



## Supporting Online Material for

### **Preschool Influences on Mathematics Achievement**

Edward C. Melhuish,\* Kathy Sylva, Pam Sammons, Iram Siraj-Blatchford, Brenda Taggart, Mai B. Phan, Antero Malin

\*To whom correspondence should be addressed. E-mail: t.melhuish@bbk.ac.uk

Published 29 August 2008, *Science* **320**, 1161 (2008)

DOI: 10.1126/science.1158808

#### **This PDF file includes**

Materials and Methods  
SOM Text  
Fig. S1  
Tables S1 and S2  
References

## **Supporting Online Material**

### **1. Materials and Methods**

#### **Study overview**

The Effective Pre-school and Primary Education (EPPE) project is one of Europe's largest longitudinal studies. Almost 3000 children have been followed from the start of pre-school (around the age of 3), through the start of school (age 5), and then followed so far through age 10. The study focuses on the factors and experiences that are related to children's progress in both intellectual and social development, followed by intensive case studies to 'un-pack' effective practices in pre-school and school. This report concerns analyses for the data from those children with pre-school experience for whom data is available through to age 10. The research is funded by the UK's Department for Children, Schools & Families. Further details are at [www.ioe.ac.uk/projects/eppe](http://www.ioe.ac.uk/projects/eppe).

#### **Participants**

One hundred and forty one pre-school centers were randomly chosen in 6 local authorities, that were selected as having a demographic make-up similar to that of England overall. From these 141 centers 2857 children were recruited into a longitudinal study. Children already in pre-schools were recruited when they became 3 years old; children starting pre-school after their third birthday were recruited at entry to pre-school. Their mean age at entry to the study was 3 years 5 months (SD = 4.6 months). Children included in the sample also needed to satisfy the following criteria:

- Stay at least 10 weeks in the study center subsequent to recruitment to the study (considered to be minimum time for a pre-school center to have an impact).
- attend three or more sessions (session = half-day or 2.5 hours) a week at their pre-school center. Additionally, if the child attended more than one pre-school center (those in dual provision) the study center must be the dominant center (in terms of time per week).

Children without pre-school experience or not attending state primary schools in England were excluded from the analyses in this paper, giving a sample of 2558. Multiple imputation (*I*) was used to compensate for missing data so that all these 2558 children were included in imputed analyses.

#### **Measures**

When children entered the study, they were assessed with four subscales from the British Ability Scales II (BAS II; block building, picture similarities, verbal comprehension and naming vocabulary) (2) to give a general cognitive ability (GCA) score. Upon entering primary school at age 5, children were assessed again with the BAS II. In addition, literacy was assessed by combining the Letter Recognition Test (3) and subscales on the Phonological Awareness assessment (4); numeracy was assessed by the Early Number Concepts subscale of the BAS II. Data were collected on the children's performance in Key Stage 1 national assessments (age 7+ years) in English, mathematics and science. At the end of the sixth school year (10+ years) standardized, teacher conducted, assessments of the children's achievement in mathematics were obtained (5).

Shortly after initial child assessments, one of the child's parents or guardians was interviewed (usually the mother). Most questions in the semi-structured interview were pre-coded, with some open-ended questions coded post hoc. The interview covered: parents' education, occupation and employment, family structure, ethnicity and languages used, the child's birth weight, health, development and behavior, the use of preschool

provision and childcare history, and significant life events. The parental interview included questions concerning the frequency that children engaged in 14 activities: playing with friends at home, playing with friends elsewhere, visiting relatives or friends, shopping with parent, watching TV, eating meals with the family, going to the library, playing with letters/numbers, painting or drawing, being read to, learning activities with the alphabet, numbers/shapes and songs/poems/nursery rhymes, as well as having a regular bedtime. Frequency of activities was coded on a 7-point scale (0=not at all; 7= very frequent). These activities were considered in the construction of a home learning environment (HLE) index. A response rate of 97% for the interview was achieved.

### **Analytic strategy**

Children and families were clustered by pre-school center and primary school and data were hierarchical. Using standard regression with such data can lead to inaccurate error variance estimates. Potentially there was greater similarity between participants within the same centers or schools so the independence of measurement assumption would be violated and misestimating levels of significance was likely. Hence cross-classified multilevel models (6) were used to overcome such problems. The statistical software used was MLwiN (7).

The dependent variable (outcome) in analyses was mathematics achievement at age 10 years. The mathematics achievement scores were adjusted for child's age and converted to z-scores a (mean=0, SD=1) with a range of -2.29 to 2.34. The independent variables (predictors) used in building the multilevel models were:

**Child Characteristics:** age at assessment, gender, ethnicity, early developmental, health, and behavior problems, and birth weight.

**Family Demographics:** maternal age at child's birth, lone parenthood (yes/no), mother's and father's education, mother's maternal work status (employed/not employed), socio-economic status (SES) of family (based on highest occupational status of parents), household language (English only, English and other language(s), other language(s) only), household income, and number of siblings.

**Area characteristics:** The Index of Multiple Deprivation (IMD) (8) was used to provide measures of area where the child lived through matching by postcode. The overall IMD score and its seven constituent domains of deprivation were used as explanatory variables in the model: Income deprivation, Employment deprivation, Health deprivation and disability, Education, skills and training deprivation, Barriers to Housing and Services, Living environment deprivation, and Crime.

**Pre-school characteristics:** months in pre-school, type of pre-school, composition of the pre-school in terms of percentage of mothers with a degree, and average cognitive ability level of children at 3-4 years old.

**Primary school characteristics:** % children with special educational needs (SEN), % children eligible for free school meals (FSM).

When the effects of these variables had been tested and those statistically significant kept in the model additional variables were added to the multilevel model. These measures of home, pre-school and school environments were the home learning environment, pre-school effectiveness and primary school effectiveness. The following sections describe how the measures of home learning environment, pre-school effectiveness and primary school effectiveness were derived.

### **Producing a measure of the Home Learning Environment (HLE)**

Parenting practices such as reading to children, using complex language, responsiveness, and warmth in interactions, are all associated with better developmental outcomes (9). This partly explains links between socio-economic status (SES) and developmental outcomes, in that higher SES parents use more developmentally enhancing activities (10). Stimulating activities may help children with specific skills enhancing development (e.g. linking letters to sounds), but also, and perhaps most importantly, by developing the child's ability and motivation concerned with learning generally. Additionally it is possible that a feedback loop is operating whereby parents are influenced by the child's level of attainment, which would lead to children with higher ability possibly receiving more parental stimulation. What aspects of the home environment produce effects upon children's school readiness in literacy and numeracy? To address this question, an interview-based measure of the home environment was developed and tested for its association with academic achievement at the start of school. Groups with unexpected levels of attainment (not achieving as expected on the basis of demographic characteristics) were examined using multilevel models to investigate sources of unexpected performance that were linked to the home environment. The methods described here are discussed in more detail elsewhere (11).

Children's characteristics and family background were included in a multilevel model (children clustered within pre-school center), to predict children's age-adjusted achievement at age 5 (start of school) using as predictors the child, family, area and pre-school characteristics (family and background characteristics). From these models, three categories of performance (unexpected over-achievers, expected, and unexpected under-achievers) for literacy and numeracy were constructed based on child residual scores deviating by at least  $\pm 1$  standard deviation. Each category of unexpected over- or under-achievement is a nominal outcome variable with *average achieving* children as the reference category. Sixteen percent of children were achieving higher than predicted from their background in both literacy and numeracy, and a similar proportion (16% literacy, 15% numeracy) were achieving less well than would be predicted.

Each of the 14 home activity items that had been derived from the parental interview was regressed in separate equations on the individual categorical variables of over- or under-achievement. The seven social/routine activities (play with friends at home, and elsewhere, visiting relatives/friends, shopping, TV, eating meals with family, regular bedtime) were not significant for under- or over-achievement in literacy and numeracy at age 5. Conversely, the seven activities providing clear learning opportunities (frequency read to, going to the library, playing with numbers, painting and drawing, being taught letters, being taught numbers, songs/poems/rhymes) each had significant positive effects on unexpected achievements. Since the items are conceptually and statistically linked a combined measure, the home learning environment (HLE) was created. The frequency of each of the seven activities was coded on a 0-7 scale (0=not occurring, 7= very frequent), and the seven scores were added to produce an index with a possible range of 0-49, which was normally distributed (mean = 23.42, SD = 7.71).

The multilevel models for age 5 outcomes treated children as clustered by preschool center, allowing the estimation and separation of residuals into individual and center variance, and estimation of the amount of variance explained by adding parameters

to the model in stepwise fashion. For age 5 literacy and numeracy, family and background characteristics explained significant individual variation between children in centers: 16% for literacy and numeracy scores. Thus, most variation in children's achievement was not due to family or background characteristics but to other unmeasured factors not considered in the demographic model.

It was hypothesized that variations in predicted achievement based upon family and background characteristics (i.e. unexplained individual-level variance) would be partially accounted for by the home learning environment. Firstly, the categories of over- and under-achievement for children and centers were examined for a relationship with home learning environment. The mean HLE scores for the over-achieving (mean= 26.44, SD=7.26), average (mean= 23.61, SD=7.45) and under-achieving (mean= 21.62, SD=7.83) groups of children appear to vary systematically for the demographically adjusted levels of achievement (i.e. unexpected over-achieving, expected, and unexpected under-achieving) in literacy. Multinomial logistic regressions confirm, as hypothesized, that children with a higher HLE are more likely to be over-achievers ( $p < 0.0001$ ), while lower HLE scores are associated with under-achievement ( $p < 0.0001$ ). For numeracy the effects were also significant but not as strong as for literacy. Children with higher HLEs had a greater likelihood of over-achieving in numeracy, and those with lower HLE had a greater likelihood of under-achieving in numeracy.

To support the conclusions that the HLE added to the prediction of achievement over that provided by family and background characteristics for children, new multilevel models for literacy and numeracy were created including the HLE index in addition to the significant family and child background factors. By adding the HLE to the demographic model, the explained variance at the child level showed a 21% increase for age 5 literacy and an 18% increase for age 5 numeracy, thus supporting the conclusion that the home learning environment is an important predictor of development.

**A measure of the Home Learning Environment (HLE):** The results support the importance of the HLE and the influence of the HLE was over and above that of standard proxy measures of parental education and SES, thus demonstrating the usefulness of this measure of variability in parenting. While other family factors such as parents' education and SES are also important, the extent of home learning activities exerts a greater and independent influence on educational attainment. The comparison of over, average, and under-achieving groups indicates that at age 5 the HLE is effective in differentiating both over and under-achieving groups from children achieving as expected, i.e. across the ability range.

The home learning environment is important for school readiness yet it only moderately associated with SES and parents' educational levels (correlations= 0.28 - 0.32), indicating that low SES homes sometimes score highly and, conversely, high SES homes at times sometimes score poorly on the HLE measure. Thus the HLE measure is substantially independent of social class related measures.

The effects of the home environment and parenting upon children's development may partly be due to the teaching and learning of specific skills, e.g. letter-sound relationships. However, the multiplicity of learning opportunities included in the HLE suggests that the effects may be related to more generalised and motivational aspects of child development, e.g. learning to learn. Also children may internalise aspects of

parental values and expectations (implicit in the activities of the HLE) as they form a self-concept of themselves as a learner. Such a perspective is congruent with Vygotsky's theory (12) that children learn higher psychological processes through their social environment and specifically with adult guidance operating within a child's "zone of proximal development" (stimulation within the child's comprehension) and reinforces the idea that children acquire cognitive skills such as literacy through interaction with others who aid and encourage skill development.

It is quite possible that the strong relationship between home learning environment and cognitive scores is mediated by some intervening unmeasured factor. Those parents, who answer the questions in a way leading to a high score, may have other characteristics that lead their children to have higher cognitive scores. Even if this were so, the HLE would still be an efficient proxy measure of such unmeasured factors.

Whatever the mechanisms, the influences of parenting upon child development are pervasive. Research involving 0-3 year-olds from the evaluation of the Early Head Start (EHS) program found that the intervention increased both the quantity and quality of parents' interaction with children, as well as children's social and cognitive development (13). A review of early interventions concluded that, to gain the most impact, interventions should include both parent and child together with a focus on enhancing interactions (14). Such work indicates that parenting behaviors are learnable, and changes in parenting are associated with improved child development. Similar conclusions derive from a UK study (15) where children showed better literacy progress when parents received a program on improving child literacy during the preschool period.

### **Producing a measure of the effectiveness of pre-school centers**

In order to investigate the impact of individual pre-school centers on young children's progress, it is essential to have accurate baseline data about children's prior attainment so that subsequent gains can be measured. Only in this way will it be possible to establish whether children attending specific centers showed greater progress by entry to school. Additionally, it is also necessary to control for variations in the background characteristics of children attending different centers, as well as their prior attainments, in order to take adequate account of the influence of intake before comparisons of center effectiveness are made (16).

### **Participants for studying pre-school effectiveness**

The pre-school stage of the study involved 2,857 children recruited from 141 pre-school centers, including 588 children from nursery classes, 606 children from playgroups, 517 children from private day nurseries, 433 children from Local Authority centers, 517 children from nursery schools and 192 children from integrated children's centers, and thus included all types of pre-school center in England at the time of the study. The number of children within a pre-school center varied from 6 to 35, with a mean of 20.26 and a standard deviation of 5.66. With substantial numbers per pre-school center, estimation of pre-school center effects from multilevel models are likely to be reasonably reliable. The children were aged between 3 years and 4 years 3 months (mean = 40.4 months; SD = 4.6 months) at the beginning of the study.

The child assessments at entry to the study and upon starting primary school and the measures of background information derived from parental interview as described earlier under measures.

### **Analysis of progress over the pre-school period**

Multilevel models are a generalized form of regression analysis, particularly suited to data exhibiting a hierarchical structure (5), and they provide a method of measuring the variation in children's attainment that can be attributed to differences between individual children and to differences between groups (e.g. resident in different areas, or attending different schools or pre-schools). Research in the school effectiveness field (17, 18) has shown that prior attainment is a stronger predictor of future attainment than other pupil background indicators. The assessments at entry to the study (age 3 years plus) on the BAS Ability Scales provide baseline measures of cognitive ability for the value added analysis. The age of the child must be taken into account at both assessment points i.e. entry to the study and also at entry to primary school. The cognitive ability assessments at entry to pre-school have been age standardised to take account of the different ages at which children are assessed at pre-school entry (reflecting variations in center entry policies and parents uptake of places). For assessments made at entry to primary school, the age in months of the child when the test was completed is controlled for in all the multilevel models. Child, parent, family, home environment, time in pre-school and composition of pre-school intakes are all likely to affect children's attainment. Consequently, the multilevel model investigates the impact of such measures on children's progress over the pre-school period while taking account of prior attainment. The predictor variables that were tested in multilevel models included:

#### **Child characteristics**

Prior cognitive ability

Age

Gender

Birth weight

Perinatal health difficulties

Previous developmental problems

Previous behavior problems

Previous health problems

#### **Parental characteristics**

Socio-economic status

Mother's level of employment

Father's level of employment

Mother's qualifications

Father's qualifications

Mother's age

Father's age

Age mother left education

Age father left education

Marital status

#### **Family characteristics**

Lone parent

Number of siblings

Birth position

Life events

**Home characteristics**

Home learning environment

Rules about bedtime

Rules about TV

Play with friends at home

Play with friends elsewhere

**Childcare history**

Total relative care before entering the study

Total individual care before entering the study

Total group care before entering the study

**Pre-school experience variables**

Number of months in pre-school

Pre-school composition - percentage mothers with degree or higher qualification

The results reveal that child, family and home environment factors continue to show a statistically significant relationship with children's progress over the pre-school period and account for additional variance at both the center and child level. In summary, when exploring the impact of pre-school centers on children's progress, in addition to baseline measures of children's prior attainment, this study demonstrates the need to obtain good data about the child, parent and home environment. This allows proper control for differences between centers in the characteristics of the children they serve. Only when differences in intake are measured can valid comparisons be drawn. Individual level characteristics show statistically significant relationships with progress over the pre-school period as summarized below.

**Significant factors for child's progress in numeracy over the pre-school period****Child measures:**

- Girls made more progress than boys.
- Children older at school entry assessment made more progress.
- Children with low birth weight made less progress than other children.
- Children with reported developmental problems in first 3 years made less progress.

**Parent/Family measures:**

- Children whose mothers have academic qualifications, especially at age 16 and 18, made more progress than children whose mothers have no qualifications.
- Children from families where the parents' highest SES is classified as either skilled non-manual or semi-skilled manual made less progress than children from professional class families.

**Home environment measures:**

- Children with a higher home learning environment score show more progress.

**Duration of pre-school:**

- Children who spent more months in pre-school made more progress.

**Composition of pre-school intake:**

- Children attending pre-school centers where there were a higher proportion of children whose mothers have degrees, higher degrees or other professional qualifications made more progress.

**A measure of pre-school center effectiveness:** Using multilevel models, the progress of children in numeracy over the pre-school period has been measured to estimate the impact of individual centers. Child characteristics, social background, home learning environment, duration of pre-school attendance and composition of pre-school were all allowed for in analyses. The analyses indicate that pre-school centers differ in their impact on children's progress in numeracy, i.e. in some centers children make significantly greater progress than in others after allowing for background and prior attainment. The center level residuals from the multilevel models provide a measure of the extent that an individual preschool affects a child's progress, and can be regarded as a measure of pre-school center effectiveness. These residuals measure the difference between the expected and actual scores, after controlling for differences in prior attainment, child, parent and home environment characteristics as well as duration and composition of pre-school. Pre-schools where children made significantly greater progress than predicted on the basis of prior attainment and background characteristics could be viewed as more effective. Pre-schools where children made less progress than predicted could be viewed as less effective. This measure of pre-school center effectiveness was used in the analyses of the effects of home, pre-school and primary school influences upon children's attainment in mathematics at age 10 years. This measure was converted to a z-score (mean=0, SD =1.0) ranging from -3.3 to 2.35.

### **Primary School Effectiveness for all state primary schools in England**

Here we consider the factors affecting pupils' progress in mathematics in primary schools in order to produce measures of effectiveness in promoting mathematics attainment for all primary schools in England. Thus the measures of school effectiveness are produced independently of the longitudinal sample used in the main analyses. The effectiveness measures for the schools in the longitudinal sample allowed the research team to place the schools that children attended in the context of all other schools in England. The effectiveness measures were derived using data from national assessments in all state primary schools in England from three years (2002-2004). Factors known to influence these results, prior achievement at age 7 (Key Stage 1) and pupil characteristics, were included in the analysis so that the measures of effectiveness reflect the schools' effectiveness rather than the composition of the school. A school effectiveness research design was used to investigate children's progress from age 7 to age 11. Multilevel models (three levels, children within schools, schools within local authorities) were used to explore primary school effects and the 'value added' by different primary schools (19).

The multilevel models produced estimates of the differences between pupils, between schools, and between local authorities (LAs). The focus of this analysis was in terms of progress in mathematics from age 7 to 11 (Key Stage 2, or KS2, in England). Looking at the results for the primary schools over three separate years enabled the stability of school effectiveness measures to be established.

### **Analysis strategy for school effectiveness**

Multilevel models were used to investigate children's progress from age 7 to 11 by controlling for a child's prior attainment, as well as a number of background influences. These analyses were used to establish the extent to which children's progress was associated with the primary school attended. The calculation of residuals (based on

differences between children's expected and actual attainments at age 11; end of KS2) for each school provided a value added indicator of each school's effectiveness in promoting progress in a given outcome. Primary schools where children made significantly greater progress than predicted on the basis of prior attainment and intake characteristics could be viewed as more effective (positive outliers in value added terms). Primary schools where children made less progress than predicted could be viewed as less effective (negative outliers in value added terms).

In the analyses of pupil's progress from 7 to 11, age 11 national assessment scores in mathematics were used as outcomes. Prior attainment in reading, writing, mathematics and science was measured using age 7 national assessment scores. Pupils' results from the two time points (age 7 and age 11) were matched, using name, gender and date of birth. Using Pupil Level Annual Schools' Census (PLASC) data, the models were developed through controlling for pupil background characteristics such as gender, ethnic group, English as first or additional language, free school meals eligibility (FSM) and Special Educational Needs (SEN) provision type. Yet further development of the models involved adding further control variables reflecting the characteristics of the area in which the pupil resided. This level of control of covariates is the most extensive that has been attained with these national data sets. The Pupil Level Annual Schools' Census (PLASC) provided the following pupil information:

- Gender
- Date of birth
- Date of admission to school
- Home postcode
- Ethnic group
- English as first language or not
- Free school meal eligibility (FSM)
- SEN provision type
- Details of permanently excluded pupils.

The child's postcode was used to relate the child's area of residence to data from the 2004 Index of Multiple Deprivation (IMD) and variables derived from the 2001 census to provide area variables as follows:

Overall Index of Multiple Deprivation (IMD)

% of adults (16-74 years) with no qualifications

% of households headed by a lone parent

% of households that are owner-occupier

% of adults (16-74 years) in managerial jobs

% of adults (16-74 years) in intermediate jobs

% of adults (16-74 years) in lower supervisory & technical jobs

% of adults (16-74 years) in routine jobs

% of adults (16-74 years) never worked & long-term unemployed.

Also the PLASC data were used for two variables of school composition. These were

% of children in the school eligible for free school meals (FSM)

% of children in the school that had special educational needs (SEN).

The data described above were used in analyses of pupils' progress from age 7 years to age 11 years, and hence provided estimates of school effectiveness adjusted for the above pupil background and area characteristics.

For each of the years 2002, 2003 and 2004, age 11 mathematics scores were analyzed using multilevel models as a function of the variables described above. These models took into account the hierarchical structure of the data, with the pupils clustered within schools, and the schools clustered within LAs. With multilevel models it was possible to estimate measures of individual school effectiveness, after taking account of the lack of independence between pupils and schools, whilst allowing for the effects of covariates to be assessed and controlled for. As the variation in each model was partitioned between LAs, schools and pupils, the estimates of school effectiveness were calculated by summing the LA and school effects. These effects reflected the effectiveness of schools after having taken into account the pupils' previous level of achievement and the pupils' characteristics and the structure of the data.

### **Sample for primary school effectiveness analysis**

It was necessary to have age 7 results, age 11 results and PLASC data matched by pupil. Complete data were available for 149-155 LAs and 14765–15846 schools over the three years. The number of schools in the LAs ranged from 1 to 522 (median=70, inter-quartile range (IQR)=70). The number of pupils in a school ranged from 1 to 240 (median=33, IQR=36). There were 538,951 to 560,168 pupils in analyses for any one year. The distributions of age 11 mathematics scores for each of the years were similar (2002: mean = 63.3, SD = 20.5; 2003 mean =22.0, SD = 22.0; 2004 mean =64.7, SD = 21.8) and the range was always 0-100.

### **Findings related to school effectiveness**

Effect sizes are given for predictor variables. Computation for effect size is as follows. For the binary explanatory variables:

$$\text{Effect size} = \beta / \sigma_{\text{pupil}}$$

Where  $\beta$  is the model parameter estimate and  $\sigma_{\text{pupil}}$  is the standard deviation at pupil level.

This is equivalent to the difference between category means, in standard deviation units.

For continuous variables, the following method (19, 20) was used:

$$\text{Effect size} = (2\beta * \text{SD}) / \sigma_{\text{pupil}}$$

Where SD = standard deviation of the independent (explanatory) variable, and  $\sigma_{\text{pupil}}$  is the standard deviation at the pupil level.

This is equivalent to the difference in values of the dependent variable corresponding to the points one standard deviation above and below the mean of the independent variable, measured in standard deviation units.

**Prior attainment:** The effects of prior ability, as measured by age 7 scores in reading, writing, mathematics and science, were allowed to vary by school in the models, thus allowing each school to have its own measure of progress. The interpretation of these random effects is discussed at the end of the results section. In this section, the average effects of prior ability on age 11 mathematics are discussed after allowing for schools to have different rates of progress and after controlling for pupil, area and school level characteristics. For all years prior ability in mathematics has the largest effect on age 11 mathematics scores (effect sizes=1.63 – 1.81). There are also significant effects of prior ability in reading, writing and science.

**Age, Language, FSM and SEN:** Although pupils’ age at the start of the academic year was significantly related to the age 11 mathematics scores across all years, with younger pupils progressing more than older pupils, the effect sizes were fairly small. Pupils who had English as an additional language showed more progress in mathematics in all three years, however the effect sizes were small (0.15 - 0.18). The progress made by children entitled to free school meals (FSM – an index of poverty) was allowed to differ between schools. So the effects reported here are the effects associated with a child attending an average school. A consistent relationship was found between free school meals status and progress in mathematics, with pupils entitled to free school meals showing significantly less progress but effects were small. SEN status had larger effects (-0.34 to -0.64).

**Area of Child Residence:** In looking at the possible influence of the characteristics of where a child lives, a wide range of area variables was initially considered. The IMD did not produce any significant effects as census-derived variables were having a stronger effect and therefore suppressed any IMD effect. All of the other variables (derived from the 2001 census) produced small size significant effects. All of these variables in some way reflect aspects of a continuum of area deprivation. The results are consistent with an explanation that the more deprived the area of residence the less progress will be made. All of these area variables are correlated with each other and the sum of their small individual effects will be greater than that indicated for each individual variable.

**School composition:** School composition variables are the percentage of children eligible for free school meals (FSM), percentage of children with special educational needs (SEN). Both show small significant effects. The greater the percentage of children eligible for FSM, or having SEN within a school the poorer children’s progress will be.

#### **Interpretation of random effects in the final models**

Consider a simplified model containing just an intercept and the age 7 mathematics score –mathematics - with the intercept random across LA and school and the slope of the mathematics variable also random across schools. Algebraically this would be written as

$$Y_{ijk} = \beta_0 + \beta_1 \text{mathematics}_{ijk} + v_{0k} + u_{0jk} + u_{1jk} \text{mathematics}_{ijk} + e_{0ijk}$$

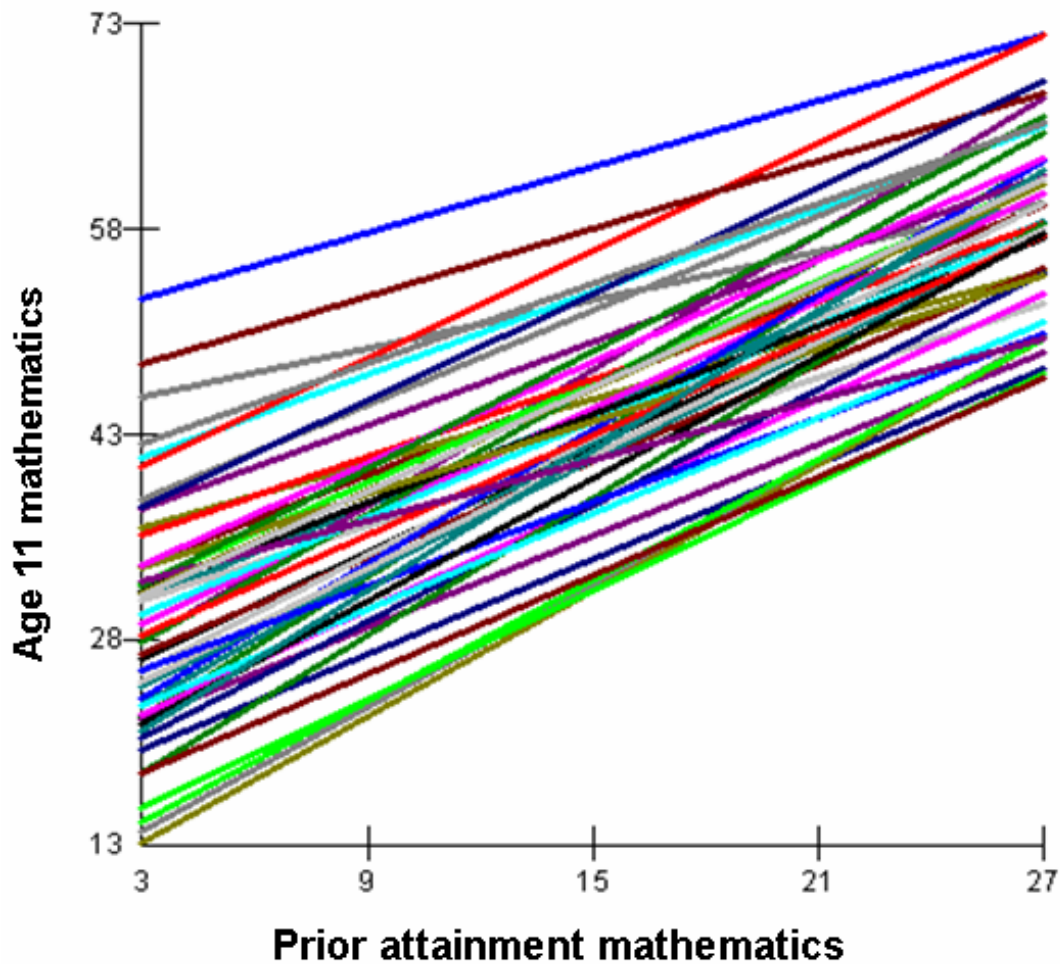
where  $Y_{ijk}$  is the age 11 mathematics score for pupil  $i$ , in school  $j$ , in LA  $k$ .

At the school level there are two residuals,  $u_{0jk}$  is the overall school effect – the residual from the overall intercept,  $u_{1jk}$  is the deviation from the slope for school  $j$  in LA  $k$ ; the change in outcome for a unit increase in the mathematics score for a pupil in this school is given by  $\beta_1 + u_{1jk}$ .

The residuals  $u_{0jk}$  and  $u_{1jk}$  have corresponding variances  $\sigma_{u0}^2$  and  $\sigma_{u1}^2$  respectively. These indicate the distribution of the intercepts and slopes for each school

around the averages. But the two residuals for each school are also related through a covariance  $\sigma_{u01}$ . The covariance indicates the tendency for the slope to be steeper when the intercept is higher a positive covariance – diversion) or for the slope to be shallower when the intercept is higher a negative covariance – conversion). The third possibility is for there to be no covariance – the relationship between intercepts and slopes is random.

There was a significant effect of school level variance associated with age 7 mathematics indicating that schools vary in their effects depending upon how pupils vary on age 7 ability, which is illustrated in Figure S1.



**Figure S1.** Examples of relationships for prior attainment and age 11 outcomes for schools. This is illustrative (for 50 schools) of the nature of differential effectiveness within schools for children of different levels of initial (age 7) ability. Each line represents the relationship within a school of age 7 score to age 11 score (eliminating the effects for all variables in the model) for the pupils of that school.

The covariance between the residuals for the intercept and prior attainment (age 7 mathematics) is -4.18. This equates to a correlation of  $-4.18/\sqrt{(173.32*0.21)} = -0.69$ . So schools with higher intercepts tend to have shallower slopes, whilst those with lower intercepts tend to have higher slopes. Figure S1 above illustrates how this relationship might look for 50 schools generated from these parameters. Each line represents the mean performance for a school for a given prior attainment level. The negative covariance results in there being less variation between schools when pupils' prior attainment is high than when it is low. The fact that the lines tend to cross each other suggests that the relative position of schools – which achieve “better” results – will vary according to the pupil's prior attainment (in this case age 7 mathematics). The schools that do better for pupils with lower prior attainment (schools with high intercepts) do not necessarily perform as well for pupils with higher prior attainment. The shallower slope means that the effect of prior attainment on the outcome tends to be smaller at schools with high intercepts schools that do better for pupils with low prior attainment). Looking at models across years similar patterns emerge.

What does this mean for school effectiveness? The shallow slopes of the highly effective schools could be brought about by either a relative boosting of age 11 scores of low ability (low age 7 score) pupils or a relative suppression of the age 11 scores of high ability (high age 7 scores) pupils. The correlation between a schools average age 11 attainment and the school effectiveness scores derived from the models are 0.47 – 0.53 for mathematics. Hence it is clear that more effective schools achieved higher age 11 results. This is consistent with the view that if such schools were suppressing pupils' scores then they would not be highly effective. Hence, the conclusion is that they are highly effective through their relatively greater boosting of the age 11 achievement of lower ability pupils. Conversely, the steeper slopes of the less effective schools could be brought about by the relative suppression of the age 11 scores of low ability pupils or the relative boosting of the age 11 scores of high ability children. If such schools were boosting (relative to other schools) pupils' achievement they would not be less effective. Hence it appears that they are producing a relative suppression of the ability of the low ability pupils. Overall, attending a highly effective versus an ineffective school matters much more to low ability pupils than high ability pupils, and that the differential effect for low ability pupils is a major factor in differences in school effectiveness.

The degree of stability of school effectiveness across years was measured with correlations between school residuals for each subject across years. These correlations were 0.55-0.56 for adjacent years and 0.42 between 2002 and 2004, which indicates moderate levels of instability in primary schools' levels of effectiveness. Hence averaging effectiveness scores across years is likely to provide measures more indicative of longer-term school performance.

### **Summary of school effectiveness results**

In analysing progress, predictor variables are used that might explain progress in addition to measures of children's ability at age 7. After allowing for effects associated with the individual level predictor variables and school composition there was significant school level variance indicating that progress in mathematics differed significantly from school to school. The school level residuals were a measure of such differences in school effectiveness.

Also the analysis of random effects reveals that there are marked differences in the amount of progress that schools produce dependent upon the level of initial ability of pupils. The difference in the consequences for progress of being in a school that is dependent upon the initial pupil ability can be termed differential effectiveness. The level of differential effectiveness is markedly different for different primary schools. This phenomenon has been described and discussed in the literature on secondary schools (22, 23) but not in the literature on primary schools. The analysis reveals that overall measures of school effectiveness are associated with the differential effectiveness within a school. This comes about because those schools with higher overall effectiveness produce a relatively greater boost to low ability pupils than to high ability pupils and that this differential boost is greater for the overall more effective schools. This is not a tautology. Rather it indicates that effective schools are particularly successful with their lower ability pupils than are less effective schools. While differences are also present between relatively effective and ineffective schools for high ability pupils they are less marked than for the low ability pupils. The consequences of differential effectiveness are that while all children benefit from being in an effective primary school rather than an ineffective one, the consequences are considerably greater for low ability children than for high ability children.

#### **Selecting measures for primary school effectiveness**

The analyses described here clearly demonstrate that schools can show different levels of effectiveness in mathematics teaching for pupils of different initial ability. In order to take account of this phenomenon, the multilevel models were run for the top third, middle third and bottom third (based on initial mathematics ability) of pupils separately, for the three years. This provided measures of primary school effectiveness for high medium and low ability pupils for all schools, for each of the three years. As there was moderate instability in schools' effectiveness across years, averaging effectiveness scores across years provided measures more indicative of longer-term performance. Thus the measures for high ability pupils were averaged across the three years, and similarly for measures for medium and low ability pupils. These primary school effectiveness measures in relation to child's ability were used in subsequent analyses. The effectiveness measures were converted to z-scores ranging from -3.12 to 5.7 (mean = 0, SD = 1). For the children in the longitudinal study the primary school effectiveness scores were extracted for their particular primary schools and matched to individual children by ability.

Having described the derivation of the materials and methods including the derivation of the measures of the home learning environment, pre-school effectiveness and primary school effectiveness we now return to the results of the main analyses for the longitudinal sample.

## **2. Supporting Text**

### **Results of main analyses for longitudinal sample**

There were 2558 children who had attended pre-school and a state primary school in England. The analyses concerned the prediction of these children's mathematics scores at age 10 years. The sample's descriptive statistics for the significant predictor variables in the final model are shown in Table S1.

Multiple imputation was used to counter missing data bias by imputing missing data for all children in the study who attended a state primary school. Separate analyses using either complete cases data or multiple imputation data revealed that differences in results for complete and imputed data were negligible; therefore imputed data results are reported. In the multilevel models, child, family, area and pre-school and school composition effects were controlled for. The home learning environment, pre-school effectiveness and primary school effectiveness all showed significant effects on children's mathematics achievement at age 10. Total variance accounted for was 0.22. Effect sizes are calculated from the final model, indicating effects having allowed for all other variables. Table S2 presents the effect sizes from the final model for the total sample and separately for the low income and high income groups.

### **Interpretation of results**

Can we interpret effects as causal? In a randomized controlled trial, with a sufficiently large sample size the random assignment to treatment or control groups is assumed to result in groups being equivalent on background characteristics. Therefore any differences found are assumed to be causally related to the treatment. Randomised controlled trials have been used to investigate pre-school effects for small very disadvantaged samples (24) receiving experimental interventions, but such a methodology would not be feasible for investigating effects that apply to the general population receiving generally available provision. Observational studies, such as the study here, do not have random assignment so it is always possible that results may reflect selection bias and/or the operation of unmeasured variables. Hence making causal attribution on the basis of observational evidence is always subject to uncertainty. One way to reduce such uncertainty is to statistically control for covariates that may also be related to the outcome. In this study the analyses have tested a large number of covariates, and effects only reported when adjustments have been made for covariates. Nonetheless it is always possible that there are unmeasured variables that are affecting the results. However for any unmeasured variable to have an influence upon results in a study such as that reported here it has to have an effect upon the outcome and NOT have a substantial relationship with any of the measured covariates. The fact that a large number of covariates were tested in analyses reduces the probability that such an unmeasured variable has influenced the results. Ultimately the test of the causal explanation resides in the plausibility of the alternative explanations and their implications, rather than the analyses themselves.

Another way of expressing the limitation discussed above is in terms of selection-treatment interactions or selection bias. In this case are the variables home learning environment, pre-school effectiveness and primary school effectiveness exogenous or relatively independent of other influential predictors? Considering the home learning environment, the highest levels of association with any covariate in the study are with measures of social class such as parental occupational status, educational status or

income, and these show only a modest association ( $r=0.28-0.32$ ). Such low associations mean that the analysis methods can easily adjust for the association with other variables in establishing associated effects upon outcomes. For pre-school effectiveness and primary school effectiveness the associations of either of these variables with any of the covariates in the final model are all low ( $<0.13$ ) and indicate that they are relatively independent or exogeneous variables for the purposes of this study. Of course there remain the possibility of other unmeasured variables that may covary with the variables showing effects, and while these can never be ruled out, until they are found to exist and identified, the variables used here function as good proxy measures for any such unmeasured factors.

### 3. Supporting Tables

**Table S1.** Descriptive statistics for predictor variables

<b>Variable</b>	<b>Mean</b>	<b>S.D.</b>
<i>Birth weight</i>	3310	621.7
<i>Pre-school average cognitive ability</i>	-0.11	6.5
<i>Primary school % SEN</i>	2.69	6.0
	<b>N</b>	<b>%</b>
<i>Gender</i>		
Female	1029	48.0%
<i>Ethnic Group</i>		
White UK	1658	77.4%
White European	63	2.9%
Black Caribbean	92	4.3%
Black African	46	2.1%
Indian	41	1.9%
Pakistani	53	2.5%
Bangladeshi	17	0.8%
Other Ethnicity	52	2.4%
Mixed Ethnicity	121	5.6%
<i>Highest Parental Occupational Status (SES)</i>		
Professional	165	7.7%
Other professional	564	26.3%
Skilled non-manual	773	36.1%
Skilled Manual	276	12.9%
Semi-skilled	274	12.8%
Unskilled	44	2.1%
Never worked	47	2.2%
<i>Mother's Education</i>		
No Qualifications	367	17.1%
Vocational only-mother	358	16.7%
16 academic-mother	872	40.7%
18 academic-mother	182	8.5%
Degree or professional qual.	364	17.0%
<i>Father's Education</i>		
No Qualifications	319	14.9%
Vocational only-father	264	12.3%
16 academic-father	517	24.1%
18 academic-father	167	7.8%
Degree or professional qual.	354	16.5%
father absent	522	24.4%
<i>Family Income</i>		
No earned income	617	28.8

2500-15000 salary	427	19.9%
17500-27500 salary	375	17.5%
30000-35000 salary	230	10.7%
37500-132000 salary	494	23.1%
<i>Home Learning Environment (HLE)</i>		
High	386	18.0%
Medium	1435	67.0%
Low	322	15.0%
<i>Pre-school effectiveness</i>		
High	286	13.3%
Medium	1525	71.2%
Low	332	15.5%
<i>Primary school effectiveness</i>		
High	262	12.2%
Medium	1557	73.0%
Low	324	15.1%

**Table S2.** Effect sizes for significant variables

<b>Predictor Variable</b>	<b>Full Sample (n=2558)</b>	<b>Low Income (n=1344)</b>	<b>Higher Income (n=1214)</b>
Birth weight	0.12*	0.18**	0.05
Female (compared to male)	-0.09*	-0.13*	-0.06
<b>Socio-economic Status (professional/managerial as reference):</b>			
Other professional	0.01	0.10	-0.03
Skilled non-manual	-0.09	0.03	-0.15
Skilled manual	-0.29*	-0.14	-0.38*
Semi-skilled	-0.29*	-0.17	-0.44*
Unskilled	-0.36	-0.31	-0.13
Never worked	-0.19	-0.19	0.77
<b>Ethnicity (white UK as reference):</b>			
White European	-0.26*	-0.29*	-0.20
Black Caribbean	0.00	0.15	-0.28
Black African	-0.15	-0.19	-0.11
Indian	0.35*	0.56*	0.13
Pakistani	0.11	0.15	0.01
Bangladeshi	-0.06	-0.09	0.14
Other Ethnicity	0.07	0.14	0.00
Mixed Ethnicity	-0.06	-0.04	-0.02
<b>Mother's Education (no qualification as reference):</b>			
Vocational only	-0.05	-0.11	0.18
16 academic	0.13	0.05	0.37**
18 academic	0.32**	0.25	0.56***
Degree or professional qualification	0.50***	0.50**	0.68***
<b>Father's Education (no qualification as reference):</b>			
Vocational only	0.13	0.01	0.26
16 academic	0.05	-0.04	0.16
18 academic	0.00	0.02	0.03
Degree or professional qualification	0.23*	0.14	0.33*
Father absent	0.03	0.00	0.09
<b>Total Family Annual Salary (no income as reference):</b>			
2500-15000 salary	0.06	0.06	
17500-27500 salary	0.21**		<b>(reference)</b>
30000-35000 salary	0.21*		-0.01
37500-132000 salary	0.31***		0.08
<b>Composition Effects:</b>			
Pre-School average cognitive ability	0.19**	0.13	0.22**
Primary School % SEN	-0.20***	-0.21***	-0.19*
<b>HLE and School Effectiveness (low as reference):</b>			
HLE high	0.40***	0.45***	0.40**
HLE medium	0.21**	0.21*	0.26*
Pre-school effectiveness high	0.26**	0.25*	0.29*
Pre-school effectiveness medium	0.05	0.06	0.05
Primary school effectiveness high	0.33***	0.33**	0.27
Primary school effectiveness medium	0.39***	0.36***	0.32**

\*p&lt;0.05; \*\*p&lt;0.01; \*\*\*p&lt;0.001

## References

1. J. L. Schafer, *Analysis of Incomplete Multivariate Data* (Chapman & Hall, London, 1997).
2. C. Elliot, P. Smith, K. McCulloch, *British Ability Scales* (NFER-Nelson, Windsor, 1996).
3. M. Clay, *An observation survey of early literacy achievement* (Heinemann, Portsmouth, NH, 1993).
4. P. Bryant, L. Bradley, *Children's reading problem* (Blackwell, Oxford, 1985).
5. *Primary Mathematics 10 Test*. (NFER-Nelson, Windsor, 1994).
6. H. Goldstein, *Multilevel Statistical Models*. (Arnold, London, ed.3, 2003).
7. J. Rasbash, F. Steele, W. Browne, B. Prosser, *A user's guide to MLwiN Version 2.0* (Inst of Ed., London, 2004).
8. *The English Indices of Deprivation 2004 (revised)* (ODPM, London, 2004).
9. R. Bradley, in *Handbook of Parenting*, M. Bornstein, Ed. (Erlbaum, Hillsdale, N.J., ed.2, 2002).
10. R. Hess, S. Holloway, G. Price, W. Dickson, in *Families as Learning Environments of Children*, L. M. Laosa, I. E. Sigel, Eds., (Plenum, New York, 1982).
11. E. C. Melhuish, et al., *J. of Social Issues*, **64**, 157-188 (2008).
12. L. S. Vygotsky, *Mind in Society* (Harvard Univ. Press, Cambridge, MA, 1978).
13. J. Love et al., *Dev. Psych*, **41**, 885-901 (2005).
14. J. Barnes, A. Freude-Lagevardi, *From pregnancy to early childhood: early interventions to enhance the mental health of children and families* (Mental Health Foundation, London, 2003).
15. P. Hannon, C. Nutbrown, A. Morgan, *Early literacy work with families* (Sage, London, 2005).
16. P. Sammons et al., *Measuring the impact of pre-school on children's cognitive progress over the pre-school period* (Inst. of Ed./DfES, London, 2002).
17. H. Goldstein et al., *Oxford Review of Education*, **19**, 425-433, (1992).
18. S. Strand, *British Education Research Journal*, **28**, 63-78, (2002).
19. E. Melhuish, et al., *The Effectiveness of Primary Schools in England in Key Stage 2 for 2002, 2003 and 2004*, (Inst. of Ed., London, 2006),  
[www.ioe.ac.uk/schools/ecpe/eppe/eppe3-11/eppe3-11%20pdfs/eppepapers/Tier%201%20full%20report%20-%20Final.pdf](http://www.ioe.ac.uk/schools/ecpe/eppe/eppe3-11/eppe3-11%20pdfs/eppepapers/Tier%201%20full%20report%20-%20Final.pdf)
20. P. Tymms, in *What Does it Mean? The Use of Effect Sizes in Educational Research*, K. Elliot, I. Schagen, Eds. (NFER, Slough, 2004).
21. K. Elliot, P. Sammons, in *What Does it Mean? The Use of Effect Sizes in Educational Research*, K. Elliot, I. Schagen, Eds. (NFER, Slough, 2004).
22. H. Goldstein, S. Thomas, *J. of the Royal Statistical Society*, **159**, 149-163, (1996).
23. P. Sammons, *Educational Research and Evaluation*, **2**, 113-149, (1996).
24. L.J. Schweinhart, H., Barnes, D. Weikhart, D. (Eds.), *Significant benefits: The High/Scope Perry Pre-school Study through age 27* (High/Scope Press, Ypsilanti, Michigan, 1993).