Inequality in the early cognitive development of British children in the 1970 Cohort.

Leon Feinstein *

7 December 2000

Abstract

This paper uses the 1970 cohort to develop an index of development for 1292 UK children assessed at 22, 42, 60 and 120 months. The paper discusses the importance of these early scores as measures of human capital formation and argues that they can provide insights for growth or labour economists as well as those concerned with social equity. Position in the distribution of this index at 22 months (in 1972/3) is shown to predict final educational qualifications at age 26 (1996). The position at 22 months is shown to be related to family background. However, the children of educated or wealthy parents who scored poorly in the early tests, had a tendency to catch up whereas children of worse off parents who scored poorly were extremely unlikely to catch up and are clearly shown to be an at-risk group. As children mature and do more discriminating tests, the family background association strengthens.

Tony Blair has famously made education the priority of his Government and it is clear that human capital production plays a central role in the Government's thinking about inequality. It is well understood that the Government wishes to reduce income inequality through reducing educational inequality. However, social and family background factors influence or are associated with the development of children before they have entered school or, even pre-school. Liaw *et al.* (1994), for example, show that "at risk factors", such as family mental health or problem behaviours related to poverty, influence the IQ of children as young as age three. Klebanov *et al.* (1998) show that these risk factors influence the development of North American one-year olds and that, moreover, poverty significantly affects children by age two. By age three, even neighbourhood effects have played a significant role.

At face value, this evidence suggests that educational interventions after children have already entered school may come too late. Family background or genetic factors have already played their role in generating intergenerational inequality. However, such a conclusion would place a strong weight on early IQ and developmental scores, and it is important to know to what extent

* University of Sussex and Centre for Economic Performance, London School of Economics. I would like to thank the ESRC Data Archive at Essex University for permission to use the BCS data and two anonymous referees for many helpful comments. Barbara Maugham of the Institute of Psychiatry also made many interesting and helpful points. I would like to thank James Symons, Steve Machin, Costas Meghir, Marco Manacorda, as well as participants at the Labour Economics Seminar at the CEP, the Econometrics Workshop, UCL and the Economics Subject Group seminar, Sussex University

early measures of ability are correlated with later ability and qualifications. This paper, therefore, attempts to answer three questions: a) whether indicators of pre-school development are associated with final, adult educational outcomes; b) whether performance in these pre-school indicators is stratified by social class; and c) how this stratification changes as children mature.

There are many studies in the literatures of developmental psychology, psychometrics and behavioural genetics that have sought to consider the descriptive questions addressed here, but none of which I am aware that have sought to do all three in a single, large and representative data-set. Studies have typically been on small or specific samples not representative of the general population, and most studies have been on American children for which the patterns of association with family background may be different to those for UK children. Moreover, the majority of early development studies, such as those of the High/Scope programme in the U.S., have been based on specially selected, non-random samples (Berrueta-Clement *et al.* (1984), Schweinhart, *et al.*, (1986) among others). This paper exploits the 1970 Birth Cohort Survey (BCS), a longitudinal data-set with unparalleled cross-sectional and longitudinal richness. Particularly useful are the tests of development given in two sub-samples, when the children were 22 and 42 months old. Selection was random subject to attrition and to the important restriction that all the children in these early sub-samples were from two-parent familes. This restriction is important but, nonetheless, the sample is wider and more representative than those mentioned above. Children from the full income range are included and the sample is national.

However, the results of this study would not surprise those working in the disciplines mentioned. Answers have already been developed to these three questions in other data that may lack the representativeness of the data used here, but I do not claim to add any striking, new information to those disciplines. Rather, the objective of this paper is to report these results to economists, through investigation of the nature of the stratification in the data set to be introduced and examined below.

One may ask why economists should be interested. I have already pointed to an important policy question but one might argue, more broadly, that it is important for economists to be aware of the process of human capital formation. Human capital plays an important role in explaining individual wages, and there have been many attempts to use endogenous human capital to explain growth at the macroeconomic level. It is likely that complexities in the process of human capital formation account for the failure of many human capital models of growth to be empirically validated. Human capital is not just the result of schooling investments but is formed through a series of genetic, parenting and wider social institutions. Better understanding of these issues might lead to more successful modelling. This paper does not attempt to address in detail the general question of how human capital is important for Labour or Macroeconomics but describes one of the complexities in its formation, namely the stability and meaning of tests of human capital development through childhood.

The answers given by this paper are descriptive; the paper does not make the excessively strong assumptions necessary to attempt estimation of the degree of state-dependence that would clarify somewhat the nature of the causal mechanisms underlying these patterns of association. The reasons for this are given in Section 1, which also describes the data and addresses the issue of the meaning of indices of childhood development. Arguments are put forward supporting the use of the first principal component as such an index. Since this is a fraught issue in psychological research there will be some discussion of debates in this area, followed by discussion of how the scores are modelled. Section 2 will assess the stability of this index as children mature, and Section 3 will assess the extent of social stratification. Family background variables may pick up genetic or environmental effects, so to facilitate interpretation of the meaning of stratification Section 3 includes a brief discussion of the wider evidence for each. Section 4 concludes and offers some avenues for future research.

1. The Data

Table 1 reports the ages at which the 1970 cohort have been sampled, together with sample numbers. Of particular value, here, is the data collected when the children were 22 and 42 months old. Due to medical concerns about the effect of fetal malnutrition on brain cell proliferation, a sub-sample of BCS children were studied at these ages. A ten percent random sample of all births was taken together with those children who were considered to be most at risk from fetal malnutrition. Numbers from each of these sub-groups within the 22 and 42 month sub-sample are given in Table 1.

Table 1: *Observations in first four sweeps of BCS*

	Total		Test scores	Test scores in common
Birth	17196			
22 month sub-sample,	2457		2436	
Random control sub-sample		1125		
At-risk sub-sample:				
Twins		228		
Post-mature babies		748		
Small-for-dates babies		567		
42 month sub-sample,	2315		2297	2045
Random control sub-Sample		1093		
At-risk sub-Sample:				
Twins		211		
Post-mature babies		676		
Small-for-dates babies		527		
5 years	13135		11738	1672
10 years	13871		12308	1292
26 years	9003		8395	

Notes: The primary objective of the data collectors was to study the effects of fetal malnutrition. Twins, post-mature and small-for-dates babies were included in the sub-sample because these groups were thought to be at risk. Post-mature denotes children born after term and at risk of fetal malnutrition because placental growth is said to cease from the 39th week of pregnancy. Small-for-dates babies are those showing evidence of fetal retardation at every gestational age. Groups are not mutually exclusive.

Although there were over 17000 children in the full cohort, this paper only makes use of the information about children in the pre-school sub-sample, 2457 children. There is information about test scores at all four ages for 1292 of these children and this is the sample frame for the paper.

There are two main concerns about the sub-sample data. Firstly, the 10% random group suffered 27% non-response due to health visitor staff shortages, parental refusal or tracing difficulties. This would lead to concerns about attrition if response was non-random, perhaps if parents who cared little for their children were less likely to submit their children for examination. However, the Institute of Child Health compared the random group to its equivalent in the British Births Survey and found no differences at the 5% level in terms of mortality, birth-weight, length of gestation, social class, sex, maternal height, maternal age and the mother's country of birth. There is evidence from elsewhere in this survey that parental care is often found to be associated with many of these factors and so there are grounds for optimism that the control sample is representative of the wider population.

Second, inferences about mobility and explanation of test scores in the general population using the non-random components of the sub-sample are likely to be biased if fetal malnutrition is indeed linked both to the development of brain activity and so to performance in developmental tests, particularly since fetal malnutrition is also linked to unobserved aspects of family background. Two strategies are adopted. Firstly, analysis was undertaken on each sub-group separately to test whether results varied from those for the control group. Generally it was found that there was no significant model estimation divergence between the sub-groups and the control group, although the at-risk children did show evidence of slower cognitive development. Secondly, for regressions a weighted least squares procedure was used, with unreported sub-group dummy variables to control for membership of a particular sub-group. These two strategies mean that results can be considered to be representative of the educational development of the wider population of children.

One remaining sampling issue cannot be overcome. Only children from two-parent families were included in the sub-sample. This seriously limits the representativeness of these results, particularly for those concerned with family breakdown. Nonetheless, bearing this exclusion in mind, analysis of these data still sheds light on the questions of the importance and explanation of early ability differences between children of different backgrounds. Twenty-four children who were in special schools at age ten were also excluded from the subsequent analysis on the assumption that they represent particular educational problems.

1.1. Test scores

At each age BCS children were assessed by a wide range of tests of intellectual, emotional and personal development. The full list of tests is given in Appendix 1. At 22 months the children were asked by the health visitors administering the survey to complete a range of different tasks, for example: pointing to their eyes to illustrate understanding of language; putting on their shoes, indicative of personal development; stacking cubes and drawing lines as tests of locomotor ability. These tests, together with those at 42 months, were intended to indicate the general development of children based on the tests used for screening in child health clinics (Chamberlain *et al.*, 1976). A pilot study found high correlation between the BCS tests and similar, standard tests of development such as the Bayley Scale of Infant Behaviour or the Newcastle Survey (Neligan *et al.*, 1969). At 42 months counting and speaking could be tested, and further copying tests were administered such as drawing simple geometrical shapes. At age five copying was again assessed, together with tests of basic vocabulary. Harris (1963) and Koppitz (1968) show these scores to have good properties of discrimination and reliability. Standard age ten scores for maths and reading are also available.

All these scores are appropriate for the age of the children being tested. It used to be thought that cognitive development could not be tested before children were five years old (Bayley, 1949.) In fact, recent tests of attention or response to novelty in the first six months of life have been shown to be correlated with cognitive test scores in later childhood (Bornstein *et al*., 1986.) Nonetheless, intelligence changes qualitatively over early maturation. In a review of psychiatric research, Zeanah *et al.*, (1997) emphasise three periods of major structural reorganisations in infancy. The last of these qualitative shifts, involving the entry into verbal and symbolic representation, ends at around 20 months, after which changes can be more easily characterised quantitatively. At 22 months children will still be consolidating after the most recent shift but by 42 months will have the skills much more firmly at their disposal. More stability from 42 months might, therefore, be expected and development can be more readily assessed quantitatively. There is still considerable instability in scores because very young children do not stay on-task for long and because intelligence continues to undergo change. Therefore, the early tests are likely to contain some information about developing intelligence but a great deal of subsequent mobility is expected.

1.2. Principal components analysis and the development of an ability index at each age.

In order to maximise the information available at each age while reducing the number of dependent variables, test scores at each age were combined by principal components analysis. This technique is common in most behavioural and social sciences but is perhaps less well-known in economics so a brief review is provided in Appendix 2. Broadly, principal components analysis can be considered as the eigenvalue decomposition of the correlation matrix, R. The first principal component is given by:

$$y_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1p}x_p = \sum_{i=1}^{p} a_{1i}x_i$$

where $x_1,\ldots, x_p$ are the variables to be combined, in our case test scores. The weights a_{11}, …, a_{1p} maximise the variance of y_1 and satisfy the normalising constraint:

$$\sum_{i=1}^{p} a_{1i}^2 = 1$$

It is possible to calculate as many components as there are test scores, but subsequent components must be uncorrelated with previous components and will account for less of the variation in test scores.

This method has the virtue of combining scores into a single index of development at each age that is easier to understand and use in subsequent analysis than the full set of scores. It is appropriate here because, as Appendix 1 shows, the test scores are sufficiently correlated to support the hypothesis that they are measuring manifestations of a similar process but sufficiently distinct that each contributes valuable information when they are combined.

There is no assumption here that this index identifies any uni-dimensional notion of intelligence such as Spearman's g, which is central to psychometric studies and has aroused considerable controversy. Charles Spearman, a British psychologist, noted (Spearman, 1904) that people who scored well on intelligence tests usually did well in all cognitive areas — whether verbal, mathematical or spatial in nature. He hypothesised that some general or g factor contributes to this success[1].

Recently Jensen [1998] has claimed that the g-Factor is a biological phenomenon and, crucially, that, therefore, individual and population (i.e. race) differences are a function of evolutionary processes. Jensen shows that psychometric g has direct biological correlates with brain size, brain evoked potentials, nerve conduction velocity, and the brain's glucose metabolic rate during cognitive activity. However, there is far from general agreement among neurologists that the science has evolved sufficiently to support the claim that these variables can be invoked as an explanation of g, or that g is, therefore, a biological variable.

While some neuroscientists are attempting to use brain imaging techniques to locate the part of the brain responsible for success in tasks of g-intelligence, other neuroscientists and psychologists are concerned that the psychometric model describes the results of statistical analyses without

[1] Factor analysis differs from principal components analysis in that the former isolates only the common variance of variables. Whereas principal components analysis decomposes R, factor analysis decomposes the reduced correlation matrix R-U where U is a diagonal matrix of the unique variances associated with the variables. The underlying assumption is that each variable in the variable set can be expressed as a linear combination of

explaining what the abilities are. The answer to the question, what is g? is that it is that which correlates with lots of different tests of intelligence. Modern theories of intelligence, such as the multiple intelligences of Gardner (1983) or the triarchic model of Sternberg (1985), are more concerned with understanding the different aspects of intelligence, such as practical, analytical and creative intelligences in the triarchic model, and considering how these different aspects interact. L.L. Thurstone, a follower of Spearman, recognised the problem for g-intelligence in his 1934 Presidential address to the American Psychological Association:

> there can be debate as to whether we *should* describe the tests by a single factor even though one factor is sufficient. It is in a sense an epistemological issue. Even though a set of intercorrelations *can* be described in terms of a single factor, it is possible, if you like, to describe the same correlations in terms of two or three or ten or any number of factors[2].

This epistemological question can be seen as one of identification, but whether or not one wishes to identify the first factor with intelligence, it is clearly a useful statistical artefact. Since it is not the purpose of this paper to test for hypothetical commonalities but rather to consider the importance of early signals of development, principal components analysis is preferred to factor analysis. As has been said, principal components analysis maximises the variance of the underlying data, in other words its signal, thus giving the early scores the best chance to predict later outcomes. In any case, I have no grounds on which to dismiss the variance unique to any particular score as factor analysis would do. The analysis is, therefore, less in the tradition of Spearman than might at first appear, although the first factor and first principal component will in any case converge if a sufficient number of postively correlated variables are combined. More importantly, this paper does not make any claim that the particular first component index used here identifies any biological entity or that intelligence is uni-dimensional.

---

hypothetical, unobservable common factors plus a factor unique to that variable. Factor analysis extracts linear combinations based on these hypothesised commonalities, such as, for example, Spearman's g-Factor.

[2] Thurstone, L.L.(1934) "The Vectors of Mind. Address of the president before the American Psychological Association, Chicago meeting, September, 1933." Psychological Review, 41, 1-32

1.3. Modelling the distributions of abilities.

A large degree of instability in scores may be expected because of qualitative changes in the ability being proxied at different stages of development and because the growth rate for cognitive abilities is not common for all children . A child whose IQ score remains the same throughout childhood does not exhibit the same performance at 6 and 16. Steady gains in ability will be observed but the relative performance is constant. Conducting analysis on rank position rather than actual scores increases stability and that is the procedure followed here. This also makes sense because our concern is with educational inequality. The children are ranked in normalised reverse order, a rank of one for the lowest scoring child and a rank of one hundred for the child scoring highest. This gives four outcome variables that reflect children's position in the distribution of observed development at the ages of 22, 42, 60 and 120 months. Although the rank varies between one and one hundred there are potentially as many positions within this range as there are children in each sweep who completed the tests.

We can write:

$$R_m = \alpha_m' F_m + \varepsilon_m , \qquad m=22,42,60,120 \qquad (1)$$

where F_m is a matrix of family inputs and the rank position are subscripted m rather than t to emphasise that even when combined through the use of principal components the scores do not represent movement along a single axis of ability over time. The development of attainment through childhood is clearly not akin to a time-series of, for example, individual wages because the variable itself changes as children mature. Intellectual and behavioural development is qualitatively different at each age. Different abilities are tested and these are not necessarily functional equivalent. It would be uninformative to test, say, reading skills at 22 months or, conversely, block stacking at 60 months because of the qualitative change in children's abilities. Functional equivalence for reading would demand tests of abilities at 22 months that fed particularly into reading ability but the tests used here have not been devised in that way. (See Table 2 and related discussion, below.)

Thus, although the rank positions at different ages are related, this paper does not attempt parametric estimation of such relationships, considering instead the mobility of rank positions and the association at each age with the elements of F_m. It may be econometrically tempting to employ panel data techniques that treat rank positions at different ages analogously to, say, a panel of wages and ignore the qualitative change in development in childhood. Such an approach would consider changes in rank positions but would beg the question of what was undergoing change. In these circumstances, it is not clear that inference would be meaningful.

Instead, the R_m are treated as samples of observations of 4 different random variables so that (1), therefore, describes 4 different equations. The position in the distribution of abilities at each age is a linear function of family inputs up to that age. This is standard in the economics of education following the Coleman report (Coleman et al., 1967), based on theoretical foundations provided by the education production function.

1.4. The relations between underlying test scores at different ages.

Principal components analysis is appropriate here because there is no particular connection between scores in tests of specific abilities at early ages and subsequent performance in more demanding tests of abilities, in the same dimension of ability. Rather, early test scores, particularly those at 42 months, appear to prefigure later ability, but in a more general manner. This can be observed in Table 2, which addresses the issue of how particular early scores prefigure later scores for particular aspects of ability. The table reports the raw correlation of the individual components of each age's test score with test results at age 10.

Table 2: *Raw correlation of individual test scores with scores at 120 months*

	120 months	
	Reading	Maths
22 month scores		
Cube stacking	0.21	0.11
Language use	0.22	0.12
Personal dev.	0.20	0.13
Drawing	0.15	0.14
42 month scores		
Counting	0.29	0.13
Speaking	0.28	0.17
Copying designs I	0.32	0.16
Copying designs II	0.27	0.14
60 month scores		
Copying designs	0.40	0.19
Vocabulary	0.40	0.18
Human Figure Drawing	0.31	0.13

It is well known that reading scores are more easily predicted than maths scores (for example, Stevenson *et al.*, 1986). Table 2 also shows that no single early test dominates in terms of degree of association with subsequent scores, even, as with age 10 reading and age 5 vocabulary, in what one might expect to be related subjects. As a second example, cube stacking and language scores at 22 months are equally associated with reading at age ten and this is not due to differences in the variance of the two 22 month variables. There is, therefore, no evidence that any single test score is an obvious candidate to proxy development.

The use of principal components allows us to exploit all of the information available in the test scores without discriminating on unsupported a priori grounds between different tests. Since there is no evidence of functional equivalence, we also have empirical justification for treating the set of observations at each age as observations from different populations, i.e. as different random variables.

1.5. The latent vectors.

Table 3 reports the latent (weight) vectors for the first and second principal components at each age. From the discussion of Principal Components Analysis, above, it should be clear that the squared weights sum to unity.

Table 3: *Latent vectors from principal components analysis of test scores, first and second factors*

	22 month		42 month		5 years		10 years	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd
Cube stacking	0.34	0.54						
Language use	0.37	-0.64						
Personal dev.	0.40	-0.38						
Drawing	0.33	0.63						
Counting			0.37	0.36				
Speaking			0.34	0.68				
Copying designs I			0.39	-0.32				
Copying designs II			0.35	-0.70				
Copying designs					0.48	-0.26		
Vocab					0.40	0.95		
Human Figure Drawing					0.46	-0.56		
Reading							0.34	-0.04
Maths							0.27	0.67
Picture language							0.29	-0.47
British Ability Scale							0.34	-0.09
Eigenvalue	1.92	0.80	1.90	0.85	1.65	0.77	2.58	0.68
Proportion of variance	0.48	0.20	0.48	0.21	0.55	0.26	0.64	0.17

Notes: The Table reports 2 latent vectors at each of the 4 ages for which test scores are available. Principal components analysis was undertaken using the principal components factor method based on analysis of the correlation matrix, R, of test scores at each age. The i^{th} latent vector provides the weights for the i^{th} principal component and together with the latent root, l_i, is the solution to $Ra_i = l_i a_i$ where $a_i'a_i=1$ and, A, the full matrix of latent vectors, satisfies A'A=I. The latent root or eigenvalue, l_i, is the variance of the i^{th} component. The

proportion of variance in the final row is derived as the latent root of the ith component divided by the sum of the variances of the underlying variables in the linear combination. Since these are all standardised, the proportion of variance explained by each principal component is given by l_i/p, where p is the number of underlying variables.

It can be seen that no individual test dominates the first principal component but that at each age, all scores contribute to the variation of components. This follows from the fact that the elements of the correlation matrices at each age are exclusively positive and reasonably similar. The second components explains additional variance of the underlying scores but are not interpretable as general indices of development since, under the stated assumption of zero covariance between first and second components, one or more of the elements in the second components must have a negative weight. The second components are not used in subsequent analysis but they are nonetheless described in Table 3 because they help clarify some of the features of the underlying data. It may be observed, for example, that the second component at 22 months contrasts the motor and spatial skills of cube stacking and drawing on the one hand with the more intellectual/behavioural skills of language use and personal development on the other. This can be thought of as a specifically intellectual rather than spatial component. Its eigenvalue is 0.8, indicating that this component explains 20% of the variance of the four standardized scores at 22 months. Jensen finds that the first factor is most correlated with neurological proxies of intelligence and uses this to support his argument that g is a biological entity. It may be noted that, here, the second component is not significant in regressions of the first principal component factor at age ten on age 22 factors. In fact, of the second factors at 22, 42 and 60 months, only the age five second factor, which emphasises vocabulary at the expense of human figure drawing tests, is a significant predictor of age ten development. This is unsurprising given that second components have lower variance than first components.

1.6. The measure of educational outcomes at age 26.

The final educational variable used is an ordinal measure of highest educational qualification as reported in the age 26 sweep of the BCS and coded as a stripped down version of the more detailed Schmitt schema (Schmitt, 1993) which gives an ordinal scale of educational/vocational attainment, based around years of education. The three groups with sample proportions are

None/miscellaneous (14.4%), Lower/Middle vocational (46.2%) and A'Level or above (39.4%). This ordering has been shown to be strongly linked with earnings in these data (see Feinstein, 2000).

2. The strength of the signal provided by the development indices

2.1. Predicting final educational qualifications

Perhaps the clearest picture of the relevance of the early position in the index of development is given by the results in Table 4 which show how the position in the distribution at each age predicts final qualification level at age 26. Taking the 22 month panel first, the first row reports the highest age 26 qualifications of the children who were in the bottom quartile at 22 months. For example, 15% of the bottom quartile at 22 months, obtained no or miscellaneous qualifications. The third row reports that of those who were in the top quartile at 22 months, 8% ended up in the lowest qualifications group.

Table 4: *Age 26 educational and vocational qualifications by quartile position in early development scores.*

		Age 26 Highest Qualifications			
		None/ Misc.	Lower/ Middle	A'Level or higher	Total
22 month rank					
Bottom Quartile	%	15.00	52.80	32.30	100
	(s.e.)	(2.20)	(3.10)	(2.90)	
Top Quartile	%	8.10	48.60	43.30	100
	(s.e.)	(1.50)	(2.80)	(2.80)	
z-stat on difference*		2.6	1.0	2.7	
42 month rank					
Bottom Quartile	%	25.80	57.30	16.90	100
	(s.e.)	(3.00)	(3.40)	(2.60)	
Top Quartile	%	6.50	41.00	52.50	100
	(s.e.)	(1.30)	(2.70)	(2.70)	
z-stat on difference*		6.4	3.7	8.4	
5 Years rank					
Bottom Quartile	%	30.20	51.80	18.00	100
	(s.e.)	(1.40)	(1.50)	(1.10)	
Top Quartile	%	5.20	36.40	58.50	100
	(s.e.)	(0.50)	(1.10)	(1.10)	

z-stat on difference*		18.8	8.3	21.8	
10 years rank					
Bottom Quartile	%	35.00	53.50	11.50	100
	(s.e.)	(1.40)	(1.50)	(1.00)	
Top Quartile	%	2.40	30.70	67.00	100
	(s.e.)	(0.30)	(1.00)	(1.00)	
z-stat on difference*		25.5	12.6	29.8	

[*] The final row for each panel reports a test statistic for the difference between cell proportions. This has a standard normal distribution, under the null. The z-statistic on the difference in proportions in the first column of the first panel is 2.6, i.e. the difference is significant at 1%.

The z-statistic in the final row of the panel is a test of the null hypothesis of equality between these two proportions, the proportions of top and bottom quartile children ending up in each age 26 qualifications group. Under the null hypothesis, the z-statistic has a standard normal distribution.

A chi-squared test was used to check whether results were biased by over-sampling of low birth-weight and foetally under-nourished children. This was a goodness of fit test of the difference between proportions in the Schmitt scale for the control group and full sub-sample including over-sampled groups at each age. The data do not reject the null hypothesis of uniformity between samples at any age.It is striking that even measured at 22 months, children in the bottom quartile of this development index are significantly more likely not to get any qualifications than those in the top quartile. Moreover, more than three times as many of those in the top quartile at 42 months as those in the bottom quartile go on to get A'Level qualifications or above. Given the young age of the children tested, these are strong findings, suggesting that the index picks up clear signals of educational development; before children have even entered school, very substantial signals about educational progress are contained in standard tests of development.

2.2. Transition matrices.

A second approach to the initial question of the stability of the distribution of scores as the children develop is to consider transition matrices. These group children by their quartile position at each of two ages, giving a table of sixteen cells. Given the large degree of instability in scores at these ages, it is not obvious that movement to cells adjacent to the leading diagonal are informative of genuine mobility. Perhaps more interesting are movements from top to bottom and vice versa. Therefore, only these large movements are shown in Table 5. The top panel of Table 5 shows movements from the quartile position at 22 months, the bottom panel shows movements from the 42 month position.

Again, as with Table 4, it might have been expected that the degree of movement observed would be affected by the over-sampling of children at risk from fetal under-nourishment. If such children were hindered in early years but subsequently caught up, mobility would be over-stated in this sub-sample relative to that in the population. On the other hand, if such children were persistently affected, mobility might be under-stated. Chi-squared tests for contingency tables have been applied and are presented in Table 5. These suggest, as before, that there is no significant difference between the transition matrices for the full sub-sample and those for the control group. Other experiments were undertaken with mobility indices such as those of Bartholomew (1973) which weights cells by their distance from the leading diagonal, a high overall score indicating a large degree of mobility; or of Shorrocks (1978). These also showed that the mobility results described in the text are not substantially altered by over-sampling.

Table 5: *Selected cells from quartile transition matrices*

		Quartile at 42 months		Quartile at 10 years		
		Bottom	Top	Bottom	Top	Obs
Quartile at 22 months	Bottom	39.9	13.7	41.5	15.4	304
		(1.4)	(1.0)	(1.4)	(1.0)	
	Top	10.8	43.5	13.4	34.6	306
		(0.9)	(1.4)	(1.0)	(1.4)	
	Chi-squared (dof=15):	2.5		8.7		
				1st	Top	
Quartile at 42 months	Bottom			44.8	10.1	306
				(1.4)	(0.9)	
	Top			6.2	43.8	306
				(0.7)	(1.4)	
	chi-squared (dof=15):			10.8		

Notes: Standard Errors are in brackets. Only extreme quartile cells are reported, i.e. top and bottom quartiles. The reported chi-squared test is a test of the difference between transition matrices of the control group and full sub-sample. The critical level at 5% with 15 degrees of freedom is 25.0

The first row shows that of the 25% children scoring lowest at 22 months, 39.9% were still in the lowest quartile at 42 months. On the other hand, 13.7% had entered the top quartile. Clearly there is considerable movement within the distribution over these twenty months. By 120 months, even more children had made large movements across the distribution. From these sample data, a child in the bottom quartile at 22 months would have a probability of 0.42 of being in the bottom quartile at 10 years but a probability of 0.15 of reaching the top quartile by then.

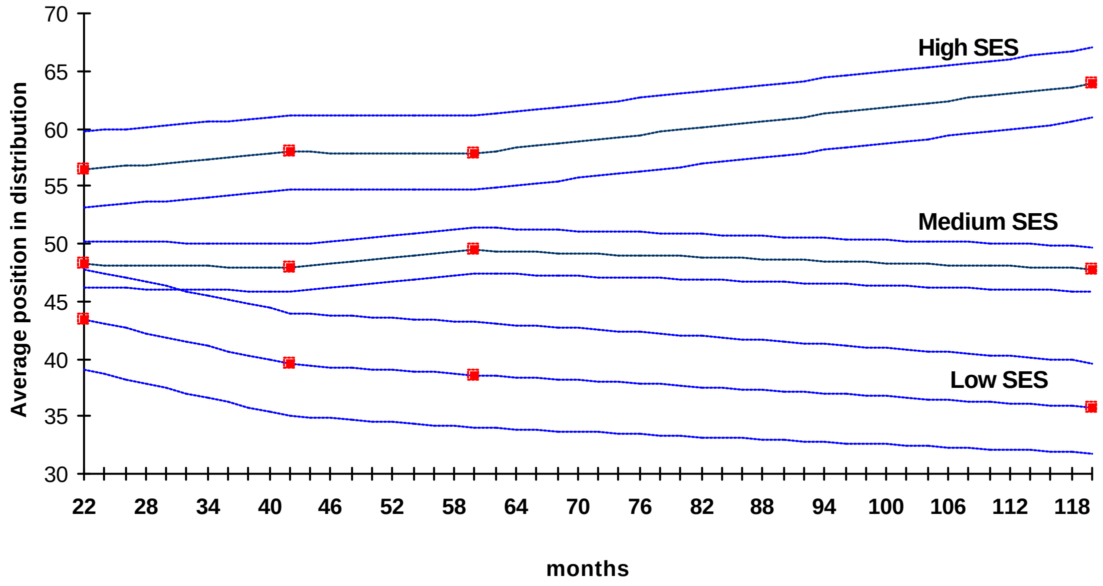
There is more clear persistence of scores between 42 months and ten years, particularly in terms of the proportion of large movements. Thus, as expected, the position at 42 months seems to be more firmly fixed than that at 22 months. However 10% of the bottom group at 42 months had reached the top quartile by age ten. This emphasises the interpretation of the development indices as signals of development and not as stronger classifiying mechanisms. Plenty of scope still remains for children to catch-up and overtake other children who may be out-performing them early on. As we see, below, high SES children who under-perform early on are very likely to catch up in this way. Nonetheless, the 22 and 42 month scores provide a meaningful guide to subsequent performance. The development index at 42 months is the preferred indicator. Other experiments have shown that for girls it is slightly more stable than for boys .

3. The association of test rank with social class

Figure 1 maps the average position of children from different social backgrounds in the distribution of test ranks at the four survey ages. Social class classifications are made here on the basis of both parents' occupational classification (Socio-Economic Status, SES) at the child's birth. It should be remembered that all the children in this sample are from two-parent families. Details of categorisation are given in the notes to the Figure. No allowance is made for changing occupational classifications over time because it is not possible to differentiate between genuine changes and mis-coding. In any case, social class at birth provides a good indicator of the material, genetic and educational inputs that the children can be expected to receive through childhood.

Observations are only made at 22, 42, 60 and 120 months. As noted above, the sample is restricted to those 1292 observations for whom test scores are available at all ages. This increases the standard errors of differences between groups because we discard all age 5 and 10 observations that are not included in the pre-school sub-sample. The advantage is that we can observe the average rank of three fixed groups of children as they mature.

Figure 1: Average rank of test scores at 22, 42, 60 & 120 months, by SES of parents



Dotted lines represent intervals of two standard errors. The definition of categories with sample observations are as follows: High SES – Father in professional/managerial occupation and mother similar or registered housewife (307 obs.) Low SES – Father in semi-skilled or unskilled manual occupation and mother similar or housewife (171 obs.) Medium SES - Those omitted from the high and low SES categories (814 obs.) Thus, children whose mothers were housewives were categorised by the SES of fathers.

At 22 months, the difference between the average rank of children in top and bottom social class groups is 13 percentage points (standard error of difference=2). At 10 years the difference between top and bottom groups is 28 points (s.e., 2.5). The average rank of the low SES group falls over time but this does not mean that actual development has been retarded. The interpretation of this change as one of increasing polarisation must be tempered by the fact that the ranks are positions within the distribution of different tests at the two ages. The finding may be explained either by declining relative performance or by the hypothesis that the later tests are more effective at discriminating between children. It may also be that the later tests assess tasks at which children in the low SES group are less able.

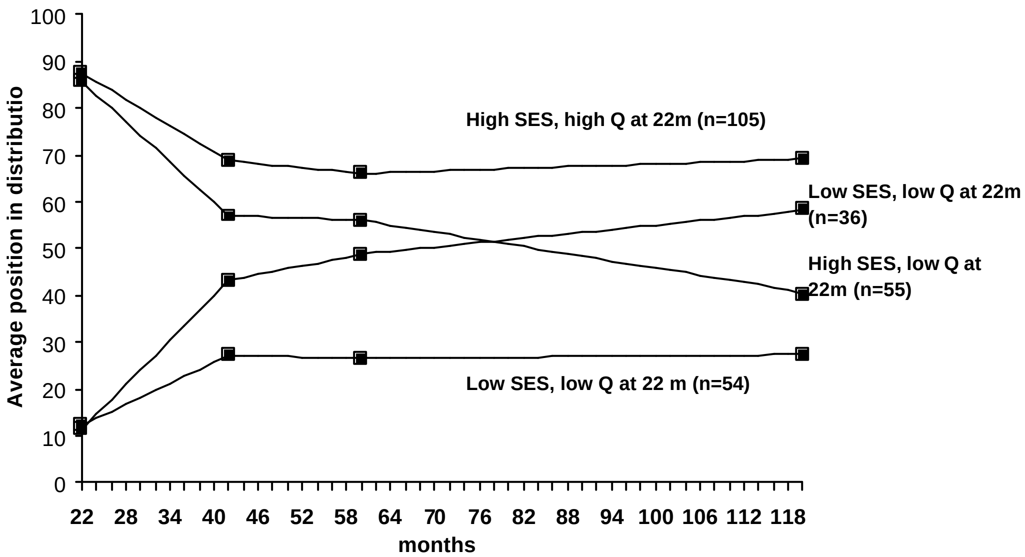
However, in a seminal paper, Wilson (1983) considers the congruence of IQ scores of twin pairs in the Louisville Twin Study between the ages of 3 months and 15 years. For identical (monozygotic) twins, the correlation between scores at 24 months was 0.81. For dizygotic twins the correlation was 0.73. By age 15, the correlation coefficients were 0.88 and 0.54, respectively. Thus, as children matured the genetic component of performance in the tests came to the fore. This may either be because the later tests were better tests or because, as Wilson argues, as with the development of height, the action of the genes is not completed until adulthood.

The pattern of polarisation here is, therefore, not surprising. Of course, there is considerable dispute between psychologists and behaviour geneticists about the relative causal importance of genes and environment and this is not the place to enter that debate. For balance, however, it is important to point out, firstly, that identification of this correlation with genetic rather than environmental causation hinges crucially on the "equal environments" assumption, that relevant environmental variation is the same within identical and non-identical twins. Second, Plomin (1984) demonstrates that twin and adoptee studies consistently show that a substantial proportion of the variance in normal psychological traits is due to non-genetic factors. Third, there is considerable evidence that the parenting methods and social institutions that the children will have experienced are very different (for example, Bee, 1969, Ramey, Campbell et al., 2000.) Finally, one need only point to the steady and increasing world-wide increase in IQ scores, the well-documented and fairly universal "Flynn effect" (Flynn, 1987) for evidence of environmental effects on performance on tests of cognitive ability. In conclusion, the evidence in this paper is of a gene-environment correlation and there is no case here for extending the conclusions beyond that point.

Crucially, the graph clearly shows that although children are already stratified by social class in standard tests of intellectual and personal development at 22 months, this stratification has become more extreme by 10 years as assessed by the standard tests for academic development appropriate at that age. There is certainly no evidence here that entry into schooling in any way overcame the polarisation of children in the late 1970s. The most generous statement that may be made for schooling is that it may or may not have minimised the deepening effects of parental background.

It should be remembered, however, that Figure 1 shows the mean rank positions within each of three groups of children, as they mature. There is, however, considerable and important within-group variation. This is brought out in Figure 2 which groups children not just by their family background but also by their 22 month quartile position. The advantage over Figure 1 is that this graph shows something of the distribution within the SES groups.

Figure 2: Average rank of test scores at 22, 42, 60 & 120 months, by SES of parents and early rank position



The definition of SES categories is as for Figure 1 with medium SES children omitted. Children in the 2nd and 3rd quartile at 22 months are also omitted. Standard error intervals are not shown to ease clarity of exposition. Important details of significant differences are given in the text.

Having a low score at 22 months doesn't matter greatly for future performance in test scores unless you are low SES as well. If you are, the implications are serious. However, if you are low SES, even having a top quartile score at 22 months, won't help you all that much. It is tempting to conclude from this that it is SES, rather than the early scores which makes the difference. This is supported by a consideration of the separate transition matrices of low and high SES kids. 60% of low SES children who were in the bottom quartile at 22 months were still there at age ten. On the other hand, high SES kids who happened to be in the bottom quartile at 22 months, were more likely to be in the top quartile at 10 years than to still be in the bottom quartile!

Does this suggest that early scores don’t matter? The answer is no, for two reasons. Firstly, from Figure 2, it is still the case that children within each of the SES groups who are in the top quartile at 22 months, score better at ten years than children in the same SES group who were in the bottom quartile at 22 months. The difference is still thirteen points at age ten for the low SES group and eleven for the high SES group, differences that are significant at 1%. For the omitted middle SES group, there is also some convergence over time for the high and low quartile groups at 22 months but although by 42 months the difference had fallen to 26 points in the distribution, at age ten it was still 22 points.

Second, we can also reconsider Table 5. This showed the probabilities of obtaining final educational qualifications in the abridged Schmitt range on the basis of quartile position at the different ages. For the middle SES group (the majority of children), children in the bottom quartile at 22 months are significantly more likely to get no qualifications than children in the top quartile and significantly less likely to get A'Levels or higher qualifications. For the top and bottom SES groups, differences at 42 months predict final educational qualifications. So, conditioning on SES, the pre-school score still matters. Nonetheless, as well as influencing early ability, family background clearly plays a tremendously important role in determining the continued development of ability of UK children.

It might appear to be a natural extension of the conditioning process to consider an ordered probit regression of the age 26 Schmitt variable on the rank at 22 months and SES dummies and other family background variables. However, we expect from equation (1) that the Schmitt position be indicated by 22 month rank partly because the 22 month rank picks up SES effects. Since the intention is to test whether or not the 22 month score is an indicator of real development, that is precisely the point. Even if the 22 month rank picked up only SES effects and nothing else, it would still indicate development. The problem would, rather, be in the opposite direction, if the rank position did not pick up SES effects at all which would suggest that it was a poor indicator[3].

More interesting is Figure 2b in the Appendix which repeats Figure 2 but classifying by quartile position at 42 months instead of that at 22 months. This again shows the importance of SES but also that, conditioning for SES, the pre-school quartile positions provide a good guide to age ten outcomes. The key difference is that whereas the high SES/low quartile group at 22 months had overtaken the low SES/high quartile group by age ten, when the conditioning is by 42 months quartile, the high SES/low quartile group are still significantly below the low SES/high quartile group by age ten.

The question set in the introduction was whether or not it matters for future development that disadvantaged children under-perform their peers in pre-school tests or whether such tests are poor indicators. I now answer by saying that the tests are clearly not unrelated to final outcomes. SES continues to be important as children mature and outweighs the importance of 22 month scores for high and low SES children but for all children, the early score is a guide, together with family background information, to final educational qualifications and academic performance. The lesson for policy makers is clear from Figures 2 and 2b. There is mobility (as one would expect) after 22 or 42 months, but this is mainly for high or medium SES children. Low SES children do not, on average, overcome the hurdle of lower initial attainment combined with continued low input. Even high SES children find it hard to escape from poor performance at 42 months.

3.1. The importance of different aspects of social class

The ability trajectories show that as children mature and do more discriminating tests, the family background association strengthens. Figures 1b-1d in Appendix 3 show that this result does not appear to depend on which conditioning variable is selected from the matrix of family background variables F. It remains when children are grouped by the education rather than SES of their parents or when they are grouped by the backgrounds of one parent only.

It would be interesting to know which aspect of family background dominates, either as a genetic marker or as a proxy for key environmental inputs. However, given the strong correlations between the elements of F, there is no unambiguous way to identify separate contributions to explaining the variation of test scores. For, continuous, independent regressors one approach is to consider the partial correlation coefficients. Together with simple coefficients, this gives a guide to the relative importance of each variable. In the current case, however, the regressors are a set of dummy variables and so the netting out process introduces further ambiguities that are not obviously resolved.

Neglecting, therefore, the importance of partial correlations, Table 7 reports OLS estimates of the α_m vectors from equation (1), which is reproduced here:

$$R_m = \alpha_m' F_m + \varepsilon_m , \qquad m=22,42,60,120 \qquad (1)$$

Weighted OLS is used to reduce the importance of over-sampled observations but none of the conclusions described depend on sampling bias, transformations of the data or problems of

---

[3] The reader might nevertheless be interested to note that, in fact, even conditioning on all the background variables in Table 6, below, rank at 22 months is still significant at 5%.

discreteness or censoring[4]. For the reported regressions, observations were grouped across genders. Experiments with running separate regressions did not bring to light any differences that are important to the conclusions here.

Table 6: *Test ranks regressed on background variables at birth.*

	22 months		42 months		5 years		10 years	
	Est.	t	Est.	t	Est.	t	Est.	t
Father's SES								
3,4,5	2.4	0.8	-2.9	1.1	0.4	0.1	-7.2	2.9
6	2.1	0.4	-1.4	0.3	-5.9	1.1	-12.1	2.6
Father's highest qualification								
Vocational or other	3.0	1.0	3.4	1.1	-0.5	0.2	-0.1	0.0
O'/A'level, SRN or Cert. Ed.	3.0	1.0	3.9	1.4	7.3	2.7	8.4	3.4
Degree	5.1	1.5	3.5	1.0	9.0	2.5	12.8	4.3
Mother's SES								
3,4,5	-1.3	0.3	1.9	0.5	6.1	1.5	5.0	1.5
6	-0.9	0.1	-19.9	2.8	-12.8	1.5	-22.0	4.5
Housewife	-1.9	0.5	1.5	0.4	5.0	1.2	3.0	0.8
Mother's highest qualification								
Vocational or other	3.1	1.0	4.3	1.4	1.7	0.6	4.2	1.6
O'/A'level, SRN or Cert. Ed.	7.9	2.8	7.9	2.9	8.0	3.0	13.1	5.5
Degree	21.4	3.4	20.5	2.6	19.8	2.8	25.2	4.2
Siblings								
1 older	-4.1	1.4	-0.8	0.3	-3.9	1.4	-8.9	3.5
2 older	-4.3	1.2	-5.5	1.5	-8.7	2.4	-13.9	4.2
3 or more older	-7.3	1.7	-13.3	3.0	-9.7	2.3	-18.2	4.9
1 younger			-3.0	1.1	-4.7	1.9	-3.7	1.6
2 or more younger			-14.2	2.1	-9.5	2.2	-6.9	1.5
girl	7.1	3.3	5.7	2.7	0.7	0.3	-1.8	1.0
Mother's age	0.2	0.6	0.4	1.5	0.5	1.9	0.7	3.4
Constant	-27.5	1.0	-82.4	2.0	37.7	4.7	38.9	5.3
obs		1194		1194		1194		1194
R^2		0.08		0.13		0.12		0.25

Notes: Observations are re-weighted by the formula $w_i = \hat{s}_i / s_i$ for i=1,2 where w_1 and w_2 are the weights of the control and at-risk groups in the early sub-sample, $\hat{s}_i$ is the number of observations of type i in the sub-sample predicted on the basis of the full sample proportions and s_i is the actual number of observations of type i. Controls for reason for inclusion in the sub-sample and precise age when test was taken are also included but are not reported here.

In this regression framework children for whom neither parent had qualifications were 38 points lower in the distribution at age ten than children whose parents both had degrees. Add in a

[4] If variation in the control group is higher than for the fetally undernourished groups then parameter estimates based on the latter groups might be biased downwards but the pattern of results described below changes very little if only control group observations are used. Inferences are also unchanged if the rank score

couple of older siblings and the effect rises to 52 points. Even at 22 months the effect of two parents with degrees is 26 points. This compares with negligible effects of SES until age ten. From 42 months, the association of mothers' social class group 6 and rank positions is strongly negative but it must be borne in mind that this is, in fact, the average association for a group of only 15 women.

The association with mothers' education is particularly striking early on and dominates the effect of paternal education but, again, this is partly due to the smaller numbers of women than men with degrees (13% of men as opposed to 2% of women). However, the mother's education variables are jointly significant at 1% while those of fathers are not jointly significant even at 20%. The father's education variables do not become significant at 5% until age 5 and the father's SES variables until age 10. In fact, the father's SES and education variables are not jointly significant at 5% until age 5.

Overall, it appears that for this sample the education of mothers is the best indicator of expected development and so may be most the useful variable in determining at-risk groups.

4. Summary and some directions for future research

This paper finds, firstly, that there were significant differences in the educational performance of children from different social groups in these data, even at 22 months. In this sense there was pre-school educational inequality in the UK between 1970 and 1975. Given that economic inequality has since worsened considerably, and public expenditure on education has not fully responded to this change, it is reasonable to assume that the degree of pre-school educational inequality has not improved. Second, performance in tests of ability at 22 months are correlated with ultimate schooling outcomes at age 26. The scores at 22 months are, therefore, meaningful measures of development in the sense that they provide real signals of development. This suggests that it would be wrong to ignore the evidence of inequality in pre-school scores solely on the grounds that the scores have no real meaning. However, the paper finds that 42 month scores provide a better guide than those at 22 months.

Third, family background plays a large role in influencing the mobility of children within the distributions of ability at different ages. Most low SES children who are in the bottom quatile at 22 months are still there at age ten. High SES children show considerably more upward mobility and

---

dependent variable is replaced by the continuous test score variables using tobit regression to correct for some are more likely to be in the top quartile than the lowest quartile by age ten, even if they were in the bottom quartile at 22 months. This might suggest, in a strong behavioural genetic interpretation, that early scores do not pick up the full effects of genetic inheritance and are poor guides to true potential. Alternatively, it may be that early scores are some guide to children's potentials but that environmental influences swamp the signal provided by these scores. It is not possible, here, to distinguish between these explanations. However, these results bring out the extent to which the formation of human capital in the UK is influenced by family background. It would be very interesting to know how much these associations are reproduced elsewhere, in countries with perhaps less, or more, social inequality.

Considering the implications for macroeconomics referred to in the introduction, it would also be interesting to know how national differences in the production of human capital due to social inequality, influence growth. One other issue for growth emerges from the result that parents education is, perhaps, the key factor in the formation of the human capital of children. It may prove interesting, therefore, to model long-run growth to see the extent to which appropriately lagged increases in educational investment increase the output of succeeding generations. It may prove beneficial to model human capital by the variables that explain performance in Table 6, rather than by years of schooling which has been shown to be a poor measure of genuine educational investment because of the wide differences in the quality of the schooling received.

In terms of labour economics, these results also show why family background explains earnings above and beyond that part explained by years of schooling. A large proportion of adults have no or low levels of qualifications. For these individuals, years of schooling provides no guide to formed human capital and it is necessary to assume that all individuals without additional education have identical human capital. Earnings differences are, therefore, be hard to explain and models incorporating family background as proxies of human capital formation may sometimes be more informative.

Returning to the general conclusions, the paper has also found that the early differences in attainment are not appreciably reduced by entry into the schooling system and become more extreme as children mature. However, it is not possible to conclude from this that by the time children enter school the position is irreversible. The test instruments change as children mature and

evident lower censoring which might also have caused downward bias.

the degree of stratification may be affected by this. It is not sensible, therefore, to consider changes in the rank position between stages of development as standard first differences.

In order to suggest that by the time children have entered school it is too late to intervene it would be necessary to show that the later scores were to a large extent state-dependent and not solely due to individual heterogeneity - in other words, that the subsequent positions were due to the early performance. If, rather, rank positions at different ages were correlated solely by virtue of some unexplained aspect of ability that was picked up by tests at all ages, and that was amenable to intervention at later stages, then clearly, there would be no over-whelming case for getting in early.

In the absence of alternative instruments, this identification could not be attempted without treating the scores at different ages as a panel which, with tests at sufficient ages, would allow one to include and instrument a lagged dependent variable in the education production function. Such an approach may be possible where tests are of a single entity such as maths skills and is also left to future research. Plewis (1996) provides a multi-level modelling framework for observations of test scores for a common attribute during development. Although this paper has not estimated such a model, the remainder of the remarks here will consider the relevance of state dependence and individual heterogeneity. It is hoped that these reflections will support understanding of the meaning of the persistence of test ranks.

A weak form of state dependence would result from the fact that children choose learning opportunities on the basis of previously accumulated skills. More advanced children are better able to select new learning tasks and to understand new tasks. A child who has trouble with counting will clearly find the transition to multiplication difficult if not impossible. Early educational inequality that was state dependent in this way would provide a barrier to the effectiveness of schooling interventions against adult inequality unless schooling interventions could address the nature of the barrier to future learning. This observation has led to the development of important learning programmes such as Rightstart in the US which was designed to remedy the observation that a significant number of low-income children started school without as good an intuitive understanding of number as their middle-income peers, and that this knowledge gap was both responsible for further falling behind and correctable (Griffin et al., 1994).

Some educationalists have pointed to the neurological finding that synaptic proliferation occurs mainly during the first three years to suggest that the pre-school era is a 'critical period' for interventions. This stronger form of state dependence is due to a classic paper by Wiesel and Hubel

(1965) whose animal experiments showed that kittens who had one eye covered from birth to three months had insufficient synaptic development in the areas of the brain concerned with sight. The sight of older cats was not persistently affected by the same visual deprivation imposed at a later age. Thus, the first three months were a critical period for the development of brain activity required for sight. However, it is a long step from these animal experiments to the conclusion that there is no brain plasticity in humans after the first three years. For example, Pascual-Leone *et al.* found that adults given a two-week piano training course experienced an enlargement of the area of the brain concerned with control of the fingers. There may well be state dependence in the weak form suggested above, based on the idea that development is cumulative. From this it follows that the early years are particularly important. However, the case should not be over-stated.

Alternatively, explanations of rank positions by individual heterogeneity would suggest that early rank positions were not immutable if aspects of the individual's experience were sufficiently open to policy influence. This might not require that heterogeneity be fully environmental, as opposed to genetic. The classic case of this is phenylketonuria, a genetically determined condition that can result in a substantially lower IQ. The effects of the condition depend crucially on diet. An environmental manipulation that involves changing the amino-acid constitution of the diet normalises intelligence dramatically.

The policy problem, however, is to determine which aspects of early life are open to environmental influence. Research summarised by Waldfogel (1999) suggests that there is considerable room for optimism about intervention programmes, but that success cannot be had cheaply. A well-replicated set of randomised experiments in the U.S. (Ramey and Ramey, 2000) suggest that to be successful interventions must pay top salaries in order to recruit well-qualified staff and keep staff turnover to a minimum. They must also follow children over time because the benefits of programmes that start in pre-school but do not continue for at least the first two years of school are highly liable to decay.

However, other kinds of programmes about which there was considerable optimism such as two-generation interventions or programmes targetted at entire families, have not produced the hoped-for gains. For example, Barnett (1995) shows that if interventions do have positive effects on the performance of children this is not due to effects on parents. This conclusion has been replicated in many other studies.

Returning to the UK, the Blair administration has developed the £540 million Sure-Start programme to bring together child-care organisations so that communities have access to organised and co-ordinated systems of support. Professionals and carers are provided with evidence-based guidance about practice. The non-coverage of those who do not choose to get involved in programmes or do so only indirectly will clearly be a concern. It is also important to note that Sure-Start is an area-based intervention and will, therefore, competely miss those families that happen to live outside targetted areas. This suggests that, as well as through Sure-Start, such skills might be taught at school, rather than waiting until the period of compulsory schooling is over.

5. References

Barnett, W., (1995) "Long-Term Effects of Early Childhood Programmes on Cognitive and School Outcomes." *The Future of Children: Long-Term Outcomes of early Childhood Programs*, 5(3), 25-50. Los Altos, CA: Centre for the Future of Children, the David and Lucile Packard Foundation.

Bartholomew, D. (1973) *Stochastic Model for Social processes.* London: Wiley.

Bayley, N. (1949) "Consistency and Variability in the Growth of Intelligence from Birth to Eighteen Years." Journal of Genetic Psychology, **75**, 165-196.

Bee, H.L. (1969) "Social Class Differences in Maternal Teaching Strategies and Speech patterns." Development Psychology, **1**, 6

Berrueta-Clement, J., Schweinhart, L.J., Barnett, W.S., Epstein, A.S. and Weikart, D.P., (1984). "Changed Lives: the Effects of the Perry Pre-school Programme on Youths through Age Nineteen." Monograph of the High/Scope Educational Research Foundation. No. **8**.

Bornstein, M.H. and Sigman, M.D., (1986). "Continuity in Mental Development from Infancy." Child Development, **57**, 251-274.

Chamberlain, R. and Davey, A., (1976). "Cross-sectional study of Developmental Test Items in Children aged 94-97 Weeks: Report of the British Births Child Study." Developmental Medicine and Child Neurology, **18**, 54-70.

Coleman, J.S. et al. (1966) Equality of Educational Opportunity. Washington D.C.

DfEE, (1999). Sure-Start: A Guide for Trailblazers. DfEE, London

Dougherty, T.M. and Haith, M.M., (1997). "Infant Expectations and Reaction Time as Predictors of Childhood Speed of Processing and IQ." Developmental Psychology, **33**, 1, 146-155.

Feinstein, L. (2000) "The relative economic importance of academic, psychological and behavioural attributes developed in childhood." Centre for Economic Performance Discussion Paper, No. **443**, February.

Feinstein, L., Robertson, D. & Symons, J. (1999) "Pre-School Education and Attainment in the NCDS and BCS." Education Economics, Vol. **7**, no. 3, 1999.

Flynn, J.R.,(1987). "Massive IQ gains in 14 nations: What IQ tests really measure." Psychological Bulletin. **101**(2): 171-191

Gardner, H. (1983) *Frames of Mind: The Theory of Multiple*. NY: Basic Books.

Griffin, S., Case, R. and Siegler, R., (1994). Rightstart: Providing the Central Conceptual Prerequisites for First Formal Learning of Arithmetic to Students at Risk of School Failure in McGilly, K. (Ed.) Classroom Lessons: Integrating Cognitive Theory and Classroom Practices. MIT.

Hanushek, E.A. (1986) "The Economics of Schooling: Production and Efficiency in Public Schools." Journal of Economic Literature. **24**, 1141-77

Harris, D.B., (1963). *Children's Drawings as Measures of Intellectual Maturity.* New York: Harcourt, Brace and World Inc.

Jensen, A., (1998). The g Factor: The Science of Mental Ability. Praeger.

Kaiser, H.F., (1960). "The Application of Electronic Computers to Factor Analysis." Educational and Psychological Measurement, **20**: 141-151.

Klebanov, K.B., Brooks-Gunn, J. McCarton, C. and McCormick, M.C., (1998). "The Contribution of Neighbourhood and Family Income to Developmental Test Scores over the First Three Years of Life." Child Development, **96**, 5, 1420-1436.

Koppitz, E.M., (1968). Psychological Evaluation of Children's Human Figure Drawings. New York: Grure and Stratton.

Liaw, F. and Brooks-Gunn, J., (1994). "Cumulative Familial Risks and Low Birth-Weight Children's Cognitive and Behavioural Development." Journal of Child Psychology, **23**, 360-372.

Liebowitz, A. (1974) "Home Investments in Children." Journal of Political Economy, **82**(2): 111-31.

Neligan, G.A. and Prudham, D., (1969). "Norms for Four Standard Development Milestones by Sex, Social Class and Place in Family." Developmental Medicine and Child Neurology, **11**, 413.

Pascual-Leone, A., Nguyet, D., Cohen, L.G., Brasil-Neto, J.P. Cammarota, A. and Hallett, M., (1965). "Modulation of muscle responses evoked by transcranial magnetic stimulation during the acquisition of new fine motor skills." Journal of Neurophysiology, **74**(3):1037--1045.

Plewis, I. (1996). "Statistical Methods for Understanding Cognitive Growth: A Review, a Synthesis and an Application." British Journal of Mathematical and Statistical Psychology, 49, 1, 25-42.

Ramey, C., Campbell, F., Burchinal, M, Skinner, M., Gardner, D and Ramey, S., (2000). "Persistent Effects of Early Childhood Education on High-Risk Children and Their Mothers." Applied Developmental Science, **4**, 1, 2-14.

Ramey, C. and Ramey, S., (2000). "Early Childhood Experiences and Developmental Competence." Chapter 4 in Waldfogel, J. and Danziger, S. (Eds.) *Securing the Future: Investing in Children from Birth to College*. NY: Russel Sage.

Schmitt, , J. (1993). “The Changing Structure of Male Earnings in Britain, 1974-1988.” Centre for Economic Performance Discussion Paper, No. **122**, March.

Schweinhart, L., Weikart, C. and Larner, M. (1986) “Consequences of Three Pre-school Curriculum Models through Age Fifteen.” Early Education Research Quarterly, 15-45.

Shorrocks, A. (1978) “The Measurement of Mobility.” Econometrica, vol **46**, pp 1013-24

Spearman, C. (1904) “General Intelligence, Objectively Determined and Measured.” American Journal of Psychology, 15, 201-293.

Sternberg, R.J. (1985). *Beyond IQ*. New York: Cambridge University Press.

Stevenson, H. and Newman, R., (1986). “Long-term Prediction of Achievement in Mathematics and Reading.” Child Development, 57, 646-659.

Waldfogel, J., (1999) “Early Childhood Interventions and Outcomes.” CASE paper **21**. Centre for the Analysis of Social Exclusion, London School of Economics.

Wiesel, T. N. and Hubel, D. H. (1965). “Extent of recovery from the effects of visual deprivation in kittens.” Journal of Neurophysiology, **28**:1060--1072.

Zeanah, C.H., Boris, N.W. and Larrieu, J.A., (1997). “Infant Development and Developmental Risk: A Review of the Past Ten Years.” Journal of the American Academy of Child and Adolescent Psychiatry, **36**:2, February.

6. Appendix

Appendix 1: *Tests undertaken by CHES, with Correlation Matrices*

22 Month											
Cube stacking	1.00										
Language use	0.27	1.00									
Personal development I	0.34	0.46	1.00								
Personal development II	0.20	0.29	0.31	1.00							
Drawing	0.31	0.22	0.30	0.25	1.00						
Gross Locomotor	0.19	0.31	0.27	0.32	0.27	1.00					
42 Month											
Counting	1.00										
Speaking	0.40	1.00									
Copying designs I	0.35	0.31	1.00								
Copying designs II	0.28	0.19	0.38	1.00							
Building	0.30	0.26	0.35	0.19	1.00						
Cube stacking	0.16	0.16	0.17	0.08	0.34	1.00					
Picture test I	0.26	0.43	0.23	0.11	0.29	0.26	1.00				
Picture test II	0.35	0.50	0.33	0.19	0.34	0.24	0.57	1.00			
Line drawing	0.25	0.31	0.26	0.11	0.29	0.19	0.27	0.38	1.00		
Gross Locomotor	0.22	0.32	0.24	0.11	0.22	0.15	0.24	0.29	0.19	1.00	
Parts of the body	0.26	0.48	0.27	0.14	0.28	0.26	0.43	0.48	0.29	0.35	1.00
5 Year											
Copying designs	1.00										
Vocabulary	0.30	1.00									
Human Figure Drawing I	0.39	0.22	1.00								
Human Figure Drawing II	0.39	0.22	0.81	1.00							
Profile drawing	0.20	0.19	0.23	0.23	1.00						
10 Year											
Reading	1.00										
Maths	0.49	1.00									
Picture language test	0.53	0.34	1.00								
British Ability Scales	0.74	0.48	0.57	1.00							

Notes: Two Human Figure Drawing tests are reported here for the children at age five. These are both based on the same test but weighted by different procedures developed in the educational literature (Koppitz, 1968 and Harris, 1963). The HFD score used in the text is the average of these two different measures of HFD This avoids the need for assumptions about which weighting procedure is preferable. The correlation between the two scores is, in any case, 0.81, perhaps too high for separate entry in the principal components analysis.

Appendix 2: ***A brief summary of Principal Components Analysis.***

Principal components analysis is the eigenvalue decomposition of the covariance matrix, C, or, where the variables are standardised as here, of the correlation matrix, R. The first principal component is given by:

$$y_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1p}x_p = \sum_{i=1}^{p} a_{1i}x_i$$

where $x_1, \ldots, x_p$ are the variables to be combined, in our case test scores. The weights a_{11}, …, a_{1p} maximise the variance of y_1 and satisfy the normalising constraint:

$$\sum_{i=1}^{p} a_{1i}^2 = 1$$

A second vector of weights, (a_{21}, …, a_{2p}), maximises the variance of the second principal component, y_2, and satisfies

$$\sum_{i=1}^{p} a_{2i}^2 = 1 \quad \text{and } \mathrm{cov}(y_1, y_2) = 0$$

It is possible to calculate as many components as there are test scores, but subsequent components must be uncorrelated with previous components and will account for less of the variation in test scores.

The variance of the i^{th} principal component, λ_i, is known as the latent root or eigenvalue since $l_i = a_i' R a_i$ where a_i is the weight or latent vector for the i^{th} component. Generalising, $\Lambda = A'RA$ where A is the p*p matrix of latent vectors, $A'A = I$ and Λ is a diagonal matrix of the corresponding latent roots ordered from largest to smallest. Pre-multiplying by A and post-multiplying by A' it can be seen that $R = A\Lambda A'$ and we have the definition of principal components analysis as the eigenvalue decomposition of the correlation matrix.

Appendix 3: ***Additional figures showing group rank positions by different classifications.***

Figure 2b: Average rank of test scores at 22, 42, 60 & 120 months, by SES of parents and 42 month rank.

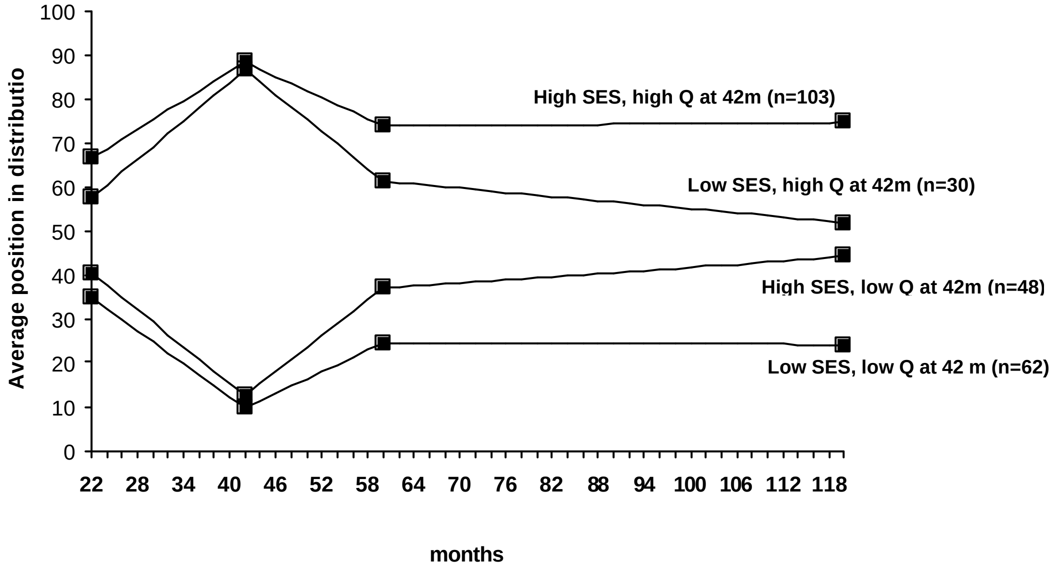
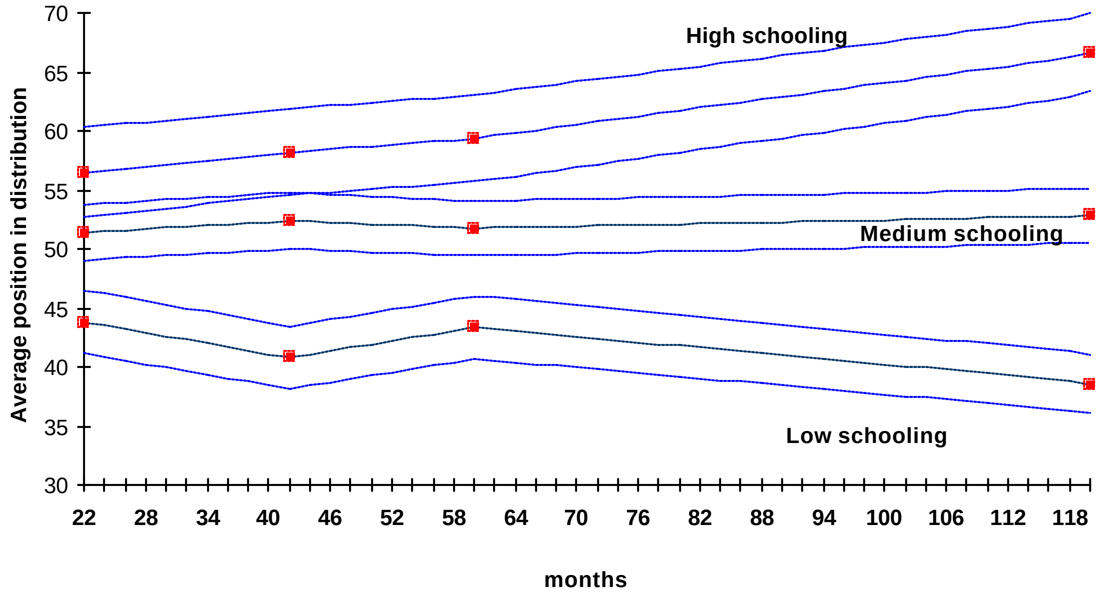
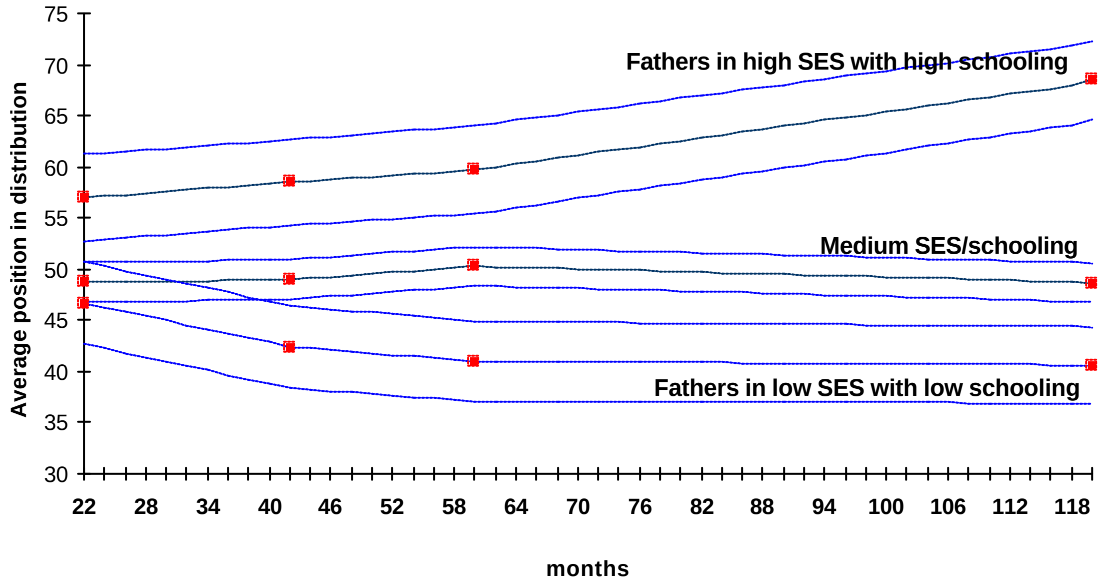


Figure 1b: Average rank of test scores at 22, 42, 60 & 120 months, by schooling of parents



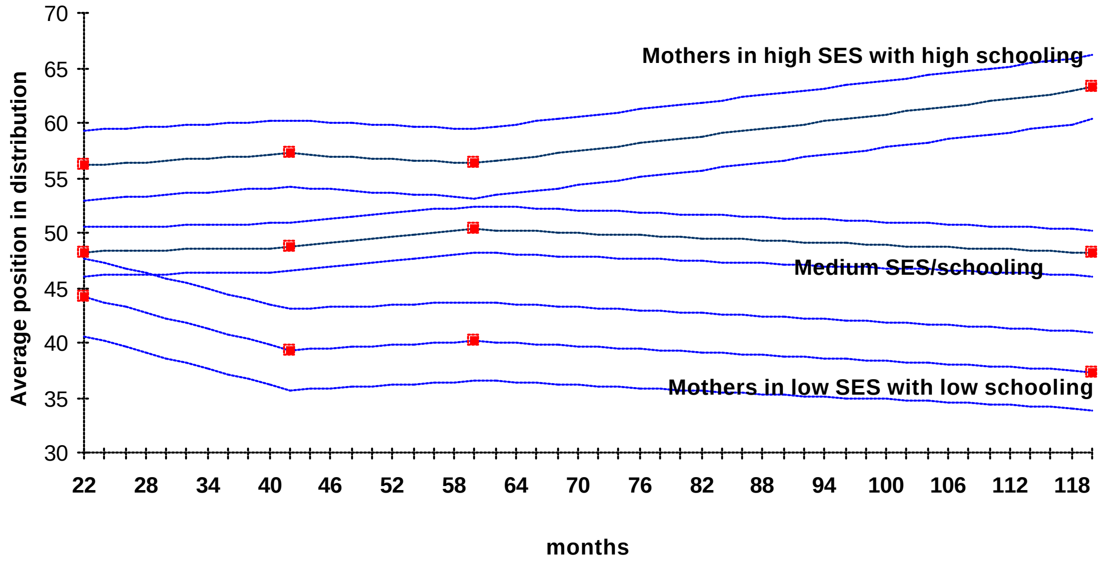
Dotted lines represent intervals of two standard errors. High schooling denotes families where both parents have A'Level or higher (474 obs.) Low schooling denotes families where neither parent has qualifications (226 obs.) Medium schooling denotes those omitted from the high and low categories (592 obs.)

Figure 1c: Average rank of test scores by SES and schooling of fathers.



Dotted lines represent intervals of two standard errors. SES and schooling are defined as in Figures 1 and 2. Fathers classified as medium SES/schooling are those omitted from high and low categories. Observations are 176, 897 and 219 in the high, medium and low categories respectively.

Figure 1d: Average rank of test scores by SES and schooling of mothers.



Dotted lines represent intervals of two standard errors. SES and schooling are defined as in Figures 1 and 2. Mothers classified as medium SES/schooling are those omitted from high and low categories. Observations are 176, 897 and 219 in the high, medium and low categories respectively. Children whose mothers were housewives were placed in the middle group.

Appendix 4: ***Basic statistics for background information in Table 7***

	Obs	Mean	S.d.	Min	Max
Girl	1194	0.46	0.50	0	1
Father in SES 1,2	1194	0.16	0.37	0	1
Father in SES 3m, 3nm, 4	1194	0.78	0.41	0	1
Father in SES 5	1194	0.06	0.23	0	1
Father has no qualifications	1194	0.47	0.50	0	1
Father has low qualifications	1194	0.14	0.35	0	1
Father has medium qualifications	1194	0.23	0.42	0	1
Father has high qualifications	1194	0.13	0.33	0	1
Mother has no qualifications	1194	0.54	0.50	0	1
Mother has low qualifications	1194	0.17	0.37	0	1
Mother has medium qualifications	1194	0.27	0.44	0	1
Mother has high qualifications	1194	0.02	0.14	0	1
Mother in SES 1,2	1194	0.10	0.30	0	1
Mother in SES 3m, 3nm, 4	1194	0.55	0.50	0	1
Mother in SES 5	1194	0.01	0.11	0	1
Mother is housewife	1194	0.34	0.47	0	1
Mother’s age	1194	26.2	5.3	16	44
No older siblings	1194	0.39	0.49	0	1
One older siblings	1194	0.35	0.48	0	1
Two older siblings	1194	0.15	0.36	0	1
More than two older siblings	1194	0.11	0.32	0	1
No younger siblings at 42 months	1194	0.65	0.48	0	1
One younger siblings at 42 months	1194	0.33	0.47	0	1
Two or more younger siblings at 42 months	1194	0.02	0.15	0	1
No younger siblings at 5 years	1194	0.57	0.50	0	1
One younger siblings at 5 years	1194	0.38	0.49	0	1
Two or more younger siblings at 5 years	1194	0.05	0.23	0	1