National Centre for Research Methods Working Paper

6/15

A reassessment of socio-economic gradients in child cognitive development using Growth Mixture Models

Katy Sindall, Patrick Sturgis, Fiona Steele, George Leckie, Rob French

NiCRM
National Centre for
Research Methods

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

A reassessment of socio-economic gradients in child cognitive development using Growth Mixture Models

Katy Sindall, *Department of Sociology, Social Policy, and Criminology, University of Southampton*

Patrick Sturgis, *Department of Social Statistics and Demography, University of Southampton*

Fiona Steele, *Department of Statistics, London School of Economics*

George Leckie, *Centre for Multilevel Modelling, University of Bristol*

Rob French, *Centre for Multilevel Modelling, University of Bristol*

AUGUST 2015

ABSTRACT

Recent social and educational policy debate in the UK has been strongly influenced by studies which have found children's cognitive developmental trajectories to be significantly affected by the socio-economic status of the households into which they were born. Most notably, using data from the 1970 British cohort study, Feinstein (2003) concluded that children from less advantaged backgrounds who scored high on cognitive tests at 22 months had been overtaken at age 5 by children from more advantaged origins, who had scored lower on the baseline test. However, questions have been raised about the methodological robustness of these studies, particularly the possibility that their key findings are, at least in part, an artefact of regression to the mean. In this paper we apply and assess the Growth Mixture Model (GMM) as an alternative approach for identifying and explaining cognitive developmental trajectories in children. We fit GMMs to simulated data and to data from the Millennium Cohort Study to assess the suitability of GMMs for studying socio-economic gradients in developmental trajectories. Our results show that GMMs are able to recover the data generating mechanism using simulated data, where the conventional approach is subject to regression to the mean. Substantively, our MCS findings provide no support for the contention that more initially able children from disadvantaged backgrounds are 'over-taken' in cognitive development by less initially able children from more affluent backgrounds. We do, however, find that cognitive developmental trajectories are related to socio-economic status, such that initial class-based inequalities increase over time.

INTRODUCTION

Evidence of growing socio-economic inequalities in the UK, as in many other countries around the world (Picketty, 2014; Stiglitz, 2012; Wilkinson and Pickett, 2009) has led to a focus in policy research on the initial causes of such disparities and, in particular, on how economic inequality is reproduced between generations (Hout, 2015; Washbrook, Gregg and Propper, 2014). This body of research has demonstrated that disparities in cognitive development and educational attainment are evident at very early stages in the life course (Cunha et al., 2006; Goodman et al., 2009). Indeed, substantial social class gradients in cognitive test scores are found when children are as young as 18-24 months, which is perhaps the earliest stage at which it is feasible to administer valid and reliable measures of cognitive ability (Feinstein, 2015). Because cognitive ability and educational attainment are so key to later socio-economic outcomes (Heckman and Mosso, 2014), it is natural that policy-makers have drawn on this evidence to develop and justify policy interventions intended to improve intergenerational social mobility (Lupton 2015). For example, policies such as 'Sure Start' and 'the Pupil Premium' in the UK, which aim to even out attainment disparities across social class groups, provide additional resources to children from disadvantaged backgrounds at earlier points in the life course than has historically been the case. A key plank in the rationale underpinning this policy agenda is that interventions must be implemented as close as possible to the point in the developmental pathway at which socio-economic gradients begin to emerge. Interventions which are targeted later in childhood, adolescence, or early adulthood may arrive too late, as intransigent inequalities and accumulation to existing advantage will already have set in.

Key to the debate over the timing of early years interventions have been studies which analyse longitudinal data on cohorts of young children to assess how cognitive developmental trajectories are related to socio-economic origins (Feinstein 2003; Schoon 2006; Blanden and Machin 2007; Parsons et al 2011). While varying in the detail of their approach, these studies have generally used a methodology which assigns children to groups based on a cross-classification of cognitive test scores at the first point of measurement (generally taken at an early age, such as 24 months) and their parents' socio-economic status (such as social class or income quantile) at the first measurement occasion. In its simplest form, this yields a four category classification: a high ability high-socio-economic status (SES) group; a low ability-high SES group; a low ability-low

SES group; and a high ability-low SES group. Cognitive ability scores are then compared across the four groups at successive points of measurement. These studies, which we shall henceforth refer to as the 'pre-assigned groups' method have consistently revealed two striking patterns. First, as noted above, children's cognitive test scores exhibit stark socio-economic gradients by as early as 22 months. Second, and most importantly for our purposes here, the 'high ability-low SES' group tends to show a decline in test scores over subsequent measurement occasions relative to the test scores of the 'low ability-high SES' group. These countervailing trends result in a 'cross-over', whereby the able but disadvantaged group is 'overtaken' by the less able but more advantaged group by the age of approximately 7 (though the exact age at which the cross-over occurs is imprecisely measured and differs across studies). The cross-over effect is depicted in the well-known chart from Feinstein's 2003 paper, which is reproduced in Figure 1 below.

FIGURE 1 HERE

While there is now little or no dispute regarding the emergence of socio-economic gradients in cognitive test scores early in childhood, scholars have questioned the validity of the methodology which produces the 'cross-over' effect (Read 2003; Tu and Law 2010; Jerrim and Vignoles 2013; Goldstein and French 2015). These authors have pointed to a number of limitations in studies that have used the pre-assigned groups approach, including non-comparability of tests across measurement occasions, conflation of average and individual effects, the arbitrary nature of the group boundary definitions, and failure to account for non-random nonresponse and attrition (see Feinstein, 2015).

Most notably, however, criticism has focused on the potential for the cross-over effect to result as an artefact of regression to the mean (RTM), rather than any substantive differences between the groups. In assigning children to high or low ability groups on the basis of a single test score, some will be mis-classified as a result of measurement error (we note that it is not necessary for a test score to contain measurement error for a regression to the mean effect to be observed (Goldstein and French, 2015)). That is to say, some children will have achieved higher or lower test scores relative to their 'true ability' purely by chance. Thus, when the mis-classified children are re-interviewed in subsequent waves, those in the high ability groups will tend to show a decline in test

performance, while those in the low ability groups will tend to show an improvement in test performance. In other words, there will be a regression to the mean effect.

Because test scores at the first point of measurement are already stratified by SES, regression to the mean should be most pronounced for the high ability-low SES and the low ability-high SES groups, because to have been allocated to the high and low ability groups, respectively, more children in these groups must have scored further from their 'true ability' compared to the other two groups. And this pattern is exactly what the Feinstein analysis reveals. Jerrim and Vignoles illustrate this theoretical expectation using simulated data, showing that the cross-over effect is obtained using the pre-assigned groups method when the simulated data were generated from a population containing two groups with different intercepts but equal (zero) slopes (Jerrim and Vignoles, 2013).

These findings lead Jerrim and Vignoles to conclude that the pre-assigned groups method "can induce substantial bias in estimates of the educational achievement trajectories of different SES and ability groups and thus lead to the wrong conclusions being drawn from trends in the data" (2013, p904). They advocate instead the use of a measurement error correction procedure which they apply to data from the Millennium Cohort Study (MCS). Application of the error correction procedure reveals the familiar social class gradient in cognitive test scores at the first measurement occasion in the MCS, but this gradient does not change appreciably over time (and does not, therefore, exhibit the cross-over effect).

However, the Jerrim and Vignoles procedure requires a quite restrictive assumption that a parallel test of cognitive ability is available at the first measurement wave. A parallel test is one which measures the same underlying construct but whose errors of measurement are uncorrelated with the alternative test score. As Goldstein and French (2015) note, however, these conditions are unlikely to be met very frequently in practice. And, from a practical perspective, even if it were feasible to develop parallel measures, their inclusion in study questionnaires which are already long and complex may well represent an unacceptable additional burden on cohort members, potentially leading to higher rates of attrition. Moreover, the error correction procedure would clearly not be applicable to important existing data sets such as the British Cohort Study and the National Child Development Study, for which only single measures of cognitive ability will ever be

available. It is important, therefore, that new methodological approaches are explored and evaluated which are capable of overcoming the regression to the mean effect. That is our objective in this paper.

We assess the utility of growth mixture modelling (GMM) (Muthén, 2004) as an alternative methodology for evaluating the association between socio-economic origin and cognitive developmental trajectories. The GMM extends the linear growth curve model (e.g. Singer and Willett, 2003; Goldstein, 2011) by identifying latent classes of individuals who follow qualitatively distinct developmental trajectories, with different growth parameters estimated for each of the latent classes. Latent class membership is identified by applying finite mixture models to individual variation in growth parameters from the single latent curve model. In addition to providing a means of determining the number of different latent trajectory groups in a population and the shapes of their trajectories on the repeated outcome over time, GMMs allow trajectory group membership to be predicted as a function of covariates via multinomial logistic regression.

In principle at least, the GMM framework offers a number of attractive properties for studying the relationship between socio-economic status and cognitive development. First, it is not necessary to assign cohort members to high and low ability groups using a discrete threshold on the ability measure at the first point of observation. Instead, trajectory groups are defined as a function of test scores across all measurement occasions. For this reason, we should not anticipate that the trajectory groups thus defined will be subject to the regression to the mean effects experienced by the pre-assigned groups approach. Second, membership of the latent developmental trajectory groups is treated as probabilistic rather than determined, which is preferable from both a conceptual and a measurement perspective. Third, because trajectory groups are derived only as a function of the repeated cognitive ability measures and not of socio-economic origin, it is possible to include multiple predictors of trajectory group membership. Finally, GMMs use all the available data. While existing methods generally discard data from cohort members who are not defined as being in the high or low ability groups, GMMs use all available observations. It is also straightforward in the GMM framework to implement procedures which correct for non-random nonresponse, attrition, and item missing data (Muthén et al., 2011).

The remainder of the paper is structured as follows. First, we present a formal account of the GMM and explain how it can be implemented to study socio-economic gradients in cognitive development. We then describe and set out the results of GMMs fitted to the simulated data used by Jerrim and Vignoles (2013) to assess whether GMMs can recover the data generating mechanism. Next, we fit GMMs to real data from the Millennium Cohort Study (MCS) and evaluate how the parameter estimates of our preferred model relate to the findings of existing studies. We conclude with an assessment of the substantive and methodological implications of our findings and of the suitability of GMMs for studying social class effects on cognitive development.

THE GROWTH MIXTURE MODEL

The GMM (Muthén, 2004; Muthén and Asparouhov, 2009; Vermunt, 2007) consists of a joint model for analysing repeated measures, conditional on individuals' latent classes, and for individuals' probabilities of belonging to each class. In the current context the repeated measures are children's attainment scores across test occasions, while the classes represent subpopulations of children following qualitatively distinct developmental trajectories. Let y_{ij} and t_{ij} denote a continuous attainment score and age at occasion i ($i = 1, \dots I$) for child j ($j = 1, \dots, J$) and let C_j denote the latent class, the values of which are indexed by k ($k = 1, \dots, K$). For the current application, an appropriate GMM is where y_{ij} given $C_j = k$ is a second degree polynomial function of t_{ij}

$$y_{ij|C_j=k} = \beta_0^{[k]} + \beta_1^{[k]} t_{ij} + \beta_2^{[k]} t_{ij}^2 + u_{0j}^{[k]} + u_{1j}^{[k]} t_{ij} + e_{ij}^{[k]}. \tag{1}$$

The regression coefficients $\beta_0^{[k]}$, $\beta_1^{[k]}$ and $\beta_2^{[k]}$ measure the intercept, linear, and quadratic terms of the average attainment trajectory in latent class k, while the random-intercept and -slope effects $u_{0j}^{[k]}$ and $u_{1j}^{[k]}$ allow the intercept and linear component of this trajectory to vary across children. As our application covers only three measurement occasions, we do not include a child-specific quadratic coefficient, although it is straightforward to add further random coefficients with more measurement occasions. The occasion-specific error or residual $e_{ij}^{[k]}$ allows attainment scores to deviate from perfectly quadratic trajectories.

The random effects in each latent class are assumed to be bivariate normally distributed

$$\begin{pmatrix} u_{0j}^{[k]} \\ u_{1j}^{[k]} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^{2[k]} & \\ \sigma_{u01}^{[k]} & \sigma_{u1}^{2[k]} \end{pmatrix} \right\}, \tag{2}$$

where $\sigma_{u0}^{2[k]}$ and $\sigma_{u1}^{2[k]}$ denote the random-intercept and -slope variances in class k and $\sigma_{u01}^{[k]}$ their covariance. The residuals in each class are assumed normally distributed with occasion-specific variances, $e_{ij}^{[k]} \sim N\left(0, \sigma_{ei}^{2[k]}\right)$. Thus, in the most general specification, each class is characterised not only by its own average quadratic trajectory, but also by the extent to which children vary around their average trajectories, and in the degree to which the actual attainment scores vary about the child-specific trajectories. We also estimate $\pi_j^{[k]}$, the posterior probability that child j belongs to each class, k, given their attainment scores (y_{1j}, y_{2j}, y_{3j}).

When $K = 1$, model (1) collapses to a standard growth-curve model (Goldstein, 2011; Raudenbush and Bryk, 2002; Singer and Willet, 2003). When $\sigma_{u0}^{2[k]} = \sigma_{u1}^{2[k]} = \sigma_{u01}^{[k]} = 0 \ \forall k$, the model collapses to a latent class growth analysis (LGCA) or group-based trajectory model (Nagin, 1999, 2005) where the members of each latent class trajectory group are constrained to have identical intercepts and slopes. Predictors of latent class membership can be introduced by inclusion of covariates in a multinomial logistic regression for latent class membership. This can be done in different ways with each approach having advantages and disadvantages. We consider the one-step approach (Muthén, 2004; Muthén and Asparouhov, 2009; Vermunt, 2007), the three-step approach (see Vermunt 2010; Asparouhov and Muthén, 2013), and the so-called pseudo-class approach (Wang et al., 2005).

In the one-step approach, a multinomial logistic regression model for C_j is fitted jointly with the above GMM. The model can be written as

$$p_j^{[k]} \equiv \Pr\left(C_j = k \middle| \mathbf{x}_j \right) = \frac{\exp\left(\mathbf{x}_j' \boldsymbol{\alpha}^{[k]}\right)}{\sum_{l=1}^{K} \exp\left(\mathbf{x}_j' \boldsymbol{\alpha}^{[l]}\right)}, \tag{3}$$

where $\mathbf{x}_j$ denotes the vector of child-level covariates and $\boldsymbol{\alpha}^{[k]}$ denotes the associated vector of regression coefficients where the last class is set to be the reference category (i.e., $\boldsymbol{\alpha}^{[k]} = \mathbf{0}$).

In the three-step approach, Step 1 consists of fitting the above GMM. Step 2 assigns children to their modal latent classes $\hat{C}_j$ based on their predicted posterior class membership probabilities, $\hat{\pi}_j^{[k]}$. Step 3 involves fitting a standard multinomial logistic regression to $\hat{C}_j$. A correction is applied in this final step to account for the classification error introduced in Step 2.

In the pseudo-class approach, children are randomly allocated to classes as a function of their predicted posterior class membership probabilities and then, as before, a standard multinomial logistic regression is fitted to children's assigned classes. This approach should be repeated a large number of times and the results combined via Rubin's (1987) multiple imputation rules in order to propagate the uncertainty which arises from probabilistically assigning children to classes (Wang et al., 2005).

The main disadvantage of the one-step approach is that adding predictors of class membership may lead to changes in the number and sizes of the latent classes and therefore their substantive interpretation (Muthén and Asparouhov, 2009; Vermunt, 2007). This may not fit well with the logic of most applied researchers who view introducing covariates as a step that comes after the classification model has been built. However, a notable advantage of the one-step approach over the other two approaches is that it fits the growth curve and class membership components of the model jointly and by full-information maximum likelihood estimation and therefore assumes missing achievement and covariate data to both be missing at random (MAR). In contrast, the growth curve component in the three-step and pseudo-class approaches assumes that missing data is MAR and only with respect to the observed test score data not the observed covariate data. Moreover, the multinomial logistic regression component drops children with missing covariates and so makes the more restrictive assumption that the missing covariates are missing completely at random (MCAR). Thus, when there is substantial missing data the one-step approach may be preferred.

Deciding on the number of classes is a difficult topic in growth mixture modeling. Reviews by Nylund, et al. (2007), Tofighi & Enders (2008) and Yang (2006) suggest that the sample-size adjusted BIC (Sclove, 1987) and LMR statistic (Lo, Mendell & Rubin, LMR, 2001) tend to perform well in extracting the correct number of classes. The adjusted BIC measures the goodness of fit, penalized for model complexity (i.e., number of model parameters and sample size) while the LMR is a modified version of a standard likelihood ratio test which recognizes that LRT statistics which compare models with different numbers of classes have non-standard chi-squared distributions. We fit all models in the Mplus software Version 7.2 (Muthén and Muthén, 1998-2013).

APPLICATION TO SIMULATED DATA

We begin with an application of GMMs to simulated data to assess how well they recover features of the measurement model from which the data were generated. Perhaps the most compelling line of evidence in Jerrim and Vignoles' (2013) critique of Feinstein (2003) is that they find the crossing pattern using the pre-assigned groups method using simulated data, even though no such pattern existed in the simulated population from which the sample data were drawn. A key first step in assessing the utility of GMMs in this context then is to evaluate whether they can successfully recover the data generating mechanism, or whether the cross-over effect is also incorrectly produced.

The data were simulated using the same assumptions as Jerrim and Vignoles (2013). Denote by A_{ij} the 'true ability' at occasion i for child j and suppose that children are in one of two equally-sized SES groups indicated by x_j such that $A_{ij|x_j=g} \sim N(\mu_g, \delta_g)$ for $g = 0, 1$. Following Jerrim and Vignoles we assume $\delta_g = \delta = 1$, and that a child's ability is constant over time, i.e. $A_{ij} = A_j$ for all i. Suppose that A_{ij} is measured with error and that we observe y_{ij} which is related to A_{ij} by the linear measurement model

$$y_{ij} = A_{ij} + \epsilon_{ij}$$

where the measurement error $\epsilon_{ij} \sim N(0, \gamma)$. Based on these assumptions, the data generating model (DGM) for y_{ij} can be written

$$y_{ij} = \theta_0 + \theta_1 x_j + u_j + \epsilon_{ij} \qquad (4)$$

where $u_j \sim N(0,1)$, $\theta_0 = \mu_0$ and $\theta_1 = \mu_0 + \mu_1$. Thus the measurement model takes the form of a random intercept multilevel model with a single predictor x_j and no time trend.

We consider how the performance of the GMM depends on (i) the mean difference between the SES groups (θ_1), and (ii) the reliability r of y where $r = \text{corr}(A_{ij}, y_{ij}) = 1/\sqrt{1 + \gamma}$. Following Jerrim and Vignoles, we generate data for each combination of $\theta_1 = 1.4$ and 3, and $r = 0.4$ and 0.75, leading to four simulation conditions. Each simulated dataset contains 20,000 observations representing children, with each measured at all three occasions.

We fit two-class GMMs to the simulated data to investigate how well the latent classes map on to the observed SES categories. The fitted GMM has a more complex form than the DGM of equation (4). In particular, we specified a model with a class-specific quadratic time trend (equivalent to including a dummy variable for each occasion for 3 time points), and allowed for class-specific between-individual (intercept) variances. Class and occasion-specific residual variances were also considered, but these models failed to converge. We therefore present results from models with the residual variances constrained to be equal across classes, i.e. $\sigma_{ei}^{2[k]} = \sigma_{ei}^2$ for $k = 1,2$. Convergence problems were experienced with several of the models which is to be expected because the simplicity of the DGM means that more general specifications are not supported by the simulated data. Nevertheless it is important to release parameter constraints to assess the extent to which the parameter estimates are consistent with the DGM.

The fitted model has the same form as equation (1) but excluding the random slope on t_{ij} (and constraining the occasion specific residual variances to be class invariant):

$$y_{ij|C_j=k} = \beta_0^{[k]} + \beta_1^{[k]} t_{ij} + \beta_2^{[k]} t_{ij}^2 + u_{0j}^{[k]} + e_{ij}^{[k]}, \qquad k = 1, 2$$

where $u_{0j}^{[k]} \sim N(0, \sigma_{u0}^{2[k]})$ and $e_{ij}^{[k]} \sim N(0, \sigma_{ei}^2)$. If the GMM perfectly reproduces the DGM, and the latent classes map on to the categories of x_j, we would have $\left| \beta_0^{[1]} - \beta_0^{[2]} \right| = |\theta_1|$, $\beta_1^{[k]} = \beta_2^{[k]} = 0$ for $k = 1, 2$, $\sigma_{u0}^{2[1]} = \sigma_{u0}^{2[2]}$ and $\sigma_{e1}^2 = \sigma_{e2}^2 = \sigma_{e3}^2$. Furthermore, the average probability of class membership would be 0.5, $\bar{\pi}^{[1]} = \bar{\pi}^{[2]}$.

The GMMs were fitted to standardised scores. While it is common to apply the pre-assigned groups method to percentiles, this is inappropriate for the GMM because percentiles follow a uniform distribution while a GMM assumes that residuals and random effects are normally distributed. Under the model of equation (4) and using the result that $\text{var}(x_j) = 0.25$ for a binary variable with equal sized categories, the marginal variance of the observed responses is given by

$$\text{var}(y_{ij}) = v^2 = 0.25\,\theta_1^2 + 1 + \gamma. \tag{5}$$

Thus we compare estimates of $\left| \beta_0^{[1]} - \beta_0^{[2]} \right|$ with $\theta_1^* = \theta_1/v$.

After estimating the GMM, we assessed the correspondence between the two latent classes and categories of x_j by fitting a multinomial logistic regression of modal class membership $\widehat{C}_j$, as in the three-step approach discussed above, with x_j as the only predictor. As the third step of this approach did not converge for one of the simulation conditions, we account for uncertainty in class allocation using the pseudo-class approach. We anticipate that the association between $\hat{C}_j$ and x_j, as measured by the odds ratio $\exp(\alpha_1)$, will be stronger when there is greater separation of the classes. Therefore we expect $\exp(\alpha_1)$ to be highest for the conditions where $\theta_1 = 3$ and $r = 0.7$.

Simulation Results

Table 1 and Figure 2 show the results from fitting two-class GMMs for three of the simulation conditions. The GMM did not converge for the fourth condition, $\theta_1 = 1.4$ and $r = 0.4$, which corresponds to the situation where the groups defined by x_j are least well distinguished. Figure 2 shows that in each case the predicted trajectories are constant, as in the DGM; thus $\beta_1^{[k]}$ and $\beta_2^{[k]}$ are estimated close to zero in the GMM.

We also find that for each condition the estimated between-class difference in the intercepts is close to θ_1^*, the standardised mean difference assumed in the DGM (Table 1). Estimates of the intercept variances are also similar for the two classes, estimates of the residual variances are similar across occasions (Appendix Table A1), and estimates of the total variance within each class are close to the variance conditional on x_j under the DGM (Appendix Table A2). In terms of the correspondence between the latent classes and the observed SES indicator x, the classes are of roughly equal size (Figure 2) and class membership is strongly associated with x (Table 1). As expected, the classes are more closely aligned with SES when the groups are better distinguished.

Although the results suggest that the GMM is able to recover the key features of the DGM, we note that entropy is low (Table 1) when the mean difference between SES categories is small or reliability is low. This may seem surprising given that the number of classes matches the number of groups in the DGM but entropy is a measure of the separation between classes and precision of the classification rather than model fit. It is possible for a correctly specified model to have a low entropy, as for two of the scenarios considered here (Muthén, 2004; Petras and Masyn, 2010). In such cases, the latent classes may not be useful and it is especially important to allow for uncertainty in class allocation when modelling the effects of covariates on class membership. Use of entropy-based fit indices (e.g. Celeux and Soromenho, 1996), might lead researchers to add spurious classes to the model in order to achieve more homogenous classes and therefore higher entropy. However, for all three simulation conditions, the addition of a third class led to non-convergence because, as expected given the DGM, the average probability of class membership for the additional class was very low (<0.5%).

TABLE 1 HERE

FIGURE 2 HERE

Despite these limitations, it is clear that the GMM represents a significant improvement over the pre-assigned groups method when applied to the simulated data; the DGM is accurately recovered in three of the four conditions, with no evidence of the cross-over pattern or other indicators of regression to the mean.

APPLICATION TO MILLENIUM COHORT STUDY DATA

Data and Measures

Having shown that the GMM is not subject to the same regression to the mean effect as the pre-assigned groups method when applying a simple 2-class model to simulated data, we proceed to fit more complex GMMs to data from the Millennium Cohort Study (MCS). The MCS is a longitudinal survey that began in 2000 and tracks the social, economic, and health status of a nationally representative random sample of children born between 2000 and 2001. The first wave survey had a response rate of 68% and an achieved sample size of 18,552. The following three waves of data collection were conducted when the children were 3, 5 and 7 years old, with response rates (conditional on being present in the first wave) of 78%, 79% and 72% and achieved sample sizes of 15,590, 15,246 and 13,857 respectively. The survey design and fieldwork outcomes are described in detail elsewhere (Hansen, 2012).

For our measures of cognitive ability we use the scales administered and derived by the MCS team, which measure vocabulary and reading skills at ages 3, 5, and 7. At ages 3 and 5 the naming vocabulary subset of the British Ability Scale (BAS) was used. Children were shown brightly coloured pictures and asked to name the object in the picture. At age 7, a word reading test was administered, in which children were shown a series of words on a card and asked to read them aloud. The children were shown a maximum of 90 words, but if a child read 8 words in a block of 10 incorrectly the test was stopped (Hansen, 2012). In order to make comparisons between these tests over time, test scores were standardised to a mean of zero and standard deviation of one. The test at age 7 was an assessment of slightly different language skills to those tested at age 3 and 5, but all tests are nonetheless indicative of a child's overall language ability. These tests have been used both by Jones and Schoon (2010) and Jerrim and Vignoles (2013) to compare changes in children's language skills over time.

As noted earlier, an advantage of the GMM is that we are not restricted to using only one measure of socio-economic status as a predictor of cognitive developmental trajectories and, therefore, include three such measures in our models. The first is a measure of household income at wave 1. This is an equivalised income measure that adjusts household income for the number of adults and children in the household (Hansen, 2012). The second is a binary indicator of whether families were in receipt of one or

more of the following benefit payments at wave 1: Job Seekers Allowance, Income Support, Work Families Tax Credit, or Disabled Person's Tax Credit. The third is the National Statistics Socio-Economic Classification (NS-SEC) of the current or, in the case of parents who were not in work, their most recent job held by the interviewed parent at wave 1.

We also include the following covariates measuring aspects of family structure: the interviewed parent's marital status at wave 1, their (banded) age at child's birth; and whether or not the interviewed parent had a longstanding illness, disability or infirmity at wave 1. Sex of cohort members is also included as a covariate in all models.

Model selection and interpretation of latent trajectory groups in the GMM

Models are fitted using the one-step procedure and full-information maximum likelihood (FIML) estimation. The one-step approach using FIML is preferred for the MCS data because it derives the latent trajectory groups assuming missing data to be missing at random (MAR), conditional on the covariates in the multinomial regression. The one-step procedure in MPlus does not currently permit the covariates to be used to adjust for non-random missingness in the latent class model, although this is a software not a theoretical limitation. We also fitted the models using the three-step procedure with missing data deleted case-wise and found that, with some small variation, the latent class solutions were very similar to those presented here.

We begin by examining the BIC, sample size adjusted BIC, and LMR p-values for models with an increasing number of latent classes to ascertain the optimal number of latent trajectory groups (Table 2). Entropy values are also presented as a measure of the separation of classes but, in light of the results from the analysis of simulated data, we note that entropy will in general be low when mean differences among classes are small or residual variance (measurement error) large. The fit statistics suggest that a five group model is optimal, taking into account fit to the data, substantive utility, and parsimony (Muthen 2004). Although the p-value of the LMR becomes non-significant with the addition of the 6th trajectory group, the BIC and sample size adjusted BIC statistics decrease, although these decreases are marginal. Moreover, the parameter estimates for the 6-class model reveal the additional trajectory group to be a 'splinter' class,

16

representing just 3% of children, which added little in terms of substantive insight. We therefore prefer the somewhat less well fitting but more parsimonious 5-class model.

TABLE 2 HERE

Table 3 presents the posterior probability table for the five class model. The posterior probability table is a cross-tabulation of the most likely (i.e., modal) class for each child by the mean posterior probability of class membership. A model with clearly distinguished classes should have high values, approaching one, along the main diagonal and low off-diagonal values, approaching zero (Nagin 1999). Large off-diagonal values are indicative of indeterminacy between classes. In support of our decision to settle on a five class model, values on the diagonal in Table 3 are high, ranging from 0.71 to 0.82, and those on the off diagonal are low and close to zero.

TABLE 3 HERE

Figure 3 shows fitted values for the latent trajectory groups for the five class model. The parameter estimates for this model are presented in Table A3 in the Appendix. The GMM solution in Figure 3 bears strong similarities to the pattern in the Feinstein (2003) chart (reproduced in Figure 1). A group is identified which scores high on the baseline cognitive test and continues to achieve high scores over subsequent measurement occasions (group 1, 34% of observations), while a second group is identified which initially scores low and continues to do so over the next two measurements, with some evidence of a decline in achievement between the second and third measurement occasions (group 4, 17% of observations). There is also a group which achieves a high score on the test at age 3 but whose performance on the test declines over time (group 2, 16% of observations) and a group which initially scores low but test scores then increase over successive waves (group 3, 29% of observations). These four groups appear to replicate the pattern shown in the Feinstein chart, including the 'cross-over' effect between groups 2 and 3. Additionally, the GMM produces a small group of children (group 5, 4% of observations) who achieve very low scores on the age 3 test but who then show a marked improvement over successive waves. As we noted earlier, the emergence of additional classes of this kind is likely to result from including the full set

of sample observations, as opposed to selecting only those in the high and low ability groups at wave 1.

FIGURE 3 HERE

Correlates of latent trajectory group membership

We turn next to an assessment of the correlates of group membership via multinomial logistic regression of group membership on the socio-economic status and household structure covariates. Table 4 presents the coefficient estimates and the associated standard errors for each covariate. Coefficients are the log odds of membership in the first group in the column header, relative to group 3, the reference group, for each unit increase in the covariate (additional contrasts are presented in table A4). So, for example, 0.254 in the first cell of the second column (labelled 1 v 3) is the expected change in the log-odds of being in trajectory group 1 relative to trajectory group 3 for each unit increase in equivalised household income. Given the large number of contrasts in Table 4 and A4, we do not interpret all of them here. Rather, we focus our attention on the contrasts which pertain most directly to the question of whether declining and increasing performance on the tests is associated with socio-economic (dis)advantage. We also separately consider the correlates of group 5, given that it was not anticipated to emerge *a priori*. In particular, we wish to exclude the possibility that this group arises as an artefact of measurement or analysis operations.

TABLE 4 HERE

The intercepts of the latent trajectories for all groups represent the mean score for the group at the first point of measurement. These are ordered by income, as would be expected from existing research. For example, the odds of membership in the highest scoring group (group 1) relative to all other groups increases with parental income, social class, and (non) receipt of benefits. The only exception to this pattern is the contrast between groups 3 and 2, the two groups which cross over time. None of the three socio-economic variables show a significant difference in the odds of being in the rising (group 3) rather than the declining group (group 2). This is, of course, counter to the pattern found using the pre-assigned groups approach. We also fitted these models using each socio-economic predictor in isolation to check whether the null results might have arisen due to multicollinearity, or some form of suppressor effect. However, none of the socio-

economic variables significantly differentiates between groups 3 and 2 when considered on its own either. We therefore conclude that, while the GMM does detect groups which exhibit a reversal in their achievement on the cognitive test scores between age 3 and age 7, there is no evidence to support the claim that these groups differ in their level of socio-economic disadvantage.

However, although socio-economic status does not differentiate between the two 'cross-over' groups (groups 3 and 2), there is evidence from the model to suggest that socio-economic (dis)advantage is related to change in test performance over time. Groups 1 and 2 both have large intercepts, indicating approximately equivalent high performance in the age 3 baseline test. However, while group 1 maintains this high level of achievement over time, group 2 declines significantly and substantially over the two successive waves. The odds of membership in group 1 relative to group 2 are smaller for children from households with lower equivalised household income, in lower social class groups, and whose parents are in receipt of state benefits. Similarly, group 3 and group 4 have small intercept coefficients, indicating a low score on the baseline test. Yet, while group 4 continues to perform poorly, the test scores of the children in group 3 increase significantly over time on average.

The pattern of covariate relationships for these two groups mirrors that found for the contrast between groups 1 and 2. The odds of membership in the rising group (group 3) relative to the static group (group 4) increase with income and social class, though benefit receipt does not discriminate significantly between these groups. We therefore conclude that socio-economic status was significantly related to the cognitive developmental trajectories of this cohort during this period. Amongst initially low scoring children, those from more affluent backgrounds were more likely to experience a subsequent improvement on their test scores, while amongst initially high scoring children, those from less affluent backgrounds were significantly more likely to experience a decline in their test scores at ages 5 and 7.

While the socio-economic variables did not discriminate between the two cross-over trajectory groups, parental marital status at baseline did. Children whose parents were married at the time of first interview have significantly higher odds of being in the rising ability group (group 3) compared to the declining ability group (group 2), relative to children whose parents were divorced/separated, widowed, or had never married. It

would be unwise to place too much interpretative emphasis on this association, based as it is on a single coefficient in a large and complex model. However, it may be taken as suggestive that aspects of family structure and dynamics, such as marital breakdown, may play a role in influencing children's cognitive developmental trajectories.

We also find that gender is significantly related to group membership for all but one contrast (between groups 4 and 5). We therefore also fitted models separately for male and female cohort members and found essentially the same latent class solutions as presented here for the joint model. This suggests that the latent class solution for the joint model is not driven by gender differences in cognitive development but, rather, that the gender mix varies across an essentially invariant pattern of latent trajectory groups.

Lastly, we turn to group 5, which is a small (only 4% of children have group 5 as their most likely group based on the posterior probabilities), but potentially substantively interesting group, which would not have been detected using the pre-assigned groups approach. Group 5 shows the lowest mean score on the age 3 test of all five groups but this is followed by rapid improvement in test scores over the ensuing two waves. The odds of membership in group 5 decrease with parental income and social class, for children whose parents are single and never married, and whose parents do not have a longstanding illness. This pattern of covariate associations is suggestive of trajectory group 5 comprising children from more economically disadvantaged backgrounds but who have a stable family structure.

DISCUSSION

Public policy in the UK has moved in recent years toward making interventions at earlier points in the life course than has historically been the case. This reflects a growing recognition that quite substantial gradients in cognitive development and academic achievement are evident very early in childhood (Crawford, Macmillian & Vignoles, 2014; Cunha, Heckman & Lochner, 2006). Thus, the argument goes, if redistributive policy interventions are to have maximum impact on equalising important life outcomes, they need to be implemented at or before the point in the developmental pathway that socio-economic gradients begin to emerge.

A key plank in the evidence base supporting this early years policy framework has been provided by studies which track representative samples of children from birth, through childhood and into adolescence, such as the British Cohort Study and the Millennium Cohort Study. Researchers have analysed these and other data sets to show not only that large differences in cognitive test scores are evident at early points in the life course, but also that children from different socio-economic backgrounds appear to pursue quite different cognitive developmental trajectories. Of particular significance has been the finding that initially less able children from more affluent backgrounds 'overtake' initially more able children from less advantaged families between the ages of 5 and 7 years (Feinstein 2003; Schoon 2006; Blanden and Machin 2007; Parsons et al 2011). This stylised fact has been cited by key political actors and referenced in a number of important UK government reports in support of early intervention policies (Lupton, 2015).

However, scholars have questioned the validity of the methodology underpinning this key conclusion, arguing that the 'cross-over' effect is a statistical artefact caused by regression to the mean (Tu and Law 2010; Jerrim and Vignoles 2013). Yet, while these studies have convincingly demonstrated that the conventional approach to analysing this kind of data - what we have termed the pre-assigned groups approach - is likely to be subject to regression to the mean effects, we cannot conclude from this evidence alone that the cross-over effect does not happen, nor that socio-economic status does not affect cognitive development in other, perhaps more subtle ways. As Tu and Law note in their methodological critique of Feinstein (2003) and Blanden and Machin (2007), "it is not the case that the conclusions from studies whose data analyses suffer regression to the mean are always invalid" (Tu and Law, 2009, p1249). It is essential, then, that alternative methodological approaches are developed to address this key policy question, not least because, the error correction method proposed by Jerrim and Vignoles (2013) is quite restrictive in its data requirements.

The GMM framework, we have argued, offers a number of potential advantages over existing approaches to the study of social class gradients in cognitive development. GMMs easily incorporate information from multiple waves. Where the number of time points is greater than three, it is possible to introduce quadratic and higher order polynomial growth functions, which vary across latent trajectory groups to accurately

model qualitative differences in developmental change. GMMs are also able to deal flexibly with unit and item missing data, an inevitable feature of cohort studies. In the analyses presented here, we used a full information maximum likelihood estimator (Arbuckle 1996), which is consistent and efficient assuming missing data to be missing at random (MAR) conditional on the observed data. Other approaches for dealing with missing data, such as multiple imputation, and nonresponse weighting can also be implemented in software for estimating GMMs. Most importantly with regard to regression to the mean, GMMs do not require trajectory groups to be defined at the first observation by placing an arbitrary threshold on the test score. Instead, the groups are derived as a function of change across all measurement occasions.

In our empirical applications, the GMM performed well. Applied to the same simulated data used by Jerrim and Vignoles (2013), the GMM was able to recover the data generating mechanism, under a range of mean differences between groups and levels of measurement error on the test score. However, the entropy values indicate that the groups were not well defined when mean differences between groups were small and there was a high level of measurement error on the test score. In such cases, although the selected model may be a good fit to the data, further analysis of the latent trajectory groups should be carried out with caution and must take account of classification error. Reassuringly, the addition of a third class to the GMM was not supported by data generated from a two-class model. In practice of course the DGM will be unknown and a variety of factors, including model misspecification and violation of distributional assumptions, may lead to the addition of spurious classes to achieve a good-fitting model (Bauer and Curran, 2004). It is therefore especially important with GMM to check model assumptions and to assess the validity of the extracted groups.

When applied to the MCS data, the preferred GMM produced a solution very similar in structure to the Feinstein (2003) chart. Thus, the model yielded a group with a high initial test score which was maintained over successive waves, a group with a low initial test score which did not improve over subsequent waves, a group which started with a low test score but then improved over time, and a group which scored high on the baseline test but whose performance then declined. The latter two groups exhibited the 'cross-over' pattern that has come to attract so much attention in both policy and academic debate. Additionally, the GMM produced a small 5th latent trajectory group, characterised

by a very low initial test score followed by substantial and sustained improvement over the subsequent 2 waves.

However, while the GMM produced two large groups which crossed in their trajectories, we found no evidence that membership of these groups was related to socio-economic status, whether measured using household income, social class, or receipt of state benefits. The MCS models did, though, provide evidence in support of the contention that socio-economic status is associated with widening gradients in cognitive test scores at this point in the life course. Our analyses showed that the group of children who start poorly but improve over time are more likely to be in higher socio-economic groups than those who achieve low initial scores and do not improve. Conversely, we found that children who start with high initial scores but subsequently decline are more likely to be from less socio-economically advantaged backgrounds compared to the group that achieve high initial test scores and continue to perform well in subsequent tests.

Thus, while the GMMs fitted to MCS data provide no evidence in support of the claim that "by the age of five, bright children from poorer backgrounds have been overtaken by less bright children from richer ones" (Nick Clegg, 2011), neither do they accord with Jerrim and Vignoles' conclusion that, "although family background has a major influence on the child's earliest level of cognitive development, it does not have a strengthening effect that would cause SES gaps in children's cognitive achievement to widen further beyond the age of 3 years" (2013, p905). In fact, our results are closer substantively to those of Goldstein and French (2015) who, using the same MCS data but employing linear growth curve models, also find evidence of growing disparities in cognitive performance as a function of household income. While the visually powerful image of bright working class children being overtaken by their less able but more affluent peers appears to be over-stated as a general phenomenon, our findings suggest that the critique of the evidence on which this was based may itself have been something of an over-correction. The application of growth mixture models leads us to conclude that the method has much to recommend it in this substantive context and that socio-economic status is related to change in as well as to levels of cognitive ability at this early stage of the life cycle.

REFERENCES

Asparouhov, T. and Muthén, B. O. (2013). Auxiliary variables in mixture modeling: 3-step approaches using Mplus. *Mplus Web Notes No. 15*. Retrievable from www. statmodel. com.

Blanden, J. and Machin, S. (2007) *Recent changes in intergenerational mobility in Britain*. Sutton Trust, London.

Bauer, D. J. and Curran, P. J. (2004) The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9*, 3-29.

Celeux, G. and Soromenho, G. (1996) An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification,* 13, 195-212.

Nick Clegg, Deputy Prime Minister (2011), HC Deb 5 April 2011, vol 526, col 895.

Crawford, C., Macmillan, L., & Vignoles, A. (2014). *Progress made by high attaining children from disadvantaged Backgrounds*. Research report.

Cunha, F. Heckman, J. & Lochner, L. (2006). Interpreting the Evidence on Life Cycle Skill Formation. In E. Hanushek & F. Welch *(Eds)Handbook of the Economics of Education*. Pp. 697-812. Amsterdam: Holland North.

Feinstein, L., (2003) Inequality in the Early Cognitive Development of British Children in the 1970 Cohort. *Economica, 70*, 277, 73-98.

Feinstein, L. (2015) Social class differences in early cognitive development and regression to the mean, *Journal of Longitudinal and Life Course Studies*, 6(3), 331-343.

Fitzmaurice, G. M., Laird, N. M. and Ware, J. H. (2011). *Applied Longitudinal Analysis*. 2nd edition. John Wiley & Sons, Hoboken, New Jersey.

Goldstein, H. (2011). *Multilevel Statistical Models*. 4th edition. John Wiley & Sons, Chichester, UK.

Goldstein, H. and French, R. (2015) Differential educational progress and measurement error , *Journal of Longitudinal and Life Course Studies*, 6(3), 351-358.

Goodman A., Sibieta L., & Washbook E. (2009). Inequalities in educational outcomes among children aged 3 to 16. Final report for the National Equality Panel, Institute for Fiscal Studies, London.

Hansen, K. (ed) (2012) *Millennium Cohort Study: First, Second, Third and Fourth Surveys. A Guide to the datasets (7th ed)*, London: Centre for Longitudinal Studies.

Hanushek & F. Welch (Eds)*Handbook of the Economics of Education*. Pp. 697-812. Amsterdam: Holland North.

Heckman, James J. & Mosso Stefano (2014) The Economics of Human Development and Social Mobility, Annual Review of Economics, *Annual Reviews*, vol. 6(1), pages 689-733, 08

Hedeker, D. and Gibbons, R. D. (2006). *Longitudinal Data Analysis*. John Wiley & Sons, Hoboken, New Jersey.

Hout, M. (2015) A Summary of What We Know About Social Mobility, *The Annals of the American Academy of Political and Social Science*, 657, January, 27-36.

Jerrim, J. and Vignoles, A. (2013) 'Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes', *Journal of the Royal Statistical Society Series A*, 176, 3.

Jones, E. and Schoon, I. (2010) 'Cognitive Development', in Hansen, K., Jones, E., Joshi, H. and Budge, D. (eds) *Millennium Cohort Study Fourth Survey: a user's guide to initial findings (2nd ed)*, London: Centre for Longitudinal Studies.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88, 767-778.

Lupton, R. (2015) The practice of policy-related research, *Journal of Longitudinal and Life Course Studies*, 6(3), 369-376.

Muthén, B. (2004). Latent variable analysis: growth mixture modeling and related techniques for longitudinal data. In: D. Kaplan (Ed.), *Handbook of Quantitative Methodology for the Social Sciences*, Sage Publications, CA, 345–368.

Muthén, B and. Asparouhov, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In: Fitzmaurice GM, Davidian M, Verbeke G, Molenberghs G, editors. *Longitudinal Data Analysis*. Boca Raton: CRC Press, 143–165.

Muthén, B., Asparouhov, T., Hunter, A. and Leuchter, A. (2011) Growth modeling with non-ignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods,* 16, 17-33.

Muthén, L. K., & Muthén, B. O. (1998-2013). *Mplus User's Guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Nagin, D. S. (1999). Analyzing Developmental Trajectories: A Semi-parametric, Group-based Approach. *Psychological Methods*, 4, 139-77.

Nagin, D. (2005). *Group-based modeling of development*. Harvard University Press.

Nylund, K. L., Asparouhov, T. & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14, 535-569.

Parsons, S., Schoon, I., Rush, R. and Law, J. (2011) Long-term outcomes for children with early language problems: beating the odds. *Chld. Soc.*, **25**, 202–214.

Petras, H. and Masyn, K. (2010) General growth mixture analysis with antecedents and consequences of change In *Handbook of Quantitative Criminology* eds A. Piquero and D. Weisburd), pp. 69-100, New York: Springer.

Piketty, T. (2014) Capital in the twenty-first century. Cambridge, MA. Harvard University Press.

Plewis, I. (2007) 'Non-response in a birth cohort study: the case of the Millennium Cohort Study', *International Journal of Social Research Methodology*, 10, 5: 325-334.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd edition. Sage, Thousand Oaks, CA.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley: New York.

Schoon, I. (2006) *Risk and Resilience: Adaptations in Changing Times,* Cambridge: Cambridge University Press.

Sclove, L. S. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52, 333-343.

Singer, J. D. and Willett, J. B. (2003). *Applied longitudinal data analysis: modeling change and event occurrence*. Oxford; New York: Oxford University Press.

Snijders, T. A. B. and Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. 2nd edition. Sage.

Stiglitz, J. (2012) *The price of inequality: How today's divided society endangers our future*. New York, NY; W. W. Norton.

Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), Advances in latent variable mixture models, (pp. 317-341). Greenwich, CT: Information Age.

Tu, Y.K., & Law, J. (2010). Re-examining the associations between family backgrounds and children's cognitive developments in early ages. *Early Child Development and Care, 180* (10),1243-12

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.

Vermunt, J. (2007). Growth models for categorical response variables: Standard, latent-class, and hybrid approaches. In: van Montfort K, Oud J, Satorra A, editors. Longitudinal models in the behavioral and related sciences. New Jersey: Lawrence Erlbaum, 139–158.

Vermunt, J. K. (2010). Latent class modeling with covariates: two improved three-step approaches. *Political Analysis*, 18, 450-469.

Wang, C. P., Hendricks Brown, C. and Bandeen-Roche, K. (2005). Residual diagnostics for growth mixture models: Examining the impact of a preventive intervention on multiple trajectories of aggressive behavior. *Journal of the American Statistical Association*, 100(471), 1054-1076.

Washbrook, E, Gregg, P & Propper, C 2014, 'A decomposition analysis of the relationship between parental income and multiple child outcomes'. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol 177., pp. 757-782.

Wilkinson, R. and Pickett, K. (2009) *The Spirit Level: Why more equal societies almost always do better*, London: Bloomsbury Press.

Yang, C. C. (2006). Evaluating latent class analyses in qualitative phenotype identification. *Computational Statistics & Data Analysis*, 50, 1090-1104.

TABLES AND FIGURES

Table 1. Results from fitting two-class GMMs to simulated data. The model assumes quadratic growth and random intercepts within classes.

Simulation conditions			GMM estimates		Effect of x on class membership	
θ_1	r	θ_1^*	$\left\lvert \hat{\beta}_0^{[1]} - \hat{\beta}_0^{[2]} \right\rvert$	Entropy	$\exp(\alpha_1)$	p-value
1.4	0.75	0.932	0.988	0.275	3.66	<0.001
3	0.75	1.494	1.499	0.695	44.61	<0.001
3	0.40	1.059	1.075	0.456	9.67	<0.001

Table 2. MCS model fit statistics and entropy

Classes	BIC	Sample Size Adjusted BIC	LMR p-value	Entropy
2	94905	94816	<.001	.523
3	94299	94149	<.001	.616
4	94032	93822	0.106	.632
5	93654	93383	.001	.643
6	93615	93284	0.571	.653

Table 3. Average Latent Class Probabilities by Most Likely Classes for 5 Class Model

	Groups	Average latent class probabilities				
		1	2	3	4	5
Most	1	*0.818*	0.106	0.074	0.002	0.000
likely	2	0.140	*0.710*	0.072	0.077	0.001
group	3	0.111	0.051	*0.726*	0.089	0.022
	4	0.004	0.082	0.142	*0.736*	0.035
	5	0.003	0.006	0.077	0.09	*0.824*

Table 4. Covariate coefficient contrasts with group 3 of the MCS 5 class multinomial logistic model for trajectory group membership

	1 v 3	2 v 3	4 v 3	5 v 3
Standardised income	0.254* (0.043)	0.011 (0.059)	-0.804* (0.095)	-1.44* (0.227)
benefit payments				
Yes	-0.379* (0.083)	0.016 (0.102)	0.073 (0.099)	0.035 (0.145)
NS-SEC, ref: Managerial/prof				
Intermediate	-0.408* (0.084)	-0.09 (0.115)	-0.016 (0.147)	0.314 (0.244)
Self Employed	-0.181 (0.158)	0.165 (0.204)	0.502* (0.229)	0.993* (0.322)
Technical	-0.57* (0.148)	0.215 (0.164)	0.522* (0.186)	0.872* (0.283)
Routine	-0.709* (0.091)	-0.029 (0.114)	0.671* (0.128)	1.059* (0.205)
Marital Status ref: single				
Married	-0.102 (0.077)	-0.21* (0.092)	-0.156 (0.094)	1.78* (0.224)
Divorce/Separate	0.154 (0.141)	0.08 (0.159)	0.193 (0.15)	1.157* (0.284)
Widowed	-0.442 (0.740)	0.106 (0.712)	-0.519 (0.974)	2.102* (0.775)
Parent long term illness				
Yes	-0.006 (0.073)	-0.021 (0.091)	0.031 (0.091)	-0.337* (0.147)
Parent's Age at Birth: ref: Under 20				
20-39	0.225 (0.172)	0.065 (0.163)	0.062 (0.143)	-0.022 (0.285)
40+	0.323 (0.258)	0.376 (0.282)	0.087 (0.312)	-0.009 (0.515)
Child's gender				
Female	0.599* (0.061)	0.244* (0.095)	-0.444* (0.082)	-0.302* (0.12)

*p<=0.05; Standard errors in parentheses

Figure 1. Reproduction of Figure 2 of Feinstein (2003) showing average rank of test scores at 22, 42, 60 and 120 months, by SES of parents and early rank position.
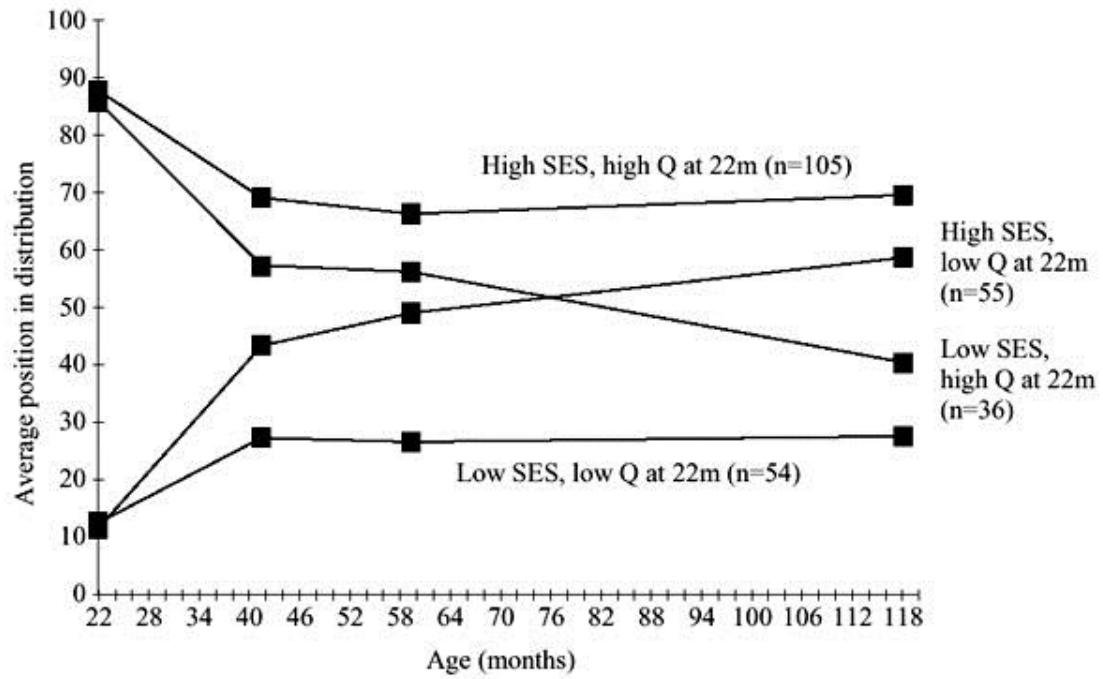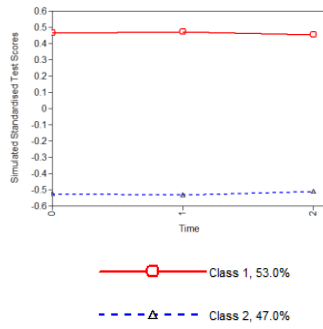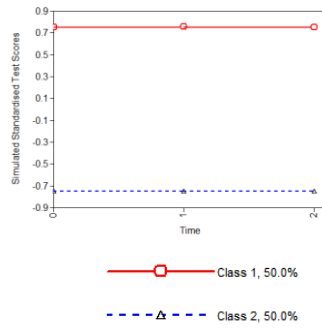
Figure 2. Predicted trajectories and latent group membership probabilities from 2-class GMM model with quadratic growth and random intercepts within classes

(a) $\theta_1 = 1.4, r = 0.75$ (b) $\theta_1 = 3, r = 0.75$ (c) $\theta_1 = 3, r = 0.40$
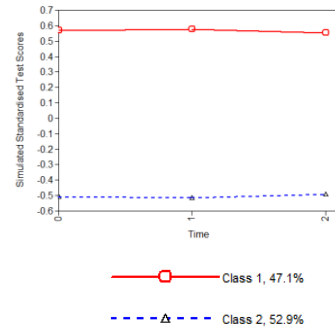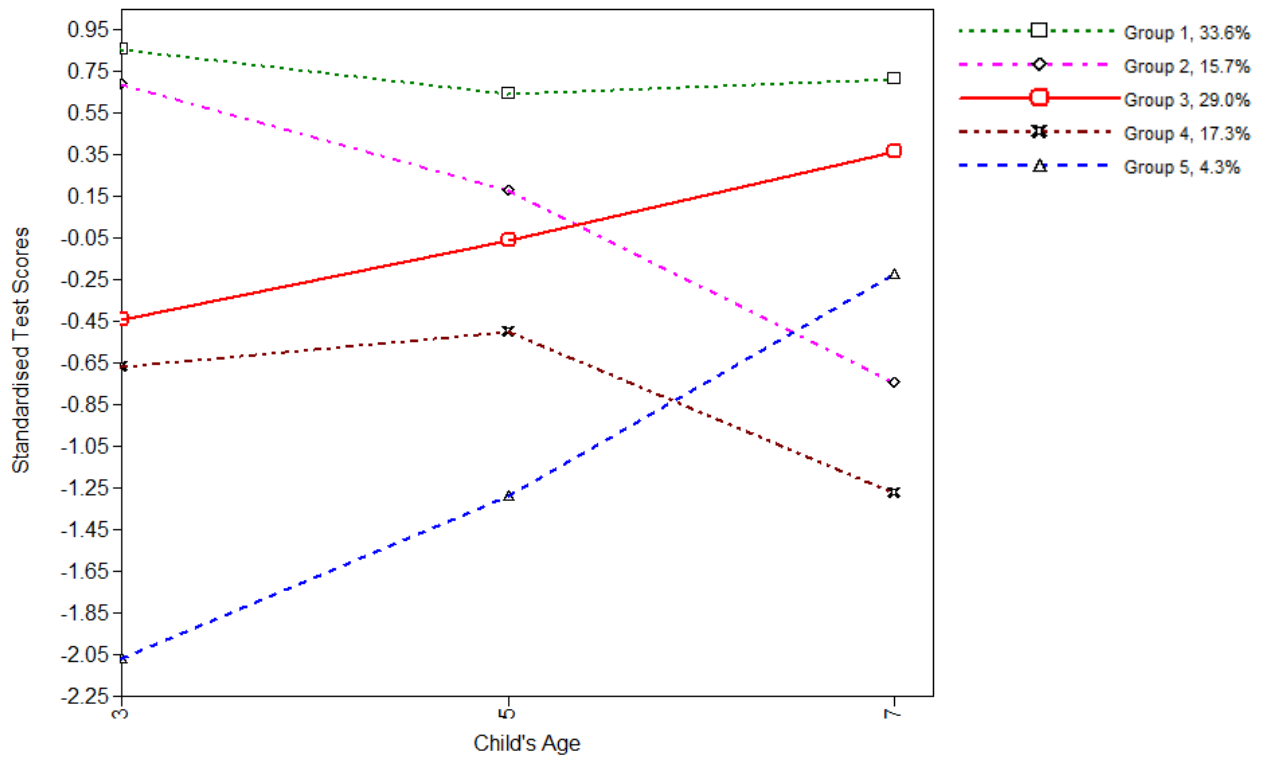
Figure 3. Latent trajectory plot for 5 class growth mixture model (MCS data)

Appendix

Table A1. Estimates (and standard errors) of intercept and residual variances from fitting two-class GMMs to simulated data. The models assume quadratic growth and random intercepts within classes.

k	$\sigma_u^{2[k]}$		σ_{e1}^2		σ_{e2}^2		σ_{e3}^2	
$\theta_1 = 1.4, r = 0.75$								
1	0.418	(0.058)	0.348	(0.006)	0.343	(0.006)	0.360	(0.005)
2	0.396	(0.059)						
$\theta_1 = 3, r = 0.75$								
1	0.237	(0.010)	0.197	(0.003)	0.194	(0.003)	0.202	(0.003)
2	0.240	(0.010)						
$\theta_1 = 3, r = 0.4$								
1	0.128	(0.019)	0.594	(0.009)	0.585	(0.009)	0.612	(0.009)
2	0.107	(0.019)						

Table A2. Comparison of estimates of the within-class variance from the GMM with the conditional variance of the standardised response y_{ij}^* given x_j under the data generating model.

θ_1	r	$\text{var}(y_{ij}^*\vert x_j)$	$\sigma_u^{2[k]} + \bar{\sigma}_e^2$	
			$k = 1$	$k = 2$
1.4	0.75	0.78	0.75	0.77
3	0.75	0.44	0.44	0.44
3	0.40	0.74	0.72	0.70

Note: Under the data generating model, the conditional variance of the standardised response given x_j is $\text{var}(y_{ij}^*\vert x_j) = \text{var}(y_{ij}\vert x_j)/\text{var}(y_{ij}) = (1 + \gamma)/(0.25\, \theta_1^2 + 1 + \gamma)$, from equation (5). The calculation of the within-class variance uses the mean of the residual variances across occasions, $\bar{\sigma}_e^2$.

Table A3. Growth parameter estimates of MCS 5 class growth mixture models

Groups	Intercept	Standard error	Slope	Standard error	Quadratic	Standard error	Estimated Posterior %
1	0.857*	0.017	-0.359*	0.033	0.144*	0.017	34
2	0.691*	0.028	-0.308*	0.073	-0.205*	0.045	16
3	-0.445*	0.016	0.367*	0.043	0.02	0.025	29
4	-0.668*	0.021	0.639*	0.051	-0.472*	0.027	17
5	-2.067*	0.085	0.637*	0.143	0.143*	0.068	4

*$p = <.05$

Table A4. Covariate coefficient contrasts of the MCS 5 class multinomial logistic model for trajectory group membership

	2 v 1	4 v 1	4 v 2	1 v 5	2 v 5	4 v 5
Standardised income	-0.244* (0.051)	-1.058* (0.091)	-0.815* (0.097)	1.694* (0.224)	1.45* (0.226)	0.635* (0.239)
benefit payments						
Yes	0.395* (0.111)	0.452* (0.093)	0.057 (0.122)	-0.414* (0.144)	-0.019 (0.155)	0.038 (0.151)
NS-SEC, ref: Managerial/prof						
Intermediate	0.318* (0.123)	0.392* (0.133)	0.074 (0.170)	-0.722* (0.24)	-0.404 (0.251)	-0.331 (0.269)
Self Employed	0.346 (0.21)	0.683* (0.204)	0.337 (0.265)	-1.174* (0.312)	-0.828* (0.337)	-0.491 (0.352)
Technical	0.785* (0.187)	1.093* (0.172)	0.308 (0.208)	-1.442* (0.283)	-0.657* (0.291)	-0.35 (0.302)
Routine	0.681* (0.12)	1.38* (0.118)	0.699* (0.151)	-1.769* (0.203)	-1.088* (0.214)	-0.389 (0.226)
Marital Status ref: single						
Married	-0.109 (0.098)	-0.055 (0.086)	0.054 (0.108)	-1.881* (0.219)	-1.99* (0.221)	-1.936* (0.224)
Divorce/Separate	-0.074 (0.171)	0.038 (0.0136)	0.112 (0.169)	-1.002* (0.279)	-1.076* (0.287)	-0.964* (0.281)
Widowed	0.549 (0.714)	-0.076 (0.921)	-0.625 (0.996)	-2.544* (0.83)	-1.996* (0.831)	-2.621* (1.067)
Parent long term illness						
Yes	-0.015 (0.098)	0.037 (0.084)	0.052 (0.107)	0.331* (0.145)	0.316* (0.152)	0.367* (0.152)
Parent's Age at Birth: ref: Under 20						
20-39	-0.16 (0.215)	-0.163 (0.158)	-0.003 (0.170)	0.246 (0.296)	0.087 (0.291)	0.084 (0.284)
40+	0.053 (0.321)	-0.236 (0.304)	-0.289 (0.344)	0.331 (0.515)	0.384 (0.516)	0.095 (0.536)
Child's gender						
Female	-0.355* (0.1)	-1.043* (0.076)	-0.689* (0.112)	0.901* (0.118)	0.546* (0.135)	-0.143 (0.129)

*p<=0.05; Standard errors in parentheses