formation.

There is evidence to suggest that general health affects girls' and boys' accumulation of ability; and the impact of general health difficulties on NEET status is indirect i.e. it impacts NEET status only via ability. Mental health is an important predicator of NEET status for girls; both directly and indirectly via ability. For boys, however, mental health has no impact on ability formation, nor does it have a significant direct impact on NEET status. Further, while past ability is an important direct predictor of persistence in NEET status for girls, for boys past ability only has an indirect effect on the persistence of NEET status.

The paper's findings have important policy implications. As noted there is a strong relationship between both prior academic ability and future academic ability, and between prior academic ability and future NEET status. Policy makers aiming to minimise the number of young people who start their working life in unemployment may do well to consider how they can best help young people develop their academic ability throughout adolescence. To some extent it could be argued that this process starts even earlier, in the early childhood of the individual. Our analysis shows, however, that when pre-adolescent academic ability is controlled for, other influences on the individual's further accumulation of ability remain significant. These include factors such as their health, the mother's education, the quality of their school and the deprivation of their local area. Interventions in early childhood and late adolescence may both have a role. The research presented supports the notion that individuals' learning opportunities need to be protected and facilitated throughout their adolescence if a society aims for all to be able to engage in its labour market.

References

Anders, J. (2012) Using the Longitudinal Study of Young People in England for research into Higher Education access. Department of Quantitative Social Science Working Paper No . 12-13 December 2012 (No. 12-13). London.

Bentler, P. M. (1990) Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246.

Bell, D. N., & Blanchflower, D. G. (2011) Young people and the Great Recession. *Oxford Review of Economic Policy*, 27(2), 241-267.

Card, D. (1999) The causal effect of education on earnings. *Handbook of labor economics*, 3, 1801-1863.

Card, D. (2001) Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5), 1127-1160.

Chowdry, H., Crawford, C., & Goodman, A. (2009) *Drivers and Barriers to Educational Success: Evidence from the Longitudinal Study of Young* (pp. 1–123). London.

Chowdry, H., Crawford, C., & Goodman, A. (2010) *The role of attitudes and behaviours in explaining socio-economic differences in attainment at age 16* (No. 10, 15). London.

Coles, B., Godfrey, C., Keung, A., Parrot, S. and Bradshaw, J. (2010) Estimating the life-time cost of NEET: 16-18 year olds not in Education, Employment or Training. Research conducted for the Audit Commission, July 2010. http://www.york.ac.uk/spsw/research/neet

Collishaw, S., Maughan, B., Goodman, R., & Pickles, A. (2004) Time trends in adolescent mental health. *Journal of Child Psychology and psychiatry*, 45(8), 1350–1362.

Cornaglia, F., Crivellaro, E. & Mcnally, S. (2015) Mental Health and Education Decisions. *Labour Economics*, 33, 1-12.

Currie, J., & Hyson, R. (1999) Is the Impact of Health Shocks Cushioned by Socioeconomic Status? The Case of Low Birthweight. *American Economic Review*, 89(2), 245-250.

Crawford, C., Duckworth, K., Vignoles, A., & Wyness, G. (2011) Young people's education and labour market choices aged 16/17 to 18/19. London: Department of Education.

Cunha, F., & Heckman, J. (2007) The technology of skill foramtion. *American Economic Review*, 97(2): 31-47.

Cunha, F., Heckman, J., & Schennach, S. (2010) Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, *78*(3), 883–931.

Department for Education. (2011) *LSYPE User Guide to the Datasets : Wave 1 to Wave 7* (pp. 1–103). Retrieved from http://www.esds.ac.uk/doc/5545/mrdoc/pdf/5545lsype_user_guide_wave_1_to_wave_7.pdf

Department for Education. (2013) About LSYPE. Retrieved from https://www.education.gov.uk/ilsype/workspaces/public/wiki/Welcome/LSYPE

Dickerson, A. W., & Jones, P. (2004) *Estimating the Impact of a Minimum Wage on the Labour Market Behaviour of 16 and 17 Year Olds*.

Ding, W., Lehrer, S. F., Rosenquist, J. N., & Audrain-McGovern, J. (2009). The impact of poor health on academic performance: New evidence using genetic markers. *Journal of Health Economics*, *28*(3), 578-597.

Duckworth, K., & Schoon, I. (2012). Beating the Odds: Exploring the Impact of Social Risk on Young People's School-to-Work Transitions during Recession in the UK. *National Institute Economic Review*, *222*(1), R38-R51.

Eisenberg, D., Golberstein, E., & Hunt, J. B. (2009) Mental Health and Academic Success in College. *The B . E . Journal of Economic Analysis & Policy*, 9(1), 1-37.

Feinstein, L., Lupton, R., Hammond, C., Mujtaba, T., Salter, E., & Sorhaindo, A. (2008) *The public value of social housing: a longitudinal analysis of the relationship between housing and life chances*.

Fletcher, J. (2008) Adolescent depression: diagnosis, treatment, and educational attainment. *Health Economics*, 1235 (December 2007), 1215–1235.

Fletcher, J. (2010) Adolescent depression and educational attainment: results using sibling fixed effects. *Health Economics*, 871 (July 2009), 855–871.

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997) The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine*, *27*(1), 191–7

Goldberg, D. P., & Williams, P. (1988) *A user's guide to the General Health Questionnaire*. (D. P. Goldberg & P. Williams, Eds.). Windsor: NFER-Nelson.

Gregg, P., & Tominey, E. (2005) The wage scar from male youth unemployment. *Labour Economics*, 12(4), 487-509.

Hankins, M. (2008) The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing? *Clinical Practice and Epidemiology in Mental Health*, *4*(10), 1–8.

Heckman, J., & Masterov, D. V. (2007) The Productivity Argument for Investing in Young Children. *Review of Agricultural Economics*, 29(3), 446–493.

Hu, L., & Bentler, P. M. (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.

Mendolia, S., & Walker, I. (2014). Do NEETs Need Grit? IZA Discussion Paper No. 8740.

Meschi, E., Swaffield, J., & Vignoles, A. (2011). The Relative Importance of Local Labour Market Conditions and Pupil Attainment on Post-Compulsory Schooling Decisions. IZA Discussion Paper No. 6143.

Mroz, T. A., & Savage, T. H. (2006). The long-term effects of youth unemployment. *Journal of Human Resources*, 41(2), 259-293.

Muthén, L.K. & Muthén, B.O. (2010). Mplus User.s Guide, Sixth Edition. Los Angeles, CA: Muthén & Muthén.

Muthén, B. (2011). Applications of causally defined direct and indirect effects in mediation analysis using SEM in Mplus. Unpublished working paper, www. statmodel. com.

OECD (2014) *Education at a Glance 2014: OECD Indicators* (www.oecd.org/edu/eag.htm).

OECD (2013) *Education at a Galnce 2013: OECD Indicators* (www.**oecd**.org/edu/eag2013%20(eng)--FINAL%2020%20June%202013.pdf)

Oreopoulos, P., & Salvanes, K. G. (2011). Priceless: The nonpecuniary benefits of schooling. *The Journal of Economic Perspectives*, 159-184.

Petrongolo, B., & San Segundo, M. J. (2002). Staying-on at school at 16: the impact of labor market conditions in Spain. *Economics of Education Review*, 21(4), 353–365.

Perri, T. J. (1984). Health status and schooling decisions of young men. *Economics of Education Review*, 3(3), 207-213.

Rees, D. I., & Sabia, J. J. (2009). The Effect of Migraine Headache on Educational Attainment. *The Journal of Human Resources*, 46(2), 317–332.

Robinson, D. (2012). Social Housing in England: Testing the Logics of Reform. *Urban Studies*, 50(8), 1489–1504.

Roeser, R. W., Eccles, J. S., & Strobel, K. R. (1998). Linking the study of schooling and mental health: Selected issues and empirical illustrations at.. *Educational Psychologist*, 33(4), 153–176.

Ryan, P. (2001). The School-to-Work Transition: A Cross-National Perspective. *Journal of Economic Literature*, 39(1), 34–92.

Steiger, J., & Lind, J. (1980). Statistically based tests for the number of common factors. *Annual Meeting of the Psychometric Society, Iowa City*.

Stroud, P. (2010). *Social housing: a launchpad, not a destination*. York.

Todd, P. E., & Wolpin, K. I. (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human capital*, *1*(1), 91-136.

University of Bristol. (2013). Avon Longitudinal Study of Parents and Children. Retrieved from http://www.bristol.ac.uk/alspac/participants/

West, P., & Sweeting, H. (2003). Fifteen, female and stressed: Changing patterns of psychological distress over time. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 44, 399–411.

Yates, S., Harris, A., Sabates, R., & Staff, J. (2011). Early occupational aspirations and fractured transitions: a study of entry into 'NEET'status in the UK. *Journal of Social Policy*, 40(03), 513-534.

Table 1: Attrition and missing variables in the LSYPE

Wave (year)	Age of YP	Total Number of YP Interviewed at each Wave	Number of YP Remaining from Wave 1 – 5*	Number of YP at each Wave with the Required Data for Inclusion+
1 (2004)	13/14 years	15,770	15,770	10,508
2 (2005)	14/15 years	13,539	13,539	7,609
3 (2006)	15/16 years	12,439	12,437	6,809
4 (2007)	16/17 years	11,801	11,425	5,967
5 (2008)	17/18 years	10,430	10,158	5,194

Notes:
*The number of young people (YP) interviewed in a given wave who were also present for all of the previous waves.
+ The number of young people present in a given wave and in each of the previous waves, with no missing data.

Table 2: Descriptive Statistics

Variable	Girls (n=2696)		Boys (n=2498)	
	mean	S.D.*	mean	S.D.*
t = 0 (LSYE Wave 1)				
KS2 Score: English	28.20	3.56	27.00	4.03
KS2 Score: Maths	27.36	4.43	28.00	4.61
KS2 Score: Science	29.06	3.24	29.20	3.32
Birth weight	3.26	0.56	3.43	0.59
Weeks born early	0.78	1.70	0.74	1.56
Mother's Highest Qualification	3.17	1.85	3.25	1.86
Non-white	0.12	-	0.12	-
School year month	6.36	3.42	6.32	3.47
t = 1 (LSYPE Wave 1)				
KS3 Score: English	35.93	5.35	34.10	5.78
KS3 Score: Maths	37.68	7.05	38.22	7.38
KS3 Score: Science	35.11	6.01	35.27	6.25
Illness not affecting schooling	0.07	-	0.09	-
Illness affecting schooling	0.04	-	0.06	-
Parental rating of school	3.34	0.74	3.34	0.73
Parental rating of teachers	2.34	0.67	2.30	0.69
Household Socioeconomic Status	4.36	2.11	4.37	2.08
t = 2 (LSYPE Wave 2)				
GHQ-12				
Recently lost sleep	0.26	-	0.13	-
Recently under strain	0.35	-	0.22	-
Recent difficulties	0.25	-	0.14	-
Recently felt unhappy	0.32	-	0.15	-
Recently losing confidence	0.25	-	0.11	-
Recently felt worthless	0.16	-	0.07	-
Recently able to concentrate	0.18	-	0.10	-
Recently not useful	0.09	-	0.07	-
Recently made decisions	0.07	-	0.03	-
Recently enjoyed activities	0.12	-	0.07	-
Recently faced up to problems	0.10	-	0.05	-
Recently felt happy	0.14	-	0.06	-
University plans	3.02	1.10	2.74	1.19
Parent thinks YP will do	2.75	0.71	2.57	0.84

Variable	Girls (n=2696)		Boys (n=2498)	
	mean	**S.D.***	**mean**	**S.D.***
Parent would like YP to do	2.84	0.54	2.75	0.59
Local Index of Multiple Deprivation	20.44	15.57	20.43	15.75
t = 2 (LSYPE Wave 3)				
GCSE Points	432.74	135.22	405.18	143.54
GCSE English	0.77	-	0.64	-
GCSE Maths	0.67	-	0.67	-
Young person's health	0.45	0.58	0.36	0.52
University plans	2.92	1.35	2.54	1.45
Parent thinks YP will do	2.84	0.55	2.69	0.69
Parent would like YP to do	2.88	0.47	2.76	0.56
Household Socioeconomic Status	4.04	2.41	4.03	2.42
t = 3 (LSYPE Wave 4)				
NEET[$]	0.04	-	0.06	-
GHQ-12				
Recently lost sleep	0.31	-	0.17	-
Recently under strain	0.42	-	0.30	-
Recent difficulties	0.27	-	0.17	-
Recently felt unhappy	0.34	-	0.19	-
Recently losing confidence	0.26	-	0.14	-
Recently felt worthless	0.15	-	0.08	-
Recently able to concentrate	0.20	-	0.12	-
Recently not useful	0.12	-	0.09	-
Recently made decisions	0.10	-	0.04	-
Recently enjoyed activities	0.17	-	0.13	-
Recently faced up to problems	0.14	-	0.06	-
Recently felt happy	0.17	-	0.09	-
Illness not affecting daily activates	0.04	-	0.05	-
Illness affecting daily activates	0.03	-	0.02	-
Household Socioeconomic Status	3.75	2.41	3.81	2.37
t = 4 (LSYPE Wave 5)				
NEET[$]	0.06	-	0.10	-
Household Socioeconomic Status	4.23	2.31	4.36	2.27

Notes:
$ "Not in Education, Employment or Training"
*Standard deviations are only reported for continuous, integer and ordered categorical variables

Table 3: NEET status a year after compulsory education (age 16/17 years): Probit Model

Variable	Girls (n=2696)	Boys (n=2498)
Latent Variables		
Past Ability	-0.24***	-0.30***
Past Mental Health Difficulties	0.12**	-0.02
Observed Variables – contemporaneous		
Illness: not affecting daily activates	-0.16	-0.40
Illness: affecting daily activates	0.37*	0.16
Household Socioeconomic Status	-0.20***	-0.02
Observed Variables – past		
University plans	-0.27***	-0.29***
Parent thinks YP will do	-0.12**	-0.13**
Parent would like YP to do	0.01	-0.01
Observed Variables – time invariant		
Ethnicity	-0.19	-0.02
Mother's Highest Qualification	-0.05	-0.06
Local Index of Multiple Deprivation	0.08	0.09*
Model Fit Statistics		
CFI	0.95	0.91
RMSEA	0.04	0.04

Notes:
*Significant at 10%; **significant at 5%; ***significant at 1%.
Table reports the standardised coefficients for the continuous control variables ($\beta^* = \beta(\frac{\sigma_x}{\sigma_y})$)

Table 4: NEET status two years after compulsory education (age 17/8 years): Probit Model

Variable	Girls (n=2696)	Boys (n=2498)
Latent Variables		
Past Ability	-0.22***	-0.14***
Past Mental Health Difficulties	0.13***	-0.04
Observed Variables - contemporaneous		
Household Socioeconomic Status	-0.15***	-0.06
Observed Variables - past		
NEET: t=3	0.69***	0.85***
Illness: not affecting daily activates	-0.27	0.28
Illness: affecting daily activates	-0.58	0.15
University plans	-0.07	-0.02
Parent thinks YP will do	0.01	-0.03
Parent would like YP to do	-0.01	0.03
Observed Variables – time invariant		
Ethnicity	0.15	-0.06
Mother's Highest Qualification	0.02	0.01
Local Index of Multiple Deprivation	0.08*	0.15***
Model Fit Statistics		
CFI	0.94	0.89
RMSEA	0.04	0.05

Notes:
*Significant at 10%; **significant at 5%; ***significant at 1%.
Table reports the standardised coefficients for the continuous control variables, $\beta^* = \beta(\frac{\sigma_x}{\sigma_y})$

Table 5: Structural Equation Model Results

Dependent Variable	Explanatory Variable	Girls	Boys
Ability (t=1)	**Latent Variables**		
	Past Ability	0.84***	0.83***
	Observed Variables - contemporaneous		
	Illness: not affecting schooling	-0.15*	0.08
	Illness: affecting schooling	-0.66***	-0.66***
	Parental rating of school	0.09***	0.09***
	Parental rating of teachers	0.13***	0.05*
	Household Socioeconomic Status	0.11***	0.11***
	Observed Variables – initial conditions		
	Birth weight	0.00	0.04*
	Weeks born early	0.01	0.03
	School year month	0.10***	0.10***
	Mother's Highest Qualification	0.15***	0.17***
	Ethnicity	-0.21***	-0.53***
Ability (t=2)	**Latent Variables**		
	Past Ability	0.80***	0.81***
	Past Mental Health Difficulties	-0.10***	-0.01
	Observed Variables - contemporaneous		
	Young person's health	-0.06**	-0.11***
	University plans	0.12***	0.11***
	Parent thinks YP will do	0.11***	0.10***
	Parent would like YP to do	0.01	0.05*
	Household Socioeconomic Status	0.11***	0.07*
	Local Index of Multiple Deprivation	-0.20***	-0.12***
NEET (t=3)	**Latent Variables**		
	Past Ability	-0.25***	-0.23***
	Observed Variables - contemporaneous		
	Illness not affecting daily activates	-0.39	-0.15
	Illness affecting daily activates	0.08	0.05
	University plans	-0.24***	-0.26***
	Parent thinks YP will do	-0.10*	-0.13***
	Parent would like YP to do	-0.02	0.02
	Household Socioeconomic Status	-0.09	0.04
	Observed Variable – time invariant		
	Ethnicity	-0.28	-0.18
NEET (t=4)	**Latent Variables**		
	Past Ability	-0.14***	-0.06
	Past Mental Health Difficulties	0.15***	-0.08*
	Observed Variable - past		
	NEET	0.34***	0.43***
	Observed Variable - contemporaneous		
	Household Socioeconomic Status	-0.15**	-0.03
	Observed Variable – time invariant		
	Ethnicity (t=0)	0.21*	-0.09
Model Fit Statistics			
CFI		0.95	0.92
RMSEA		0.02	0.02

Notes:

*Significant at 10%; **significant at 5%; ***significant at 1%.

Table reports the standardised coefficients for the continuous control variables $\beta^* = \beta(\frac{\sigma_x}{\sigma_y})$

Table 6: Direct and Indirect effects of ability and health on NEET Status at age 16/17 (t = 3)

Variable	Girls			Boys		
	Total effect	Direct effect	Indirect effect	Total effect	Direct effect	Indirect effect
Ability						
Ability (t = 0)	-0.17***	-	-0.17***	-0.16***	-	-0.16***
Ability (t = 1)	-0.20***	-	-0.20***	-0.19***	-	-0.19***
Ability (t = 2)	-0.25***	-0.25***	-	-0.24***	-0.24***	-
General Health						
Illness: not affecting schooling (t =1)	0.03	-	0.03	-0.02	-	-0.02
Illness: affecting schooling (t =1)	0.13***	-	0.13***	0.13***	-	0.13***
Young person's health (t =2)	0.02**	-	0.02**	0.03***	-	0.03***
Illness: not affecting daily activates (t=3)	-0.40	-0.40	-	-0.15	-0.15	-
Illness: affecting daily activates (t=3)	0.08	0.08	-	0.05	0.05	-
Meatal Health						
Mental Health Difficulties (t=2)	0.03***	-	0.03***	0.00	-	0.00

Notes: *Significant at 10%; **significant at 5%; ***significant at 1%.

Table 7: Direct and Indirect effects of ability and health on NEET Status at age 17/18 (t = 4)

Variable	Girls			Boys		
	Total effect	Direct effect	Indirect effect	Total effect	Direct effect	Indirect effect
Ability						
Ability (t =0)	-0.15***	-	-0.15***	-0.11***	-	-0.11***
Ability (t =1)	-0.18***	-	-0.18***	-0.13***	-	-0.13***
Ability (t =2)	-0.22***	-0.14***	-0.08***	-0.16***	-0.06	-0.10***
Health						
Illness: not affecting schooling (t=1)	0.03	-	0.03	-0.01	-	-0.01
Illness: affecting schooling (t=1)	0.12***	-	0.13***	0.08***	-	0.08***
Young person's health (t=2)	0.02**	-	0.02**	0.02***	-	0.02***
Illness: not affecting daily activates (t=3)	-0.14	-	-0.14	-0.06	-	-0.06
Illness: affecting daily activates (t=3)	0.03	-	0.03	0.02	-	0.02
Meatal Health						
Mental Health Difficulties (t=2)	0.02***	-	0.02***	0.00	-	0.00
Mental Health Difficulties (t=3)	0.15***	0.15***		-0.08*	-0.08*	

Notes: *Significant at 10%; **significant at 5%; ***significant at 1%.

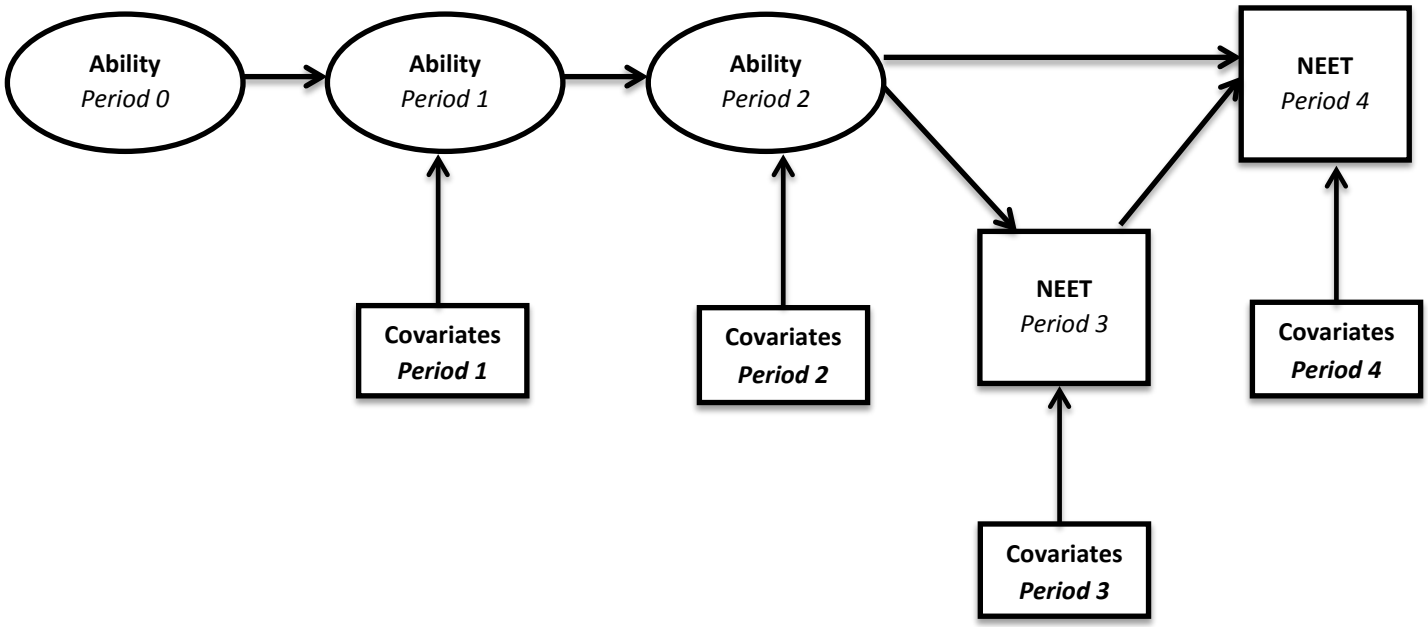
Appendix

Table A1: Variable Description

Variable (abbreviation)	Wave	Description
Outcome Variables		
English SATs score (English)	Prior to LSYPE	Continuous variable. Higher score indicates better test results. Maximum score possible is 39.
Maths SATs score (Maths)		
Science SATs score (Science)		
English SATs score (English)	1	Continuous variable. Higher score indicates better test results. Maximum score possible is 57.
Maths SATs score (Maths)		
Science SATs score (Science)		
GCSE points score (GCSE points)	3	Continuous variable. Higher score indicates better test results. Maximum varies according to the number of GCSE's: 12 GCSE's, all at A*, is equivalent to 696 points.
GCSE English C or above (GCSE English)	3	Binary Variable (0-1). 1 = if passed.
GCSE Maths C or above (GCSE Maths)		
Not in Education Employment or Training (NEET)	4, 5	Binary Variable (0-1). 1 = if not in education, employment or training.
Covariates		
Child's birth weight (Birth weight)	1	Continuous variable. Birth weight in kilograms.
Number of weeks the child was born early (Weeks born early)	1	Integer variable.
Household Socioeconomic Status	1, 3, 4, 5	Ordered categorical variable (0-7). Higher Managerial and professional is 7.
School Year Month	1	Ordered categorical variable (1-12). 1 is August 1990, 12 September 1989.
Mother's Highest Qualification	1	Ordered categorical variable (0-6). Having a degree or higher is 6 Also coded as 0 if Mother missing from household.
Parent's rating of the young person's school (Parental rating of school)	1	Ordered categorical variable: (0-4). 0 if parent very dissatisfied, 4 if very satisfied.
Parent's satisfaction with teachers (Parental rating of teachers)	1	Satisfaction by how much interest the teachers show in the young person. Ordered categorical variable: (0-3). 0 if rated as very bad, 3 if rated as very good.
Young person's ethnicity (Ethnicity)	1	Binary variable (0-1). 0 if the individual is white. 1 if the individual is not white.
Local Index of Multiple Deprivation (as above)	3	Higher score indicates greater level of local area deprivation.
Mental Health (GHQ-12)	2, 4	
All 12 questions are coded as binary variables (0-1, the value of 0 is given when the respondent replies with one of the two more positive responses to the question, the value of 1 is given when the respondent replies with one of the two more negative responses)		
Whether recently lost sleep over worry (Recently lost sleep)		
Whether recently felt constantly under strain (Recently under strain)		
Whether recently felt couldn't overcome difficulties (Recent difficulties)		
Whether recently felt unhappy and depressed (Recently felt unhappy)		
Whether recently been losing confidence (Recently losing confidence)		
Whether recently been thinking of themselves as a worthless person (Recently felt worthless)		
Whether recently been able to concentrate (Recently able to concentrate)		
Whether recently felt they have played a useful part in things (Recently not useful)		
Whether recently felt capable of making decisions about things (Recently made decisions)		
Whether recently been able to enjoy day to day activities (Recently enjoyed activities)		
Whether recently been able to face up to problems (Recently faced up to problems)		
Whether recently been feeling reasonably happy (Recently felt happy)		

Variable (abbreviation)	Wave	Description
General Health		
Health problem or disability not affecting schooling (Illness not affecting schooling)	1	Binary Variable (0-1). 0 if no health problem or problem affects schooling; 1 if health problem does not affect schooling.
Health problem or disability affecting schooling (Illness affecting schooling)	1	Binary Variable (0-1). 0 if no health problem or disability that affects schooling; 1 if health problem or disability does affect schooling.
Young Person's Self-Assessed Health (Young person's health)	3	Ordered categorical Variable: (0-3). 0 if very good, 3 if not good at all.
Health problem or disability not affecting daily activities (Illness not affecting daily activates)	4	Binary Variable (0-1):0 if no health problem or problem affects daily activities; 1 if health problem does not affect daily activities.
Health problem or disability affecting daily activities (Illness affecting daily activates)	4	Binary Variable (0-1). 0 if no health problem or disability that affects daily activities; 1 if health problem or disability that does affect daily activities
Aspirations		
Likelihood of young person applying to university (University plans)	2, 3	Whether the young person thinks it is likely they will apply to university to do a degree. Ordered categorical variable (0-4). 0 if they think it is very unlikely, 4 if they think it is very likely.
What the parent thinks the young person will do after compulsory education (Parent thinks YP will do)	2, 3	Ordered categorical variable (0-3). 0 if they think something other than training, education or an apprenticeship, 3 if they think they will continue in full time education.
What the parent would like the young person to do after compulsory education (Parent would like YP to do)	2, 3	Ordered categorical variable (0-3). 0 if they would like them to do something other than training, education or an apprenticeship, 3 if they would like them to continue in full time education.

Figure 1: Diagrammatic representation of the SEM



Notes: Ability (shown in ovals) is treated as a latent variable. Ability in period 0 represents the initial conditions. NEET status is treated as observed. Set of covariates used vary with time, we also allow for the covariates to be observed or latent. Arrows indicate the hypothesised pathways: ability in period 2 has a direct impact on NEET status in period 3; similarly ability in period 1 has a direct impact on ability in period 2 and ability in period 1 has an indirect impact on NEET status in period 3 via ability in period 2.

Figure 2: Data available across the different waves of the LSYPE

Time	**t = 0**	**t = 1**	**t = 2**		**t = 3**	**t = 4**
Age	0 to 11 years	11 to 13/14 years	13/14 to 16/17 years		16/17 years	17/18 years
LSYPE wave	Wave 1	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Covariates	Birth weight Weeks born early School year month Mothers education Ethnicity	Physical Health Parental rating of school and teachers Household SES	Mental health Aspirations Local index of multiple deprivation	Physical health Aspirations Household SES	Mental health Physical health Household SES	Household SES
Outcome	KS2 (age 11) (NPD)	KS3 (age 13/14) (NPD)	KS4-GCSE (age 16/17) (NPD)		NEET	NEET

Notes: SES = Socio economic Status; NPD = National Pupil Database; KS2 = Key Stage 2; KS3 = Key Stage 3; KS4 = Key stage 4; GCSE = General Certificate of Secondary Education; NEET = Not in Education, Employment, or Training.