**Grant number:** ES/V004328/1

**Project title:** Does father involvement increase children's educational attainment at primary school?

**MCS datasets used (in the previous wide format):**

The following MCS datasets were used in the wide format:

1. University of London, Institute of Education, Centre for Longitudinal Studies. (2023). Millennium Cohort Study: Age 9 months, Sweep 1, 2001. [data collection]. 14th Edition. UK Data Service. SN: 4683, DOI: <http://doi.org/10.5255/UKDA-SN-4683-6>
2. University of London, Institute of Education, Centre for Longitudinal Studies. (2023). Millennium Cohort Study: Age 3, Sweep 2, 2004. [data collection]. 11th Edition. UK Data Service. SN: 5350, DOI: <http://doi.org/10.5255/UKDA-SN-5350-6>
3. University of London, Institute of Education, Centre for Longitudinal Studies. (2023). Millennium Cohort Study: Age 5, Sweep 3, 2006. [data collection]. 9th Edition. UK Data Service. SN: 5795, DOI: <http://doi.org/10.5255/UKDA-SN-5795-6>
4. University of London, Institute of Education, Centre for Longitudinal Studies. (2023). Millennium Cohort Study: Age 5, Sweep 3, 2006: Foundation Stage Profile and Teacher Survey. [data collection]. UK Data Service. SN: 8785, DOI: <http://doi.org/10.5255/UKDA-SN-8785-1>
5. University of London, Institute of Education, Centre for Longitudinal Studies. (2023). Millennium Cohort Study: Age 7, Sweep 4, 2008. [data collection]. 9th Edition. UK Data Service. SN: 6411, DOI: <http://doi.org/10.5255/UKDA-SN-6411-9>
6. University of London, Institute of Education, Centre for Longitudinal Studies. (2023). Millennium Cohort Study: Age 11, Sweep 5, 2012. [data collection]. 6th Edition. UK Data Service. SN: 7464, DOI: <http://doi.org/10.5255/UKDA-SN-7464-6>

These datasets are now only available to download in the long format so will need to be reshaped to wide format (as per the instructions in syntax 1 ‘data merges and deriving key variables’).

A smaller sub-set of the merged datasets was created for the analysis (as detailed in folder 2: ‘setup’). This folder 2 lists the variables that were retained – some were variables that we derived ourselves in the larger merged dataset. The syntax for the variables that we derived, which were used in our analysis, is available in folder 1: data merges & deriving key variables’). A codebook with further information about each variable is included in folder 2.

**Folder contents:**

1. “data merges & deriving key variables” – this folder contains the syntax for (i) the MCS datasets reshaping and merges, and (ii) syntax for all the derived variables that were used to prepare the sample (father-mother households) and that were used in the Structural Equation Models (SEMs). (Note all this syntax was written when MCS dataset was available to download in the older, wide format).
2. “setup” – this folder contains:
   1. The syntax for cleaning/tidying the smaller subset of the merged MCS sweep 1-5 data in order to prepare it for analysis.
   2. A codebook (excel file) that provides detail about variable renames in the subset of data (SPSS file) so that they could be easily used in Mplus.
3. “subset dataset” – this folder contains:

Syntax for the exploratory factor analysis (EFA), which was performed on this dataset to explore the structure of the father and mother childcare involvement variables (details on how these were derived can be found in folder 1 – syntax for derived variables), the variables that measure child educational attainment (from the Foundation Stage Profile – MCS sweep 3) and variables that measure child behaviour taken from the Strengths and Difficulties Questionnaire (SDQ) – MCS sweep 3. These are the main measures in our Structural Equation Models (SEMs) - in step (4).

1. “Mplus analysis” – this folder contains:
   1. The syntax for (i) the confirmatory factor analysis and (ii) the structural equation models (SEM). These were all run in several ways to ensure the models were as robust as possible. Full details about each model (and a summary of the results of the CFA) are in the ‘summary’ excel documents in each folder.
   2. The syntax for the SEM models using MCS & Foundation Stage Profile data at age 5 and National Pupil Database data at age 7. The models were run with configural, metric and scalar invariance.
2. “Syntax for additional exploratory analysis” – this folder contains:
   1. The syntax for deriving variables for some additional exploratory analysis (i) exploring ‘time with dad’ and (ii) exploring fathers’ and mothers’ school involvement at age 7.

Please note that datasets for the analysis predicting educational attainment at ages 7 (see 4b) is secure and therefore not publicly available.