Too many children left behind: Technical appendix

31 January, 2015

Contents

Tables

Figures

Appendix to Chapter 2

Section A2.1 Overview of the datasets

Cohort selection. The datasets from the US, the UK and Australia each follow a single cohort of children longitudinally, and as a result we draw on these surveys in their entirety. The Canadian dataset is different, in that the NLSCY contains data on multiple cohorts of children, from those born around 1983 up to 2008. For this reason we use only a sub-sample of NLSCY children; namely those age 0 to 3 in the first cycle of the survey in 1994 (see below for details).

i. United States: The Early Childhood Longitudinal Study – Kindergarten Class (ECLS-K)

The ECLS-K selected a nationally representative cohort of children who attended kindergarten in the fall of 1998 and spring of 1999. The ECLS-K followed the selected cohort of kindergarteners from kindergarten to eighth grade. The first data collection of the ECLS-K was conducted in the fall of 1998, followed by succeeding surveys in springs of 1999, 2000, 2002, 2004, and 2007. Additionally, the ECLS-K randomly sub-sampled 30 percent of children from the base-year original sample to conduct the fall first-grade survey (fall 1999) that aimed to study children's summer learning experience (Tourangeau et al., 2002). The ECLS-K data were collected from multiple sources, such as children, parents, teachers, school administrators, and child school records. The child direct assessment was conducted through a Computer-Assisted Personal Interview (CAPI), the parent interview through Computer-Assisted Telephone/Personal Interviews (CATI/CAPI), and the teacher and school administrator interviews through self-administered questionnaires.

Initial sampling design. A multistage probability sampling method was employed to draw primary sampling units (PSUs) within geographic clusters that consisted of counties or groups of counties (overall, 100 PSUs were selected, taking into account the oversampling of Asians and Pacific Islanders, and partitioned into 24 self-representing and 38 non-self-representing strata), then to select schools within the sampled PSUs by using probability proportional to the measure of size that was constructed separately for public and private schools (overall, 1,280 schools were selected; 934 public and 346 private schools), and then to select up to 24 kindergarteners within each selected school by using equal probability systematic sampling (approximately, 21,260 kindergarteners were sampled and yielded about 19,170 and 19,970 completed cases in the fall- and spring-kindergarten surveys, respectively) (here and elsewhere, sample sizes are rounded to the nearest 10 in accordance with NCES rules) (Tourangeau et al., 2000).

The original base-year sample decreased in each subsequent round due to non-response and change in eligibility status (e.g., transferring schools). To collect data from school movers, in the spring first-grade survey, the ECLS-K flagged all children in a random 50 percent subsample of base-year schools, and followed children in the subsample if they transferred from their base-year school at any point in the succeeding surveys, which therefore resulted in a decrease in the original sample (Tourangeau et al., 2009).

Weighting. The ECLS-K weights were developed to take into account differential probabilities of selection at each stage of sampling and to adjust for survey non-response. The ECLS-K provides child, parent, teacher, and school-level weights. Cross-sectional weights were calculated separately at each wave. In contrast, the development of longitudinal weights was a sequential process.

Longitudinal weights for a subsequent wave were developed for children with compete data from both the wave and the previous wave by reflecting adjustments to the weights from the previous wave (see Tourangeau et al. 2000, 2009 for details of construction of the weights).

Language of assessment. Children who spoke a language other than English and passed the Oral Language Development Scale (OLDS) received the direct child assessments in English. The OLDS screener was administered in the first three surveys (i.e., the fall kindergarten, spring kindergarten, and first-grade spring surveys). Children who did not pass the language screener but spoke Spanish received a Spanish translated version of the mathematics assessment and the Spanish version of the OLDS (Spanish OLDS). The parent interview was administered primarily in English, but parents who spoke Spanish, Lakota, Hmong, or Chinese were interviewed using the questionnaire that was translated into the language they spoke.

Sponsors. The primary sponsor of the ECLS-K is the U.S. Department of Education, National Center for Education Statistics (NCES). Due to the confidentiality legislation of NCES, restricted data from the ECLS-K are accessible only through a restricted data license agreement. In accordance with NCES reporting rules, all sample sizes are rounded to the nearest 10.

ii. United Kingdom: Millennium Cohort Study (MCS)

The Millennium Cohort Study (MCS) is a nationally representative, large-scale longitudinal survey of children in the United Kingdom. It is designed to be representative of all children born in 2000-2001[1] who were alive and living in the UK at age nine months.

The MCS has surveyed cohort families five times to date, when the cohort members were roughly 9 months, then 3, 5, 7 and 11 years of age. At each sweep there were separate questionnaires for the Main Carer and the Main Carer's partner (if present in the household). Interviews were carried out using computer-assisted personal interview (CAPI) software on a laptop, and using a confidential computer-assisted self-completion interview (CASI) for sensitive subjects. Direct child assessments of cognitive ability and anthropomorphic measurements were carried out at from age 3 onwards. Additional questionnaires were completed by the study child him/herself, and by the study child's teacher, at ages 7 and 11, and by an older sibling at ages 3 and 5.

Initial sampling design. The MCS sample design allowed for over-representation of families living in areas with high rates of child poverty or high proportions of ethnic minorities which increased the power of the study to describe effects for these groups of families. The geography of electoral wards was used as a sampling frame. There were 11,090 electoral wards in the UK at the time of sampling, which were combined into 9,548 'superwards' in order to eliminate very small units with less than 24 expected births in a year. 398 of these wards and 'superwards' were selected for sampling. The sample is clustered geographically and disproportionately stratified to over-represent: (1) the three smaller countries of the UK (Wales, Scotland and Northern Ireland); (2) areas in England with higher minority ethnic populations in 1991 (where at least 30 per cent of the population were Black or Asian); and (3) disadvantaged areas (drawn from the poorest 25 per cent of wards based on the Child Poverty Index). Of the 398 wards sampled, 50% were in England, 18% in Wales, 16% in

[1] Children eligible for inclusion in the MCS were those born between 1 September 2000 and 31 August 2001 (for England and Wales), and between 23 November 2000 and 11 January 2002 (for Scotland and Northern Ireland).

Scotland and 16% in Northern Ireland; 5% were high ethnic minority wards and 48% were disadvantaged wards.

A list of all nine month old children living in the sampled wards was derived from Child Benefit records provided by the Department of Social Security (subsequently Department for Work and Pensions and then HM Revenue and Customs). Child Benefit claims cover virtually all of the child population except those ineligible due to recent or temporary immigrant status. 27,201 families were identified as eligible from the Child Benefit records, of which 24,180 (90%) were issued to the field. The achieved sample at the first (age nine months) sweep consisted of 18,552 families, yielding an initial response rate of 77%.

Weighting. Two types of weights are provided with the MCS. The sample design or probability weights are used to correct for MCS cases having unequal probabilities of selection that result from the stratified cluster sample design. These are fixed, and do not change over time and vary only by the nine stratum (advantaged and disadvantaged wards in each of the four countries, plus ethnic wards in England only). Non-response weights adjust for possible biases generated by systematic unit non-response and vary by sweep and cohort family. The non-response weight at the current sweep is the inverse of the predicted probability of responding based on a logistic regression model using data from previous sweep(s) (see Hansen 2014 for details).

Language of assessment. Roughly 4% of interviews with the main carer were conducted wholly or partly in a language other than English in Wave 1, falling to under 2% in Wave 5. Common languages used were: Welsh, Arabic, Bengali, Gujarati, Hindi, Kurdish, Punjabi, Somali, Tamil, Turkish and Urdu. The child cognitive assessments were conducted almost exclusively in English. Children in Wales were offered Welsh language versions of some assessments, but this option was taken up by only around 20 to 25 children.

Sponsors. The Millennium Cohort Study is funded by the Economic and Social Research Council and a consortium of Government Departments headed by the Office for National Statistics (ONS). Data are publicly available from the UK Data Archive.

iii. Australia: The LSAC-K

The full title of the study is *Growing up in Australia: The Longitudinal Study of Australian Children*. The study includes two cohorts, a birth cohort and a child cohort. All results here are from Release 5 of the child cohort (the LSAC-K) (March 2014 version). This cohort sampled children who were born between March 1999 and February 2000, with the first wave of interviews being undertaken in 2004. Interviews have then been undertaken every two years. Data collection takes place via a number of instruments, with most information collected from the parent who knows the child best (usually the mother). However information is also collected from the mother's partner, parents living elsewhere, child care workers and teachers and the child themselves.

The study is on-going. For general information on the survey, see AIFS (2013).

Sample design. The sampling frame was drawn from the Australian Health Insurance Commission Medicare enrolment database (after excluding known deaths), with an initial sample of around 9,925 drawn. Around 4,980 children were initially recruited to the study. The sample was stratified and clustered on a postal area basis.

Weighting. Cross-sectional and longitudinal weights are provided by the survey providers. Wave 1 weights are derived by calibrating the responding sample to known population characteristics (including mother's education and language spoken). Subsequent wave weights are derived from models estimating probability of non-response. See Cusack and Defina (2013).

Language of assessment. Most interviews were conducted in English. In wave 1, interpreters (most commonly family members) were used for 3% of interviews.

Sponsors. The LSAC is funded by the Australian government via the Department of Social Services (DSS) and its predecessor departments. The study is conducted in partnership between DSS, the Australian Institute of Family Studies (AIFS) and the Australian Bureau of Statistics (ABS). Wave 1 data collection was undertaken by Colmar-Brunton Social Research and I-view/NCS Pearson. Data collection for Waves 2, 3, 4 and 5 was undertaken by the Australian Bureau of Statistics.

iv. Canada: National Longitudinal Study of Children and Youth (NLSCY)

Unlike the other datasets used in this study, the NLSCY is a long-running longitudinal study of Canadian children, with different cohorts brought into the sample depending on the year. The so-called "original cohort" of 0 to 11 year olds were first surveyed in 1994 (Cycle 1), with information collected every two years since that time. The final round of data collection occurred in 2008 when the original cohort was 14 to 25 years of age (Cycle 8). In addition, each biennial survey adds an "Early Childhood Development" (ECD) cohort, a representative cohort of 0 to 1 year olds born in the relevant year, who were followed up only for a limited number of cycles.

Our research design requires that we observe children longitudinally at both the age of 4 or 5 and at the age of 10 or 11. None of the ECD cohorts were surveyed past the age of 8 or 9, so we draw our sample purely from the original Cycle 1 cohort. Three groups of children within this sample satisfy the survey requirement: those age 0/1, age 2/3 and age 4/5 in the first cycle. We exclude the oldest of these groups from our sample, because they were born substantially before the cohorts used for the other three countries, but pool the other two to maximize sample size.

We categorize measurement occasions in terms of the age of the child throughout this study, which means that for Canada any single measurement occasion uses data collected in two separate cycles. To illustrate, our data on 4/5 year olds are taken from Cycle 2 for the older cohort (age 2/3 in Cycle 1), and from Cycle 3 for the younger cohort (age 0/1 in Cycle 1). Similarly, data on 10/11 year olds come from Cycle 5 for the older cohort and Cycle 6 for the younger cohort. See Table A2.1 for further details.

The NLSCY interview has three broad components. The Person Most Knowledgeable (PMK) about the child and his/her spouse answered the questions in the Adult and the Child Components (for children aged 0-17). The Youth component was used for selected respondents aged 16 and above. The survey has been designed to interview up to two respondents in the Youth or Child category. The NLSCY therefore makes an important distinction between 'households' and 'children/youth'. The interviews were carried out using computer-assisted interviewing (CAI) methods and the use of paper questionnaires. The former consisted of computer-assisted personal interviewing (CAPI) with the help of a laptop as well as computer-assisted telephone interviewing (CATI).

The NLSCY also contained a school collection component in Cycles 1 to 5. After receiving written consent from parents, paper questionnaires were sent to the child's teacher and school principal. In Cycles 1 to 3, these were accompanied by a short mathematics computation test for children in grades 2 and above, which the teacher was asked to administer to the child in class. From Cycle 4 onwards the math test was instead administered at home by the interviewer, in order to minimize disruption to school time.

Initial sampling design. The NLSCY sample is meant to be representative of children and is based upon a sampling frame of all Canadian Households. Typically children are selected from households sampled by Statistics Canada's Labour Force Survey (LFS) that collects labour market data from a national sample of 54000 households on a monthly basis. The LFS is the Canadian equivalent of the US Current Population Survey (CPS). This is based upon a stratified, multistage design using probability sampling at all stages.

Weighting. The NLSCY produces three sets of weights at each cycle: two longitudinal, one cross-sectional. The NLSCY weighting strategy is based on a series of adjustments applied to the NLSCY design weight. Each child's NLSCY design weight is equal to the inverse of his/her probability of selection. The Labour Force Survey (LFS) subweight accounts for all of the sample design information for the LFS sample. For the NLSCY, children are selected from the households selected from the LFS. To reflect these additional sample design steps, the LFS subweight is multiplied by several other factors. The final weight is obtained by applying nonresponse and post-stratification adjustments to the NLSCY design weight. The 'post-stratification' weight adjustment procedure was carried out to ensure consistency between the estimates produced by NLSCY and Statistics Canada's population estimates by age, sex and province.

Language. Interviews were conducted in English or French. The interviewers were given freedom to switch between both languages if the respondent encountered any difficulty in understanding a question or a phrase. Also, a small number of interviews were conducted in languages other than French or English with the interviewer translating the questions into the respondent's preferred language.

Sponsors. The NLSCY was jointly conducted by Statistics Canada and sponsored by Human Resources and Social Development Canada (HRSDC). The data is available in Research Data Centres that were established through a joint effort by Statistics Canada and university consortia. The access to these Data Centres is restricted to researchers working on projects approved by Social Sciences and Humanities Research Council (SSHRC) and Statistics Canada.

Section A2.2 Sample selection and weights

i. General strategy

Surveys from the four countries cover children at different stages of childhood, overlapping for a common 6 year period between roughly the ages of 5 and 11. We distinguish four biennial data collection points within this range, which for brevity we refer to as the age 5, 7, 9 and 11 waves, although in practice the ages of children can vary somewhat at each measurement occasion, as shown in Table A2.1.

In our comparative work (Chapters 3 to 5) we define a *balanced panel sample* in a common way across countries. To be included a child must have been surveyed at *all waves* between ages 5 and 11 inclusive, and have a valid longitudinal weight in the age 11 wave. In practice, estimation samples deviate slightly from the balanced panel sample sizes shown in Table A2.1, due to item non-response on specific variables, but in all countries these discrepancies are minor.

The age 11 longitudinal weight is used in all analyses to correct for both non-random sampling in the initial wave of the survey, and for attrition in later waves, such that estimates can be considered representative of the relevant cohort in the country as a whole. In all cases the weights used were derived by the survey administrators, and details of their construction can be found in the documentation that accompanies each survey. The specific weights used in our analysis are documented in Section A2.2ii below.

Table A2.1 summarizes information on the samples used in our comparative analyses. There are two columns for Canada because, as explained in Section A2.1.iv above, our Canadian sample is a combination of two cohorts, which were surveyed two years apart.

Where analyses were not designed to be comparable across countries (Chapters 2 and 6), we are not restricted to analysis of children age 11 and under, and our sample selection rules are different.

Data on the proficiency levels shown in Figure 2-1 and 2-2 are taken from the 8th grade (age 14) survey wave of the ECLS-K, and are weighted using the 8th grade child longitudinal weight (C1_7FC0). This weight is defined for a sample of about 7,800 children, representative of an underlying population of about 3,840,440 children (all sample sizes for ECLS-K rounded to nearest 10, in accordance with NCES reporting rules). There is a small amount of missing data on 8th grade proficiencies within this sample, such that reading proficiencies are observed for about 7,500 children (96%) and math proficiencies for about 7,340 children (94%).

ECLS-K data on the standardized theta scores used in Figures 2-3 to 2-6 are taken from an unbalanced panel sample that is also the basis of the analyses in Chapter 6. Sample sizes in these figures are about 9,280 and 9,220 for 8th grade math and reading respectively, and about 18,630 and 18,770 for Fall Kindergarten math and reading respectively. Section A6 describes the unbalanced panel sample in more detail.

Table A2.1 Details of cohorts, timing of measurements and analysis samples in the four countries

	US	UK	AU	CA	
				Cohort 1	**Cohort 2**
Cohort birth dates	1992-1993	9/2000-1/2002	3/1999-2/2000	1/1991-12/1992	1/1993-12/1994
Cohort first surveyed	Age 5/6 years (Fall Kindergarten)	Age 9 months (MCS1)	Age 4/5 years (LSAC-K4)	Age 2/3 years (Cycle 1)	Age 0/1 years (Cycle 1)
Sample size at first survey	19,170	18,818	4,980	3,909	4,696
				Combined: 8,605	
Balanced panel sample size	8,370	11,762	3,940	1836	2510
				Combined: 4,346	
Age 5 wave:					
Name	Fall Kindergarten	MCS3	LSAC-K4	Cycle 2	Cycle 3
Fieldwork date	1998	2006	2004	1996/7	1998/9
Mean child age in years	5.7 (SD = 0.4 5 %ile = 5.2 95 %ile = 6.3)	5.2 (SD = 0.3 5 %ile = 4.8 95 %ile = 5.6)	4.9 (SD = 0.2 5 %ile = 4.6 95 %ile = 5.3)	4.9 (SD = 0.6 5 %ile = 4.0 95 %ile = 5.8)	
Age 7 wave					
Name	Spring 1st grade	MCS4	LSAC-K6	Cycle 3	Cycle 4
Fieldwork date	2000	2008	2006	1998/9	2000/1
Mean child age in years	7.2	7.2	7.1	7.0	
Age 9 wave					
Name	3rd grade	NA (no survey)	LSAC-K8	Cycle 4	Cycle 5
Fieldwork date	2002	-	2008	2000/1	2002/3
Mean child age in years	9.2	-	9.1	8.9	
Age 11 wave					
Name	5th grade	MCS5	LSAC-K10	Cycle 5	Cycle 6
Fieldwork date	2004	2012	2010	2002/3	2004/5
Mean child age in years	11.2	11.2	11.2 (10.5 at national testing)	10.9	

Sample sizes for the U.S. are rounded to the nearest 10 in accordance with NCES reporting rules.

ii. Country-specific details

United States. To take into account the complicated sampling structure and longitudinal differences in data collection spanning from kindergarten to fifth-grade, all analyses are adjusted using the kindergarten to fifth-grade longitudinal weights (C1_6FP0), strata (C16FPSTR), and PSUs (C16FPPSU). This weighting adjustment is conducted by employing the svy command in Stata: svyset [pweight = C1_6FP0], strata (C16FPSTR) psu (C16FPPSU); and the singleunit (scaled) option is used to address a stratum that has only a single PSU unit. Stata/MP 13.1 is used for all analyses.

United Kingdom. The balanced panel was selected using the survey indicators for whether a child was present at the age 5, 7 and 11 sweeps (CAOUTC00 == 1 & DAOUTC00== 1 & EAOUTC00 ==1). All comparative analyses in Chapters 3 to 5 are weighted using the age 11 longitudinal weight variable EOVWT2 . Survey design features were also taken into account in all analyses, via Stata's svy command (Stata/MP 13.1), using the variables PTTYPE2 (stratum within country), NH2 (population correction factor) and SPTN00 (fieldwork point number incorporating superwards).

In Chapter 5 we draw on data from the teacher surveys, administered when the study child was 11 years old. Paper questionnaires were sent to teachers of children in England and Wales only. The response rate among those eligible and in scope was 77%. See Gallop et al. (2013) for further details.

Australia. The LSAC study has created a set of weights to correct for differential non-response, and the longitudinal weight variable *defwt* is used to calculate all results here (standardized to sum to the sample size). SAS proc surveyreg and proc surveymeans are used to calculate all variance estimates, taking account of the stratification and clustering of the sample (variables *stratum* and *pcodes*).

In addition to the survey data, we use linked data from the national school testing program (NAPLAN) for our age 11 reading scores. These tests are undertaken in years 3, 5 and 7. We mainly use the year 5 results, but some results for the year 3 and year 7 results are shown in Section A5.4. The note to Figure A5.1 has information on the sample sizes of children with valid data on both the parental education and language/reading scores in each wave.

Canada. All estimates for Canada are adjusted using the Cycle 6 funnel weight *FWTCdW1L*. These longitudinal weights are defined only for children in the "original cohort" from Cycle 1 who participated in all of the first six cycles. As is strongly recommended by Statistics Canada, we use the 1000 bootstrap weights provided in the NLSCY (via Stata's svy command) to conduct correct variance estimation. In contrast to the other surveys, Taylor linearization is not recommended because of the high preponderance of strata with only a single primary sampling unit. See Statistics Canada (2005) for further details.

In the NLSCY, different outcomes were assessed at different ages. At the age 5 wave, all children were assessed in vocabulary, and at the age 9 and 11 waves they were all assessed in math, in all cases via an interviewer-administered test in their own homes. At the age 7 wave, however, the test a child sat depended on their age. Children who were in fact age 6 at the time of the survey, or in grade 1, took the vocabulary test again, while those in grade 2 were now eligible for the math test.

For this reason, estimation sample sizes are halved for outcomes at the age 7 wave, with vocabulary scores available only for the younger children in the cohort and math scores available only for the older children. In addition, the method used to administer the math test switched between the cycles that span our cohort at age 7. In Cycle 3 the test was taken in school (and had a fairly low response rate of 56%) while in Cycle 4 it was administered in the home in the same way as the vocabulary test (and achieved a response rate of 82%). For these reasons results on outcomes from the age 7 wave of the Canadian survey should be treated with caution.

Appendix to Chapter 3

Section A3.1 Parental education categories

Our primary measure of SES is based on the highest qualification attained by a parent who is co-resident with the child at the age 5 survey wave in each country. Where there are two parents present, but the qualifications of one are unknown, we use the highest qualification of the other parent. In our main analyses we distinguish three broad parental education groups: high, equivalent to a US bachelor's degree or more; medium, equivalent to some college in the US; and low, equivalent to a US high school diploma or less. Below we give further details of the coding system used in each of the four countries, and also provide more detailed 6-or 7-category breakdowns of the parental education distributions, which are aggregated to form the primary 3-category variables (see Table A3.1).

United States

To measure parental education, we used a composite indicator for the highest level of parental education provided in the ECLS-K. The composite indicator was created to reflect whichever parent possessed the highest education level based on both fall and spring kindergarten surveys. For each parent, a six category variable was created as follows: "8th grade or below" and "9th to 12th grade" were coded 1; "High school diploma/equivalent" was coded 2; "Voc/Tech program after high school but no Voc/Tech diploma" and "Some college but no diploma" were coded 3; "Voc/Tech program after high school" and "Associate's degree" were coded 4; "Bachelor's degree" was coded 5; "Graduate/professional school/no degree", "Master's degree" and "Doctorate or professional degree" was coded 6. These categories were collapsed into a 3-category definition as shown in Table A3.1.

United Kingdom

There are a wide variety of academic and vocational qualifications in the UK, that vary across the constituent countries (Scotland, for example, has an entirely separate system of qualifications), and across cohorts. Until recently, it was common to rank educational attainment with reference to the two dominant academic qualifications in England and Wales: General Certificates of Secondary Education (GCSEs), taken by all pupils at the end of compulsory schooling at age 16; and A-levels, which prepare pupils for university and are taken at age 18. GCSE and A-level qualifications relate to individual subjects – there is no single 'overall' qualification like a US high school diploma. Pupils usually take 8-10 GCSEs at age 16 (O-levels prior to 1988), with grades C and above typically viewed as a 'good' GCSE or a pass. A-level study normally takes places between ages 16 and 18, and it can be argued this educational stage has more in common with the first year of a US college degree than with the junior and senior years of high school. First, pupils are generally only able to enrol in A-level study if they have reached a sufficient academic standard at GCSE (typically attainment of 5 or more GCSEs at grade C or above, achieved by only roughly half the population). Second, A-level study is specialized in around 3 to 4 academic subjects, and as a consequence covers relatively advanced material. Subjects chosen at A-level determine the university courses, which are also subject-specific, for which an individual is eligible to apply.

In order to recognize the status of the wide range of vocational qualifications that exist alongside this traditional academic route, more recently a comprehensive system of ranking qualifications according to National Vocational Qualification (NVQ) levels has been developed. NVQ1 relates to qualifications equivalent to GCSE grades D and below; NVQ2 to GCSE grades C and above; NVQ3 to A-level; NVQ4 to higher education (HE) qualifications up to an including a bachelor's degree; and NVQ5 to higher degrees and postgraduate qualifications.

The education definitions used in this study are based on derived variables provided in the MCS age 5 wave (MCS3). For each parent, this variable gives the highest NVQ level achieved (1 to 5), plus categories of 'overseas qualification only' and 'none of these'. In general, the NVQ levels correspond to groups from our 3- and 6-category cross-national coding system in their entirety (see Table A3.1).

A complication arises because the NVQ4 category in the MCS dataset combines first (bachelor's) university degrees with HE qualifications below degree level. The former of these forms part of our "high" education group (comparable with a US college degree), while the latter must be classed as "medium" education as they rank below a college degree. (Qualifications in this grouping include, for example, HE diplomas and nursing and teaching qualifications below degree level.) In short, the NVQ4 category must be split between those who have a university degree and those who do not. This information was only collected in the previous waves at ages 9 months and 3 years (MCS1 and MCS2). Among the group coded as having NVQ4 at the age wave 5, therefore, we distinguish individuals recorded as having a university degree at either MCS1 or MCS2, and class these as having "high" education. The remainder are classed as having an HE qualification below degree level and fall into the "medium" education group.

Australia

Parental education level is derived from two data items for each parent: the highest level of schooling completed, and highest level of any non-school qualification. For the latter, respondents were able to choose from the following categories: Postgraduate degree, Graduate diploma/certificate, Bachelor degree, Advanced diploma/diploma, Certificate, Other, Don't know.

The term 'degree' is used in Australia in a similar fashion to the other three countries. 'Bachelor degree' generally corresponds to a US 'College degree', most commonly 3 or 4 years in duration. Graduate diplomas and graduate certificate require a bachelor degree or equivalent as a prerequisite. (Non-graduate) certificates have a number of different levels and are usually vocationally oriented and/or part of a trade qualification. Diplomas and advanced diplomas are para-professionally oriented. Examples include diplomas in nursing (though registered nurses generally require a bachelor degree), diploma of remedial massage, diploma of computing, diploma of digital media technologies (see http://cricos.deewr.gov.au/Course/CourseSearch.aspx). The accreditation process for the diploma level summarises the outcomes required as "Graduates at this level will have specialised knowledge and skills for skilled/paraprofessional work and/or further learning". In contrast the description for Certificate level IV (the highest certificate level) is "Graduates at this level will have theoretical and practical knowledge and skills for specialised and/or skilled work and/or further learning" See http://www.aqf.edu.au/aqf/in-detail/aqf-levels/.

Based on the two data items, a seven category variable for each parent is defined as follows. Parents who have completed a higher degree are coded in category 7. Those with highest qualification of a

bachelor degree are coded in category 6. Those with highest qualification of a diploma are coded to category 5. Those with highest qualification of 'certificate' and who have finished year 12 school or equivalent are coded as category 4. Those with highest qualification of 'certificate' and who have not finished year 12 school or equivalent are coded as category 2. Those with no non-school qualifications (or with a non-school qualification category of 'other' – less than 2%) are coded into category 3 if they have finished year 12 or equivalent, and category 1 if they have not finished year 12 or equivalent. This variable is created for each parent and the maximum of these two variables used. This is then grouped into the high, medium and low categories as shown in Table A3.1.

Canada

Coding for the Canadian cohort draws on derived variables provided by the survey administrators recording the highest level of education obtained by an individual, and also an indicator for whether the individual had graduated from high school. We use this information on the PMK and the partner from NLSCY Cycles 2 and 3. A 7-category variable was defined from the possible responses as follows: 1 (Lowest) = No schooling, Elementary schooling, or Some secondary schooling; 2 = No high school graduation but some post-secondary schooling (Other beyond high school, Some trade school, community college, or university, Diploma/certificate from trade school or community college); 3 = Secondary school graduation only; 4 = High school graduation plus Other beyond high school, Some trade school, community college, or university; 5 = High school graduation plus Diploma/certificate from trade school or community college; 6 = Bachelor degree (includes LLB); 7 (Highest) = Masters, degree in medicine, doctorate.

The low education group combines the first 3 of these categories, and corresponds to those who did not attain a high school diploma, plus those who did attain the diploma but did not undertake any further schooling. The middle education group combines the 4^{th} and 5^{th} categories, and corresponds to those who both graduated high school and completed some further study below degree level. The high education group combines the 6^{th} and 7^{th} categories, and corresponds to those with a bachelor's degree or higher. Breakdowns are given in Table A3.1.

Table A3.1 The detailed educational background of parents of four and five year olds

Country & education group	N	Weighted %
United States		
Low education	3090	37.1
Of which:		
1. Less than high school	*820*	*9.8*
2. High school	*2270*	*27.2*
Medium education	2750	33.1
Of which:		
3. Voc/tech program; some college - no degree/diploma	*1710*	*20.5*
4. Voc/tech program; some college - degree/diploma	*1050*	*12.6*
High education	2490	29.9
Of which:		
5. Bachelor's degree	*1460*	*17.5*
6. Above bachelor's degree	*1040*	*12.5*
Total	8340	100
United Kingdom		
Low education	4,699	40.0
Of which:		
1. NVQ1 or below (e.g. GCSE grade D or below; no qualifications)	*1,840*	*15.7*
2. NVQ2 (e.g. GCSE grade A-C)	*2,859*	*24.3*
Medium education	4,335	36.9
Of which:		
3. NVQ3 (e.g. A-level)	*1,843*	*15.7*
4. NVQ4 NOT university degree (e.g. HE diploma, nursing qual)	*2,492*	*21.2*
High education	2,721	23.2
Of which:		
5. Bachelor's degree	*1,634*	*13.9*
6. NVQ5 (e.g. masters degree, doctorate)	*1,087*	*9.3*
Total	11,755	100
Australia		
Low education	1,748	44.4
Of which:		
1. Less than Year 12	*497*	*12.6*
2. Less than Year 12 plus a certificate qualification	*768*	*19.5*
3. Completed Year 12	*483*	*12.3*
Medium education	826	21.0
Of which:		
4. Completed Year 12 plus a certificate qualification	*421*	*10.7*
5. Diploma	*405*	*10.3*
High education	1,366	34.7

Country & education group	N	Weighted %
Of which:		
6. Bachelor's degree	*640*	*16.2*
7. Post-graduate degree	*727*	*18.4*
Total	3,940	100
Canada		
Low education	932	21.7
Of which:		
1. No secondary qualification	*298*	*6.9*
2. No high school diploma but some post-secondary education	*118*	*2.7*
3. High school diploma and no post-secondary education	*516*	*12.0*
Medium education	2218	51.7
Of which:		
4. High school diploma plus some post-secondary education (no diploma/certificate)	*961*	*22.4*
5. High school diploma plus diploma/certificate from trade school or community college	*1257*	*29.3*
High education	1140	26.6%
Of which:		
6. Bachelor's degree	*812*	*18.9*
7. Post-graduate degree	*328*	*7.6*
Total	4,290	100

Breakdowns shown for balanced panel samples. Sample sizes of the U.S. are rounded to the nearest 10 due to NCES reporting rules.

Section A3.2 Parental income measures

Overview

In all four countries we draw on measures of *gross household income* measured in each of the *age 5, age 7 and age 11 waves*. In some cases, some manipulation was necessary to derive these continuous gross income measures – details are provided separately for each country below. The component income variables were all converted to constant (2011) prices using the national price index, then converted to US dollars using the OECD PPP index for 'actual individual consumption' for 2011.[2] Each component variable was equivalized for household size (N) by multiplying by the scaling factor $2/\sqrt{N}$, so that the numbers presented correspond to the reference income of a family of four. Finally, we take the mean of the real, equivalized income measures at ages 5, 7 and 11. No discounting was used. Where one or more of the component incomes is missing, the mean of the non-missing observations is used. Table A3.2 provides a variety of summary statistics on these time-averaged income measures.

[2] PPP2011 numbers are: US = 1; UK = 0.692; AU = 1.520; CA = 1.250.

Table A3.2 Family income and home ownership

	US	UK	AU	CA
Averaged income (in $2011 for a family of four)				
Mean				
All	69,500	58,100	56,700	63,100
HS or less	38,300	37,000	42,600	40,000
Some post-secondary	60,600	59,800	53,200	58,200
College degree or more	118,300	91,900	76,900	91,600
Median				
All	54,100	52,300	49,100	55,700
HS or less	31,800	30,600	38,000	35,500
Some post-secondary	54,800	57,200	49,700	53,700
College degree or more	98,100	87,900	68,500	81,200
10^{th} percentile				
All	20,500	20,900	29,200	25,200
HS or less	15,200	17,400	18,500	18,600
Some post-secondary	24,500	27,500	23,900	29,500
College degree or more	48,900	52,300	30,300	46,800
90^{th} percentile				
All	133,600	102,800	96,000	107,700
HS or less	69,800	65,200	72,100	64,700
Some post-secondary	101,600	93,400	82,600	91,000
College degree or more	218,700	138,300	128,500	137,200
Parents own home (%)				
All	66	64	69	74
HS or less	49	40	57	53
Some post-secondary	64	72	71	75
College degree or more	87	91	81	87

United States

In the ECLS-K, household income was first measured as a continuous scale at the kindergarten survey in the spring of 1999. Parents answered to the question "What was the total income of all persons in your household over the past year, including salaries or other earnings, interest, retirement, and so on for all household members?" In contrast, household income was measured as a categorical scale at the first- and fifth-grade surveys and thus parents chose one of the following thirteen categories to answer the same question: (1) $5,000 or less, (2) $5,001 to $10,000, (3) $10,001 to $15,000, (4) $15,001 to $20,000, (5) $20,001 to $25,000, (6) $25,001 to $30,000, (7) $30,001 to $35,000, (8) $35,001 to $40,000, (9) $40,001 to $50,000, (10) $50,001 to $75,000, (11) $75,001 to $100,000, (12) $100,001 to $200,000, and (13) $200,001 or more. The ECLS-K used a hot deck imputation

methodology to impute missing values in household income at each wave, which was based on all gathered information about parental socioeconomic status at each wave (e.g., parental education, occupation, and employment status).

We convert the categorical income bands of 2000 and 2004 income measures into continuous income values by assigning a dollar value to each of the income bands. Using the March CPS data for 2000 and 2004, we select families with children under 8 in 2000 and under 12 in 2004, separately; then create 13 income bands using the same boundaries of the income bands in the ECLS-K dataset, separately in 2000 and 2004; and then assign the median gross income value of each band in the CPS data to each income band in the ECLS-K data, separately in 2000 and 2004 (see Table A3.3).

Table A3.3 Median Dollar Values of Each Income Band in the CPS March 2000 and 2004 Data

	2000			2004		
	CPS		ECLS-K	CPS		ECLS-K
	Median	*N*	*N*	*Median*	*N*	*N*
$5,000 or less	1500	722	530	1280	1444	330
$5,001 to $10,000	7560	640	820	7802	1098	440
$10,001 to $15,000	12608	743	1170	12900	1378	670
$15,001 to $20,000	17500	760	1190	17900	1450	790
$20,001 to $25,000	22796	739	1210	23000	1440	880
$25,001 to $30,000	27948	672	1340	28000	1450	770
$30,001 to $35,000	32600	682	1040	32402	1313	700
$35,001 to $40,000	37702	656	1150	37800	1301	840
$40,001 to $50,000	45084	1039	1580	45000	2301	1000
$50,001 to $75,000	60500	1943	2640	61200	4850	1770
$75,001 to $100,000	84661	994	1480	85397	2967	1350
$100,001 to $200,000	123900	878	1120	124801	3317	1140
$200,001 or more	245441	240	380	362411	695	330
Total	39997	10708	15630	45000	25004	11000

Note. Descriptive statistics were adjusted using 2000 and 2004 sampling weights of the CPS and ECLS-K data, respectively. Sample sizes for ECLS-K are rounded to the nearest 10, due to NCES reporting rules.

The annual household income of the ECLS-K does not include transfer benefits. With respect to means-tested government transfers, the ECLS-K only provides information about whether or not the

family receives welfare benefits (e.g., TANF and SNAP) but not the cash value of those benefits; nor does it provide information about EITC receipt or amount. Therefore, we impute the value of TANF, SNAP, and EITC, using data from the March Current Population Survey (CPS), which provides information on receipt of government benefits in the prior calendar year.

The imputation process of the values of such welfare benefits for the household income of the 1999 ECLS-K survey is as follows. The March 1999 CPS data provide information about whether each family received TANF and also the value of their benefits. The Spring 1999 ECLS-K data provide information about whether each family received TANF. Therefore, we predict TANF values among those who received TANF in the CPS data, and then use the regression parameters to predict values among those who reported receiving TANF in the ECLS-K data. In the first step, using the March 1999 CPS data, we run a regression of TANF values on selected demographic variables within a sample of TANF recipients:

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

where Y_i is a continuous measure of TANF benefits for a reference respondent *i* of each family; and X_i indicates a vector of selected demographic variables, including family income, whether to receive TANF, number of children under age 18, number of family members, marital status, poverty status, ethnicity, education, age, and region of country. In the second step, using the ECLS-K data and the same set of demographic variables as those in the CPS data, we predict TANF values within a sample of TANF recipients.

The CPS data provide information about SNAP receipt and the market value of benefits received. The ECLS-K data provide information about SNAP receipt. Therefore, we predict SNAP values among those who received SNAP in the CPS data, and then use the regression parameters to predict values among those who received SNAP in the ECLS-K. In the first step, using the CPS data, we run a regression of SNAP market values on selected demographic variables within a sample of SNAP recipients:

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

where Y_i is a continuous measure of SNAP market value for a reference respondent *i* of each family; and X_i indicates a vector of selected demographic variables, including family income, whether to receive TANF, number of children under age 18, number of family members, marital status, poverty status, ethnicity, education, age, and region of country. In the second step, using the ECLS-K data with the same set of demographic variables as in the CPS data, we predict TANF values within a sample of TANF recipients.

Respondents in the CPS were not asked about the amount of EITC they received, so values for these variables were imputed by the Census Bureau using their tax model, which simulates individual tax returns to produce estimates of federal, state, and payroll taxes. The model incorporates information from non-CPS sources, such as the Internal Revenue Service's Statistics of Income series, the American Housing Survey, and the State Tax Handbook. To predict EITC values in the ECLS-K dataset, we first predict the likelihood that people receive EITC among those who were likely eligible for EITC in the CPS dataset, then apply the regression parameters to predict receipt among those who were likely eligible for EITC in the ECLS-K dataset, constraining total recipiency rates to match in

the two data sets. We then give each eligible family a predicted EITC value using their own family income data. In the first step, using EITC rules (Hotz & Scholz, 2001), we identify people who were likely eligible for EITC benefits based on their family income, whether or not they have income from working, and their number of children, separately in the CPS and ECLS-K data sets. In the second step, to predict the likelihood of EITC incidence in the CPS data, we run a linear probability regression for EITC receipt on selected demographic variables among those who were likely eligible for EITC:

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

where Y_i is a binary indicator with a value of 1 if a respondent *i* received EITC and 0 otherwise among those who were likely eligible for EITC; and X_i indicates a vector of selected demographic variables, including total family income, number of children under age 18, number of family members, marital status, poverty status, ethnicity, education, age, region of country, whether to receive SNAP, and whether to receive TANF. In the third step, using the ECLS-K data with the same set of demographic variables as in the CPS data, we predict the likelihood of EITC incidence among those who were likely eligible for EITC:

$$\hat{Y} = Pr(Y = 1 \mid X)$$

where $\hat{Y}$ indicates the predicted probability that an eligible respondent *i* receives EITC, given selected demographic variables are adjusted. In the fourth step, constraining total recipiency rates to match in the two data sets, we select EITC recipients in the ECLS-K data. In the fifth step, using 1999 EITC parameters (see Table A3.4), we calculate potential EITC values for all families in the ECLS-K 1999 spring dataset, and then give each eligible family the potential EITC value.

Next, we repeat the same procedures detailed above to adjust these income values for 2000 and 2004 by reflecting the value of government benefits values. Using the March CPS data from 2000 and 2004, we repeat the same procedures described above to predict TANF, SNAP, and EITC values, separately in 2000 and 2004 (see Table A3.4 for EITC parameters for 2000 and 2004).

Table A3.4 Earned Income Tax Credit Parameters, 1999, 2000, 2004

Family type	Phase-in rate (%)	Phase-in range ($)	Maximum credit ($)	Phase-out rate (%)	Phase-out range ($)
1999					
One child	34.0	0-6,800	2,312	15.98	12,460-26,928
Two or more	40.0	0-9,540	3,816	21.06	12,460-30,580
None	7.65	0-4,530	347	7.65	5,670-10,200
2000					
One child	34.0	0-7,140	2,353	15.98	12,690-27,413
Two or more	40.0	0-10,020	3,888	21.06	12,690-31,152
None	7.65	0-4,760	353	7.65	5,770-10,380
2004					
One child	34.0	0-7,660	2,604	15.98	14,040-30,338
Two or more	40.0	0-10,750	4,300	21.06	14,040-34,458
None	7.65	0-5,100	390	7.65	6,390-11,490

Source: Tax Policy Center (2013) *Historical EITC parameters*. Washington, DC: Urban Institute and Brookings Institution.

UK

Unlike the other surveys, the MCS income question elicits details of net, rather than gross income, so a conversion needs to be made for comparability with the other countries. In the questionnaire, a number of questions ask about whether income is received from various sources, then our focal survey question follows: "This card shows incomes in weekly, monthly and annual amounts. Which of the groups on this card represents you [^and your husband/wife]'s total take-home income from all these sources and earnings, after tax and other deductions. Just tell me the number beside the row that applies to your joint incomes." Respondents were asked to pick a category from 19 bands, with different bands applicable to single- and two-parent families. These were updated over time so that, for example, the top band for two-parent families increased from "80,000 or more" to "150,000 or more" between the age 5 and age 11 waves.

As is usual with income data, item non-response was an issue, so imputation was conducted by the MCS using interval regression. In addition to filling in missing values, this procedure provides a continuous income value that falls between the boundaries indicated by the respondent, but that need not be the same for all families in a particular band. Predictor variables included age, education, housing tenure, state benefit receipt and household composition. See Hansen (2014) for further details.

In order to cross-walk from these continuous net income variables to their gross counterparts, we drew on data from the Family Resources Survey (FRS) from 2006, 2008 and 2011. The FRS is a major government survey designed to collect information on the incomes and circumstances of private households in the UK. Approximately 20,000 households per year are interviewed, and details are collected on both net and gross incomes. To do the mapping, records for all benefit units containing a child aged 16 or under were first extracted from the relevant FRS. Then the parents in that benefit unit were identified from the household roster (a maximum of two were found in all cases), and variables containing their individual net and gross weekly incomes were extracted (variables NINDINC and INDINC), multiplied by 52, and summed to give total net and gross parental income respectively.

First we created a 10-part linear spline from the household net income variable provided in the MCS, with knot points at the decile boundaries. These knot points were then transferred across to create another 10-part linear spline, this time for the FRS net income variable. The FRS gross household income variable was then regressed on these splines using quantile regression, so that coefficients mapping each net income value to its associated median gross income value could be extracted. Finally, these coefficients were applied to the original splines from the MCS net income variable, and these predictions were used as our estimates of gross household income.

Australia

The Australian income concept is the usual gross weekly income of the child's mother and father (times 52.14). Negative incomes are set to zero before aggregating. Any income of other household members is not included.

Respondents were first asked whether or not they (or their partner) received income from various sources. Then a single question was asked for the total: "Before income tax is taken out, how much does … usually receive from all sources in total?" The relevant time period was recorded and amounts converted to a weekly amount.

Canada

Several income questions were asked during the NLSCY household interview. Information on income, broken down into three sources, was asked for the person most knowledgeable (PMK) and his or her spouse. Those three income sources are: wages and salary, self-employment net income, and Employment Insurance benefits. (E.g. "During the past 12 months, what was your personal income from wages and salaries (before deductions)?") Information on income, broken down into four sources was also asked at the household level. Those four income sources are: Child Tax Benefit/National Child Benefit, social assistance, child and spousal support and other sources. Total household income was derived by the NLSCY as the sum of these ten sources of income.

We use the total household income variable provided with the NLSCY, which contains imputation done by Statistics Canada to address a number of reporting problems. Some respondents refused to provide answers to the detailed income questions. Among those, some provided an estimate of their total household income or an estimate of their income using ranges. Finally, for those who responded, amounts declared in the income section were sometimes incoherent with answers provided in the labour force section (for example, an individual might have reported working in the past 12 months according to answers provided in the labour force section but no wages or self-employment income were reported in the income section). Income imputation was carried out to fill in the holes resulting from partial nonresponse as well as to rectify, when possible, these incoherencies. Imputation was also done for households whose total reported income was less than $6,000. See Statistics Canada (2005) for further details.

Section A3.3 The association between education and income across countries

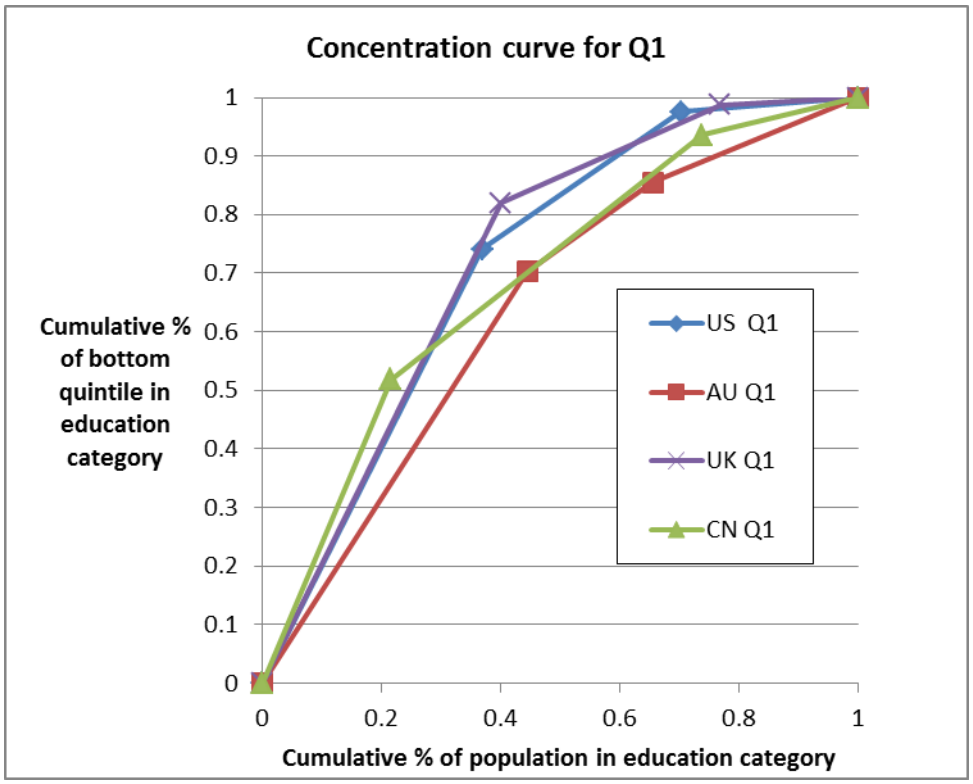
In Chapter 3 we show the median income levels of the different parental education groups. These vary most strongly in the US, followed by the UK, Canada and then Australia. This pattern of income gaps reflects both the greater income inequality in the US and the UK, but also the relatively low association between income and education in Australia (and, to a lesser extent, Canada).

For example, 45 per cent of Australian children are in the lowest parental education group, but among Australian children who are in the bottom income quintile 68 per cent are in this lowest education group – a concentration ratio of 1.5 (=68/45). In the US, on the other hand, 37 per cent overall are in the bottom education group, and 74 per cent of those in the bottom income quintile are in this group – a concentration ratio of 2.0. (The concentration ratios for the UK and Canada are 2.1 and 2.4 respectively). Australia is also an outlier for the top income quintile, where the concentration ratios for the US, UK, AU and CN are 2.6, 2.8, 1.9 and 2.3 respectively. Table A3.5 and Figure A3.1 provide more details of the relationship between income and education. At the top of the distribution, part of the Australian pattern reflects the high education but low incomes of migrants. Among native-born Australians, the concentration ratio is 2.2 for the top quintile.

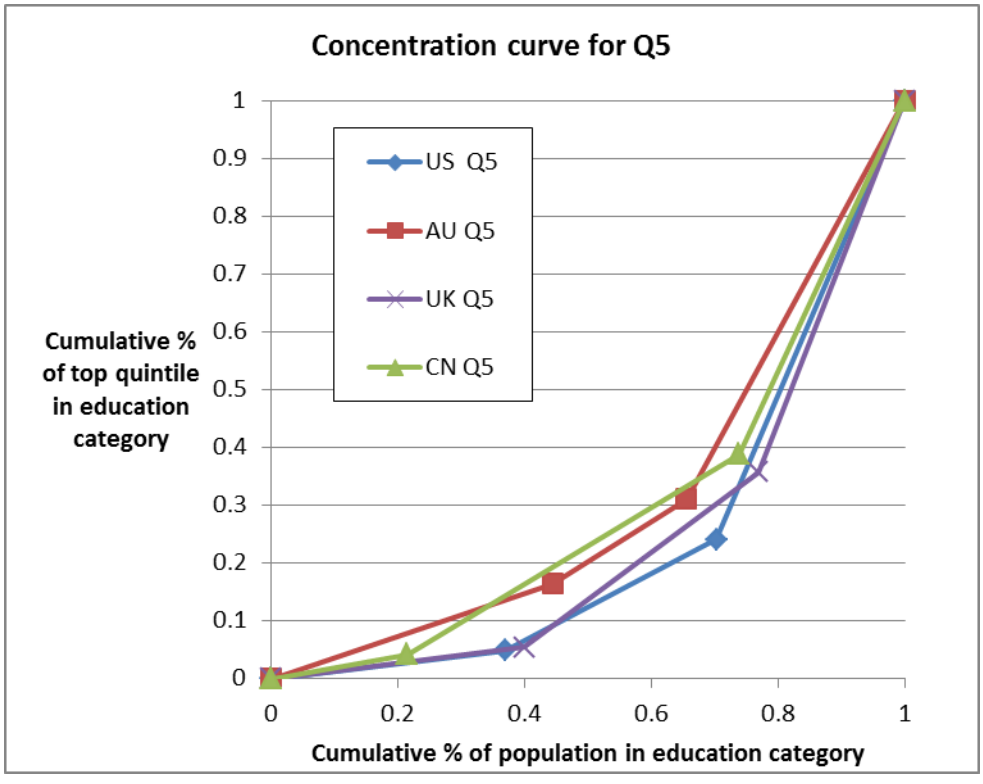
Table A3.5 The association between parental education and income

(A) Overall parental education distribution (%)				
High school diploma or less	37	40	44	22
Some post-secondary	33	37	21	52
College degree or more	30	23	35	27
Total	100	100	100	100
(B) Parental education distribution in bottom income quintile (%)				
High school diploma or less	74	82	70	52
Some post-secondary	23	17	15	42
College degree or more	2	1	14	6
Total	100	100	100	100
Relative concentration in the bottom quintile = (B) / (A)				
High school diploma or less	2.0	2.1	1.6	2.4
Some post-secondary	0.6	0.4	0.3	1.9
College degree or more	0.1	0.0	0.3	0.3
(T) Parental education distribution in top income quintile (%)				
High school diploma or less	5	5	16	4
Some post-secondary	19	30	15	35
College degree or more	76	64	69	61
Total	100	100	100	100
Relative concentration in the top quintile = (T) / (A)				
High school diploma or less	0.1	0.1	0.4	0.2
Some post-secondary	0.6	0.8	0.7	0.7
College degree or more	2.6	2.8	2.0	2.3

Figure A3.1 Concentration curves for top and bottom quintile



Source: Panels A and B of Table A3.5



Source: Panels A and T of Table A3.5

Section A3.4 Demographic characteristics

Table A3.6 The family background of four and five year olds

	US	UK	AU	CA
Mother was a teenager when child was born (%)				
All	12	6	3	3
HS or less	21	12	5	7
Some post-secondary	12	4	3	2
College degree or more	3	0	0	1
Single mother at age 4/5 (%)				
All	22	20	14	15
HS or less	33	34	24	31
Some post-secondary	23	15	8	15
College degree or more	8	5	5	4
Step-family at age 4/5 (%)				
All	13	5	3	4
HS or less	15	7	4	5
Some post-secondary	15	6	4	4
College degree or more	9	2	2	2
Both parents at age 4/5 (%)				
All	65	74	83	81
HS or less	52	59	73	64
Some post-secondary	63	79	88	81
College degree or more	83	93	93	94
Immigrant parent (%)				
All	20	15	35	22
HS or less	28	16	28	18
Some post-secondary	15	13	35	19
College degree or more	16	19	43	33

Table A3.7 Health status of mothers and children

	US	UK	AU	CA
Mother poor/fair health (%)				
All	11	14	10	4
HS or less	18	20	12	4
Some post-secondary	9	12	9	4
College degree or more	4	7	8	2
Mother mental health problems (standardized score)				
All	0.00	0.00	0.00	0.00
HS or less	0.18	0.21	0.08	0.23
Some post-secondary	0.00	-0.07	0.00	-0.03
College degree or more	-0.22	-0.21	-0.09	-0.14
Child low birthweight (%)				
All	8	8	7	7
HS or less	9	9	8	10
Some post-secondary	8	7	6	6
College degree or more	7	6	6	4
Child long-standing disability/illness (%)				
All	16	20	13	20
HS or less	16	22	14	21
Some post-secondary	16	19	13	21
College degree or more	16	18	13	17

Table A3.8 Parenting for four and five year old children

	US	UK	AU	CA
Parent reads to child every day (%)				
All		44	51	46	66
HS or less		33	43	32	55
Some post-secondary		43	51	46	65
College degree or more	58	63	62	75	

Variable definitions

Teen mother. Dummy variable equal to 1 if study child's mother was age 19 years or less at the time of the child's birth.

Family structure. These variables were coded using derived variables on the number of parents in the household at the age 5 survey wave, and their biological relationship to the study child. A child is coded as living with both parents if there are two parents in the household and both are reported to be the biological parents of the child. All other couple families are defined as step families. Where only one parent lives in the household the child is coded as having a single parent.

Immigrant parent. Mother or father born outside of the country in question. Missing counted as native-born.

Mother in poor/fair health.
"Now, I would like to ask you about your health. In general, would you say that your health is: 1 Excellent; 2 Very good; 3 Good; 4 Fair; 5 Poor (US)
"I would now like to ask about your health. How would you describe your health generally. Would you say it is: 1 excellent, 2 very good, 3 good, 4 fair, 5 or poor?" (UK)
"In general, would you say your own health is: 1 Excellent; 2 Very good; 3 Good; 4 Fair; 5 Poor" (AU)
"The following questions ask about your health. In general would you say your health is: 1 Excellent; 2 Very good; 3 Good; 4 Fair; 5 Poor" (CA)

Maternal mental health.
In the US, the instrument used was the 12-item (short) version of the Centre for Epidemiological Studies Depression scale (CES-D, Radloff, 1977). Items are scored from 0 and 3 and then summed to give a total score ranging from 0 to 36, with higher scores indicating higher levels of depression. A cut-off score of 16 or above was used to define those with "severe depression" (Radloff, 1977, p. 393).
In both the UK and Australia, parents' mental health was measured using the Kessler 6 scale (Kessler et al. 2002) which was administered to parents as a computerised self-report. The Kessler 6 scale provides a measure of psychological distress from the respondent's report of how often over the last 30 days they had felt depressed, hopeless, restless or fidgety, that everything they did was an effort, worthless and nervous. For each item the respondent indicates whether they have felt this way none, a little, some, most or all of the time which are scored from 0 to 4, respectively. The questions are summed to form a 24-point scale, with higher scores indicating greater distress. The Kessler scale has been evaluated as a screen for prevalence of serious mental illness within a community population of US adults. In this evaluation a score of 13 or more was indicated as an appropriate reference level to estimate the prevalence of serious mental illness in the population (Kessler et al. 2003), and this is the cut-off used in Table 3.3 to indicate a "mental health problem". We also use a standardized version of the raw score in our analysis. (The scoring system reported in Australia is different to the one reported here, but results in the same cut-off definition in practice.)

In Canada the instrument used was the 12-item (short) version of the Centre for Epidemiological Studies Depression scale (CES-D). Items are scored from 0 and 3, then summed, to give a total score with a maximum of 36, and with higher scores indicating greater depression. A cut-off score of 9 or above was used to define those with a "mental health problem" (Letourneau et al. 2006, p. 49).

Child birth weight. Birth weight was reported retrospectively by the child's mother. Weights below 2.5 kg were classed as low birth weight.

Child longstanding disability/illness.
In the U.S., a binary indicator with a value of 1 if the child currently has a disability was provided in the ECLS-K, which was constructed based on a set of questions: "Has {CHILD} been evaluated by a professional in response to {his/her} ability to pay attention or learn?" "Has {CHILD} been evaluated by a professional in response to {his/her} overall activity level?" "Has {CHILD} been evaluated by a professional in response to the use of {his/her} limbs?" "Has {CHILD} been evaluated by a

professional in response to {his/her} ability to communicate?" "Have you had {CHILD}'s hearing evaluated by a professional?" "Has {CHILD}'s vision been evaluated by a professional?" "Prior to this school year, did {CHILD} ever receive therapy services or take part in a program for children with disabilities?"
"Does [^Cohort child's name] have any longstanding illness, disability or infirmity? By longstanding I mean anything that has troubled [^Cohort child's name] for a period of time or is likely to affect [^Cohort child's name] over a period of time."(UK)
"Does the child have a condition which has lasted or is expected to last for at least 12 months which causes to use medicine prescribed by a doctor, other than vitamins, or more medical care, mental health or educational services?" (AU)
"In the following questions long-term conditions refer to conditions that have lasted or are expected to last 6 months or more and have been diagnosed by a health professional. Does [child] have any of the following long-term conditions? 1 Allergies; 2 Bronchitis; 3 Heart condition or disease; 4 Epilepsy; 5 Cerebral Palsy; 6 Kidney Condition or disease; 7 Mental handicap; 8 Any other long term condition." (CN)

Parents own home. Dummy variable equal to 1 if parents report they own the home in which they live outright or with a mortgage.

Child read to every day. Question wording:
Response 4 to: "How often do you or any other family member read books to child in the past week? (1) not at all, (2) once or twice a week, (3) 3 to 6 times a week, (4) everyday." (US)
Response 1 to: "How often do you... read to [^Cohort child's name]? 1 Every day; 2 Several times a week; 3 Once or twice a week; 4 Once or twice a month; 5 Less often; 6 Not at all" (UK)
Response 3 to: "In the past week, on how many days have you or someone in your family done the following with child? Read to child from a book? 0 None; 1 1 or 2 days; 2 3-5 days; 3 Every day (6-7 days)" (AU)
Response 7 or 8 to: "Currently, how often do you or another adult read aloud to him/her or listen to him/her read or attempt to read aloud? 1 Never or rarely; 2 Less than once a month; 3 Once a month; 4 A few times a month; 5 Once a week; 6 A few times a week; 7 Daily; 8 Many times each day." (CA)

Child watches 3+ hours of TV a day. Question wording:
"How many hours a day does child usually watch TV or videos on school days? Frequency (open question)" (US)
Response 4, 5 or 6 to: "On a normal week day during term time, how many hours does [^Cohort child's name] spend watching television, videos or DVDs? Please remember to include time before school as well as time after school. 1 None; 2 Less than an hour; 3 1 hour to less than 3 hours; 4 3 hours to less than 5 hours; 5 5 hours to less than 7 hours; 6 7 hours or more." (UK)
Response 4 or 5 to: "About how many hours on a typical weekday, would you say that child watches TV or videos at home? 1 Does not watch TV or videos; 2 Less than one hour; 3 1 up to 3 hours; 4 3 up to 5 hours; 5 5 or more hours" (AU)
"On average, how many hours a day does he/she watch T.V.?" Answers coded in half hour increments. (CA)

Parental employment.

US: Employment status of both parents was measured based on questions reported by parent 1 on whether each parent did any paid work in the past week ("During the past week, did {you/{NAME}} work at a job for pay?") and work hours of each parent ("About how many total hours per week {do you/does {NAME}} usually work for pay, counting all jobs?").

UK: Employment status taken from a survey-derived variable, using questions on whether each respondent did any paid work (or was on holiday) in the last 7 days. Work hours derived from responses of each parent to the question: "About how many hours a week do you usually work in your main job, excluding meal breaks but including any usual paid overtime?"

AU: Employment status of both parents reported by Parent 1. Employment status is current, using standard labor force survey question module. Fathers' hours are from the question "How many hours per week does (father) usually work in all jobs, including any paid or unpaid overtime? (If irregular hours, average over the last 4 weeks. Do not include travel time to and from place of work.)" Corresponding question for mothers.

Section A3.5 Family background in the US among non-Hispanic whites

Table A3.9 The family background of four and five year olds in the US, full sample vs. non-Hispanic Whites

	Full Sample	Non-Hispanic Whites
Teen mother at birth (%)		
All	12	9
HS or less	21	18
Some post-secondary	12	10
College degree or more	3	2
Single mother at age 4/5 (%)		
All	22	16
HS or less	33	29
Some post-secondary	23	16
College degree or more	8	7
Step-family at age 4/5 (%)		
All	13	11
HS or less	15	16
Some post-secondary	15	13
College degree or more	9	6
Both parents at age 4/5 (%)		
All	65	73
HS or less	52	55
Some post-secondary	62	71
College degree or more	83	86

Appendix to Chapter 4

Section A4.1 Outcome standardization and reliability

The outcome scores used in the four countries generally have an arbitrary mean and variance. We convert each score to a z-score (mean zero and unit variance) before analysis. This involves the following steps. All calculations are undertaken for the subpopulation of the balanced sample where there are valid values on the relevant outcome variable, and use the balanced panel weight.

1) The residuals from a regression of the variable predicted as cubic function of the child's age at the time of the time of interview are calculated. This removes any variation due to different ages within each interview round. Child age is measured in months in Canada, and in days in all other countries.

2) These residuals are divided by their weighted standard deviation.

3) The resulting variable has mean zero and a weighted standard deviation of one.

As is discussed in Reardon (2011), differential reliability of the tests used in different countries could lead to spurious differences in estimates of SES gaps across countries. In each country, the standardized score used as the basis for the gaps is given by $Y^* = (Y - \bar{Y})/\hat{\sigma}$, where Y is the observed score and $\hat{\sigma}$ is its sample standard deviation (in practice after adjustment for differences in age at testing). Measurement error in test scores will tend to inflate the variance of the test score distributions (thereby inflating $\hat{\sigma}$), meaning that the achievement gaps measured in standard deviation units will be biased toward zero. If the gaps in different countries are measured with tests that have different amounts of measurement error, then the amount of bias will not be the same in each measure of the gap, leading to potentially erroneous inferences regarding differences in the magnitudes of the gaps across countries.

If accurate information were available on the reliabilities of the different tests, our original gap estimates could be corrected for measurement error by multiplying by $\frac{1}{\sqrt{r}}$, where r is the reliability of the test. This would yield estimates of the true gaps, and eliminate any bias in the comparisons that may arise from differential reliability of the tests. Ideally, such measures of reliability would encompass all factors that might lead to a deviation between the observed scores and the child's true underlying ability. This might include the within-test variation arising from the range of question items as well as other factors such as day-to-day variation in the child's attentiveness and the testing environment. In practice, when test reliabilities are reported, they usually account for only within-test variation. Nonetheless, this is the component that we might expect to vary most between studies.

We have comprehensive within-test reliability estimates for the U.S. and some estimates for Australia. Here, we use the U.S. estimates as a reference point and conduct a sensitivity analysis exploring the degree of measurement error that would need to be present in the other countries' test scores to overturn the finding of smaller SES gaps in the US. The results are presented in Table A4.1.

Across all comparisons of the high-low (HL) SES gaps in other countries against the US, reliabilities in excess of 0.61 (and in many cases much lower) would still lead to estimated gaps lower than in the US. Given the technical expertise drawn on by all the surveys, we judge reliabilities anywhere near this low to be highly unlikely. (Of the 49 survey test scores with documented reliabilities provided in Reardon, 2011, all were equal to 0.70 or higher, and 37 were equal to 0.80 or higher.) It follows that we can be very confident that the US would have the largest SES gaps of the four countries at all ages and outcomes, even after adjustment for any reasonable amount of measurement error.

As a stricter test we calculate the minimum reliability need to maintain a *statistically significant* gap with the US (at the 5% level). These are always lower than 0.85, but in three cases they are above 0.75 (CA age 7 reading = 0.84; UK age 5 reading = 0.79; UK age 7 reading = 0.77). Recall that the age 7 Canadian samples are half the size of the usual balanced panel sample, due to differences in the tests given to the younger and older halves of the cohorts, an issue that leads to a large standard error on the estimates of the Canadian SES gap in reading at this age. The reliability required for the US-Canadian difference to become insignificant at this stage is still on the low side, but it is not implausible. With this exception, all comparisons between the US and Canada, and the US and Australia, would lead to significantly higher SES gaps in the US under any reasonable assumption about the degree of measurement error (including, of course, the actual reported Australian measurement error).

This is also the case for age 11 readings gaps and age 7 math gaps in the UK. In theory it is possible that high (but not implausible) degrees of measurement error in the UK reading scores at ages 5 and 7 could account for the finding of significant differences with US on these measures, but we would expect reliabilities well in excess of 0.8 for these UK measures, in which case significant differences remain.

Table A4.1 Minimum reliabilities required to detect different SES gaps between the US and other countries

Outcome and country	[1] HL SES gap (from Ch 5)	[2] Documented reliability	Minimum reliability required to detect	
			[3] Any gap with US	[4] Sig lower gap than US
Age 5 language/reading				
US	1.00	0.92	-	-
UK	0.79	-	0.59	0.79
AU	0.47	0.76	0.21	0.31
CA	0.62	-	0.37	0.60
Age 7/9 language/reading				
US	0.98	0.94	-	-
UK	0.78	-	0.61	0.77
AU	0.61	-	0.37	0.51
CA	0.69	-	0.48	0.84
Age 7/9 math				
US	0.92	0.95	-	-
UK	0.64	-	0.45	0.60
CA	0.36	-	0.16	0.34

Age 11 language/reading				
US	1.01	0.93	-	-
UK	0.67	-	0.42	0.56
AU	0.73	0.87	0.49	0.65
Age 11 math				
US	0.94	0.95	-	-
AU	0.68	0.92	0.51	0.71
CA	0.57	-	0.36	0.62

Sources U.S. reliability estimates: Tourangeau et al. (2009), Table 3.10, p.3-26. Australian reliability estimates: Age 5, Rothman (2005); Age 11, Australian Curriculum, Assessment and Reporting Authority (2014), Table 5.2, year 7. (Note, the age 11 reliabilities are for a different wave of the NAPLAN data to that reported here).

Minimum reliabilities in column 3 of Table A4.1 were calculated using the following formula:

$$\frac{G_{US}}{\sqrt{r_{US}}} = \frac{G_j}{\sqrt{r_{MIN,j}}} \leftrightarrow r_{MIN,j} = \frac{G_j^2}{G_{US}^2} r_{US}$$

Where G_i is the high-low SES gap for country i calculated from our data (shown in column 1), and r_{US} is the reliability for the US measure reported in the ECLS-K documentation and reproduced in column 2.

Minimum reliabilities in column 4 of Table A4.1 were calculated such that

$$\Pr(Z > |z|) = 2\{1 - \Phi(|z|)\} = 0.05$$

Where

$$z = \frac{\frac{G_{US}}{\sqrt{r_{US}}} - \frac{G_j}{\sqrt{r_{MIN,j}}}}{\sqrt{\left(\frac{\sigma_{US}^2}{r_{US}} + \frac{\sigma_j^2}{r_{MIN,j}}\right)}}$$

$\Phi(\cdot)$ is the standard normal c.d.f. and σ_i^2 is the squared standard error of the estimate of the high-low SES gap for country i. ($r_{MIN,j}$ found by grid search methods and rounded up to 2 d.p.)

Section A4.2 Cross-referencing findings for the US with ECLS-B

i. Gaps in reading vs language test scores at age 5

The UK, Australian and Canada surveys include vocabulary skills as a key cognitive outcome measure at age 5. The ECLS-K instead publishes an overall reading scale (though 8 of the 72 items are vocabulary items). Is this likely to bias our estimates of SES gaps?

We assess this by comparing the SES gaps for similar outcome measures in another US survey, the ECLS-B. This survey followed a more recent cohort of children from birth up to the start of school and included tests of vocabulary and literacy (early reading) for children at around age 4. In Bradbury et al (2012), we provided information on the SES gaps for these measures, and this is shown below in Table A4.2.

The SES measures used in this earlier study are slightly different to those used in this book. The "low education" group did not include high school graduates and the family income was averaged over the years prior to the child turning five rather than after it. Data for the middle group is also unavailable. Furthermore, the outcomes in the study were measured in the year prior to kindergarten, not during kindergarten itself as in the current study. Nevertheless, this earlier data can be used to assess whether the SES gap in literacy is the same as for vocabulary.

Table A4.2 SES gaps in vocabulary and literacy in preschool in the ECLS-B

	By parent education			By income			
	Low mean	High mean	H-L	Q1 mean	Q5 mean	Q5-Q1	Corr
Vocabulary	-0.61	0.60	**1.21**	-0.46	0.62	**1.08**	
	[-0.71, -0.51]	[0.53, 0.67]	[1.09, 1.34]	[-0.55, -0.38]	[0.54, 0.70]	[0.98, 1.18]	0.37
Literacy	-0.49	0.74	**1.23**	-0.50	0.69	**1.19**	0.40
	[-0.59, -0.39]	[0.68, 0.80]	[1.13, 1.32]	[-0.58, -0.42]	[0.62, 0.76]	[1.10, 1.28]	

Source: Bradbury et al (2012) [Online-appendix]. 95% confidence intervals shown.

With respect to parental education, the top-bottom gaps are almost identical for vocabulary and literacy. The scores for children with both low and high parental education are lower for vocabulary, which implies that the middle education group must have higher scores (since these are all z-scores with mean zero). This in turn implies that the gap in the bottom half might be larger (and in the top half smaller) if we were to use a vocabulary score in the US rather than a literacy score.

For parental income, the means of each group are similar for the two outcome measures, and the gap between the top and bottom quintile groups is slightly (but not significantly) larger for the literacy measure.

Overall, we conclude that the standardized SES gaps (particularly the overall top-bottom gap) is very similar for vocabulary and literacy in the US.

ii. Did SES gaps change between the ECLS-K cohort (kindergarten 1998) and the ECLS-B cohort (kindergarten 2006/7)?

The US (and Canadian) data are for a cohort of children born earlier than the Australian and UK children in our study. Would our conclusion about a larger US SES gap change if we were to use US data from a later cohort? We can assess this by comparing our US results with comparable results from the ECLS-B. Here, we are able to define the income and education groups in a similar way, and to assess the SES gaps in the overall reading and math scores.

Table A4.3 SES gaps in reading achievement across cohorts

	Age 5 income Q gaps			Parent education gaps		
Cohort/age	Q3-Q1	Q5-Q3	Q5-Q1	M-L	H-M	H-L
ECLS-K Fall K wave	0.60	0.69	1.29	0.46	0.54	1.00
	[0.47, 0.74]	[0.57, 0.81]	[1.16, 1.42]	[0.36, 0.56]	[0.45, 0.64]	[0.89, 1.12]
ECLS-B Kindergarten wave	0.50	0.63	1.13	0.38	0.49	0.87
	[0.43, 0.57]	[0.56, 0.70]	[1.06, 1.20]	[0.33, 0.44]	[0.43, 0.54]	[0.82, 0.92]
ECLS-B Preschool wave	0.30	0.42	0.72	0.39	0.50	0.89
	[0.23, 0.37]	[0.34, 0.49]	[0.64, 0.79]	[0.34, 0.45]	[0.45, 0.56]	[0.84, 0.95]

95% CIs in brackets. Income measured in kindergarten. Parent education: High school or less; Some college; Degree or more. ECLS-K results from balanced panel sample (weighted using 5th grade parent longitudinal weight). ECLS-B results from kindergarten sample (weighted using kindergarten parent longitudinal weight). Outcomes are age-adjusted standardized theta scores.

Table A4.4 SES gaps in math achievement across cohorts

	Age 5 income Q gaps			Parent education gaps		
Cohort/age	Q3-Q1	Q5-Q3	Q5-Q1	M-L	H-M	H-L
ECLS-K Fall K wave	0.60	0.72	1.33	0.47	0.54	1.02
	[0.48, 0.72]	[0.60, 0.84]	[1.22, 1.43]	[0.38, 0.56]	[0.45, 0.64]	[0.92, 1.12]
ECLS-B Kindergarten wave	0.51	0.73	1.24	0.37	0.56	0.93
	[0.44, 0.57]	[0.66, 0.80]	[1.17, 1.31]	[0.32, 0.43]	[0.51, 0.62]	[0.88, 0.99]
ECLS-B Preschool wave	0.34	0.51	0.85	0.40	0.55	0.95
	[0.27, 0.41]	[0.44, 0.58]	[0.78, 0.92]	[0.35, 0.46]	[0.49, 0.60]	[0.90, 1.01]

See notes to Table A4.3.

Focusing on the two sets of kindergarten data, there is some suggestion that both income and education gaps at kindergarten narrowed slightly between the ELCS-K and ECLS-B, though this difference is not statistically significant. Only the bottom-middle gap narrowed in math, while both top- and bottom-middle gaps narrowed in reading. The top-middle gap is larger than the middle-bottom gap in both cohorts for all definitions. This is more strongly marked for the more recent ECLS-B cohort than the older K Cohort (a result of the fact the gaps narrowed more between middle and bottom than between top and middle over the period).

Within the ECLS-B cohort, the income gap widened noticeably between the preschool and kindergarten waves, at both the top and bottom of the income distribution. In contrast, the education gaps remained unchanged.

Section A4.3 Socio-emotional measures

The measures used are summarized in Tables A4.5.a to A4.5.e.

Table A4.5.a Cross-national behavior coding – Instrument and coding details

US (parent report)	US (teacher report)	UK/Australia	Canada
ECLS-K Parent rating scales. Presented in 5 subscales: Impulsive/Hyperactive (IH); Self control (SC); Approaches to Learning (AL); Social interaction (SI); Sad/Lonely (SL). Plus 4 miscellaneous items not used in any scale (M)	ECLS-K Teacher rating scales. Presented in 5 subscales: Externalizing (E); Self control (SC); Approaches to Learning (AL); Interpersonal (IP); Internalizing (I)	Strengths and Difficulties Questionnaire (SDQ). Presented in 5 subscales: Conduct Problems (CP); Hyperactivity/Inattention (HI); Emotional symptoms (ES); Peer problems (PP); Prosocial behavior (PB)	NLSCY Behaviour scales. Presented in 5 sub-scales: Conduct disorder/Physical aggression (CP); Property offense (PO); Hyperactivity/inattention (HI); Emotional disorder/Anxiety (EA); Prosocial behavior (PB). (Indirect aggression scale items not used for lack of comparability with other countries).
Question wording*: I am going to read you a list of statements describing things that children sometimes do. For each statement, I want you to tell me how often {CHILD} acts in this way. How often does {CHILD}: {PROBE: Would you say never, sometimes, often, or very often?}* 1 = Never - Child never exhibits this behavior. 2 = Sometimes - Child exhibits this behavior occasionally or sometimes. 3 = Often - Child exhibits this behavior regularly but not all the time. 4 = Very often - Child exhibits this behavior most of the time.	Question wording: *For the set of items below, please think about this child's behavior during the past month or two. Decide how often the child demonstrates the behavior described. We realize that some items apply more to older children, but please answer as accurately as you can. For each item, circle one of the following responses:* 1 = Never - Child never exhibits this behavior. 2 = Sometimes - Child exhibits this behavior occasionally or sometimes. 3 = Often - Child exhibits this behavior regularly but not all the time. 4 = Very often - Child exhibits this behavior most of the time.	Question wording: *Please think about this child's behaviour over the last 6 months if you can. For each of the following statements please say whether it is:* 1 = Not true 2 = Somewhat true 3 = Certainly true of the child's behavior.	Question wording: *How often would you say that this child…* 1 = Never or not true 2 = Sometimes or somewhat true 3 = Very true DK = Don't know

US (parent report)	US (teacher report)	UK/Australia	Canada
N/O = No Opportunity - No opportunity to observe this behavior. **Note: Responses of 3 and 4 combined into a single category for comparability with other countries' 3-point scales**	N/O = No Opportunity - No opportunity to observe this behavior. **Note: Responses of 3 and 4 combined into a single category for comparability with other countries' 3-point scales**		

Table A4.5.b Cross-national behavior coding - Conduct problems scale items

US (parent report)	US (teacher report)	UK/Australia[1]	Canada
1. Fight with others (SC)	1. Fights with others (E)	1. Often has fights with other children or bullies them (CP)	1. Gets into many fights (CP)
			2. Is cruel, bullies or is mean to other children (CP)
2. Have a tantrum when he does not get his own way (SC)		2. Often has temper tantrums or hot tempers (CP)	
3. Controls temper (SC)	2. Controls temper (SC)		
4. Cooperate with family members (M)		3. Generally obedient, usually does what adults request (CP)	
5. Argue with others (SC)	3. Argues with others (E)		
6. Easily become angry (SC)	4. Gets angry easily (E)		3. When another child hurts him, he reacts with anger and fighting (CP)
	5. Disturbs ongoing activities (E)		
	6. Respects property rights of others (SC)		4. COMPOSITE ITEM[2]: Combines (1) Destroys his/her things (PO); (2) Destroys things belonging to his/her family or other children (PO); (3) Vandalizes (PO)

US (parent report)	US (teacher report)	UK/Australia[1]	Canada
		4. Steals from home, school or elsewhere (CP)	5. COMPOSITE ITEM[2]: Combines (1) Steals at home (PO); (2) Steals outside his/her home (PO)
		5. Lies and cheats (CP)	6. Tells lies and cheats (CP)
			UNUSED ITEM: Physically attacks people (CP)
			UNUSED ITEM: Kicks, bites, hits other children (CP)
			UNUSED ITEM: Threatens people (CP)

[1] Scale is the existing SDQ Conduct Problems sub-scale

[2] Coding: If ANY item response is Very, code 2. If ALL items are Not true, code 0. Otherwise code 1.

Table A4.5.c Cross-national behavior coding - Inattention scale items

US (parent report)	US (teacher report)	UK/Australia	Canada
1. Concentrate on a task and ignore distractions (AL)	1. Pays attention well (AL)	1. Easily distracted, concentration wanders (HI)	1. COMPOSITE ITEM[1]: Combines (1) Is inattentive (HI); (2) Can't concentrate, can't pay attention (HI)
2. Keep working at something until he is finished (AL)	2. Persists in completing tasks (AL)	2. Sees tasks through to the end, good attention span (HI)	2. COMPSITE ITEM[1]: Combines (1) Can't settle to anything for more than a few moments (HI); (2) Is distractible, has trouble sticking to any activity (HI)
3. Act impulsively (IH)	3. Acts impulsively (E)	3. Can stop and think things out before acting (HI)	3. Is impulsive, acts without thinking (H)
UNUSED ITEM: Hyperactive (IH)[2]		UNUSED ITEM: Restless, overactive, cannot stay still for long (HI) [2]	UNUSED ITEM: Can't sit still, is restless or hyperactive (HI) [2]

US (parent report)	US (teacher report)	UK/Australia	Canada
		UNUSED ITEM: Constantly fidgeting or squirming (HI) [2]	UNUSED ITEM: Fidgets (HI) [2]
UNUSED ITEM: Eager to learn new things (AL)	UNUSED ITEM: Shows eagerness to learn new things (AL)		
UNUSED ITEM: Help with chores (AL)			
UNUSED ITEM: Show interest in variety of things (AL)			
UNUSED ITEM: Creative in work or play (AL)			
	UNUSED ITEM: Works independently (AL)		
	UNUSED ITEM: Easily adapts to changes in routine (AL)		
	UNUSED ITEM: Keeps belongings organized (AL)		
			UNUSED ITEM: Has difficulty awaiting turn in games (HI)

[1] Coding: If ANY item response is Very, code 2. If ALL items are Not true, code 0. Otherwise code 1.

[2] Items that capture hyperactivity rather than inattention have no counterpart in the US teacher-report items, and so are excluded from the scale for other countries.

Table A4.5.d Cross-national behavior coding - Internalizing scale items

US (parent report)[1]	US (teacher report)[1]	UK/Australia[2]	Canada
1. Worry about things (M)	1. Worries about things (I)	1. Many worries, often seems worried (ES)	1. Is worried (EA)
2. Act sad (SL)	2. Acts sad or depressed (I)	2. Often unhappy, down-hearted or tearful (ES)	2. COMPOSITE ITEM[3]: Combines (1) Seems to be unhappy, sad or depressed (EA); (2) Appears miserable, unhappy, tearful or distressed (EA); (3) Cries a lot (EA)

US (parent report)[1]	US (teacher report)[1]	UK/Australia[2]	Canada
3. Show low self esteem (SL)	3. Shows low self esteem (I)	3. Nervous or clingy in new situations, easily loses confidence (ES)	
4. Appear to be lonely (SL)	4. Appears lonely (I)		3. COMPOSITE ITEM[3]: Combines (1) Has trouble enjoying himself (EA); (2) Is not as happy as other children (EA)
		4. Many fears, easily scared (ES)	4. Is too fearful or nervous (EA)
		5. Often complains of headaches, stomach aches or sickness (ES)	5. Is nervous, high strung or tense (EA)

[1] Scales are equivalent to the existing ECLSK Internalizing SRS scale with responses of Very often and Often recoded to a single category.

[2] Scale is the existing SDQ Emotional symptoms sub-scale

[3] Coding: If ANY item response is Very, code 2. If ALL items are Not true, code 0. Otherwise code 1.

Table A4.5.e Cross-national behavior coding – Prosocial behavior scale items

US (parent report)	US (teacher report)	UK/Australia[1]	Canada
1. Comfort or help others (SI)	1. Comforts or helps other children (IP)	1. Helpful if someone is hurt, upset or feeling ill (PB)	1. COMPOSITE ITEM[2]: Combines (1) Comforts a child who is crying or upset (PB); (2) Helps other children who are feeling sick(PB); (3) Will try to help someone who has been hurt (PB)
2. Listen carefully to others (M)	2. Is sensitive to the feelings of others (IP)	2. Considerate of other people's feelings (PB)	2. Shows sympathy to someone who has made a mistake (PB)
3. Have a problem being accepted and liked by others (SL)	3. Gets along with people who are different (IP)	3. Kind to younger children (PB)	3. Offers to help other children who are having difficulty with a task (PB)
4. Easily join others in play (SI)	4. Accepts peers' ideas for group activities (SC)	4. Shares readily with other children (treats, toys, pencils, etc)	4. Will invite others to join in a game (PB)

US (parent report)	US (teacher report)	UK/Australia[1]	Canada
	5. Expresses own feelings, ideas and opinions without putting down others (IP)	5. Often volunteers to help others (parents, teachers, other children)	5. Helps those who do not do as well as s/he does (PB)
UNUSED ITEM: Make and keep friends (SI)	UNUSED ITEM: Forms and maintains friendships (IP)	UNUSED ITEM[3]: Generally liked by other children (PP)	UNUSED ITEM: Spontaneously helps to pick up objects somebody has dropped (PB)
UNUSED ITEM: Respond appropriately to teasing (M)	UNUSED ITEM: Responds appropriately to peer pressure (SC)	UNUSED ITEM: Has at least one good friend (PP)	UNUSED ITEM: Volunteers to help clear a mess someone else has made? (PB)
		UNUSED ITEM: Rather solitary, tends to play alone (PP)	UNUSED ITEM: If there is a quarrel or dispute, will try to stop it(PB)
		UNUSED ITEM: Picked on or bullied by other children (PP)	
		UNUSED ITEM: Gets on better with adults than with other children (PP)	

[1] Scale is the existing SDQ Prosocial behaviours sub-scale

[2] Coding: If ANY item response is Very, code 2. If ALL items are Not true, code 0. Otherwise code 1.

[3] Unused items comprise the SDQ Peer Problems sub-scale in its entirety

Section A4.4 Obesity and health outcomes at age 5

In all four countries, height and weight are measured by the interviewer. The obesity cut-offs are based on the definitions in Cole et al (2000). In Australia these definitions are coded by the survey providers and in the other countries using the "zbmicat" function in Stata (Vidmar, Cole, & Pan, 2013).

The poor/fair health classification in the countries other than Canada (where it is not available) is based on the following question asked of the primary carer "In general, how would you say child's current health is? 1 Excellent; 2 Very good; 3 Good; 4 Fair; 5 Poor".

Section A4.5 Alternative estimates of the age 5 achievement gaps

Figure A4.1 shows the SES gaps in language/reading at age 5 when income quintiles, rather than parental education, are used as the measure of SES. Here the high SES group consists of children in the top quintile group (i.e., top fifth) of the parental income distribution; the medium SES group of children in the third (middle) quintile; and the low SES group of children in the bottom quintile group. The income measure captures gross, equivalized household incomes in constant prices, averaged over three occasions when the study child was aged 5, 7 and 11 (see Section A3.2).

Figures A4.2 and A4.3 provide additional information on SES gaps disaggregated by income and parental education for the US non-Hispanic white population. For the other countries the main results are also presented for comparison. However, though there is no simple equivalent to the US non-Hispanic white population in the countries, we would expect that gaps for the dominant ethnic groups will also be lower in these other countries.

For both sets of results for the US non-Hispanic white population, the income quintile boundaries and the variance used to standardize the outcome are those used in the full sample analysis, so that the underlying variables are identical – only the sample changes.

Figure A4.1 **Gaps in language/reading at age 5, by income**

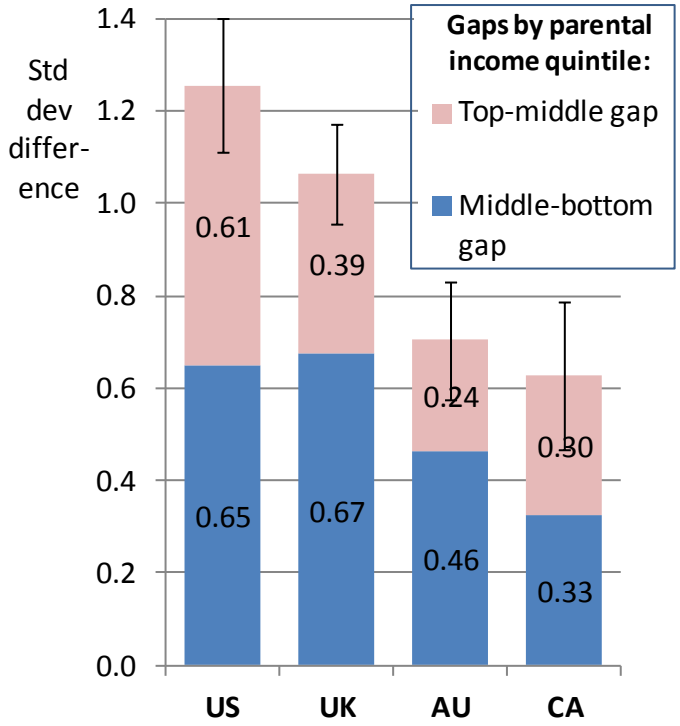


Figure A4.2 **Gaps in language/reading at age 5, by parental education, including US overall and US non-Hispanic whites only**

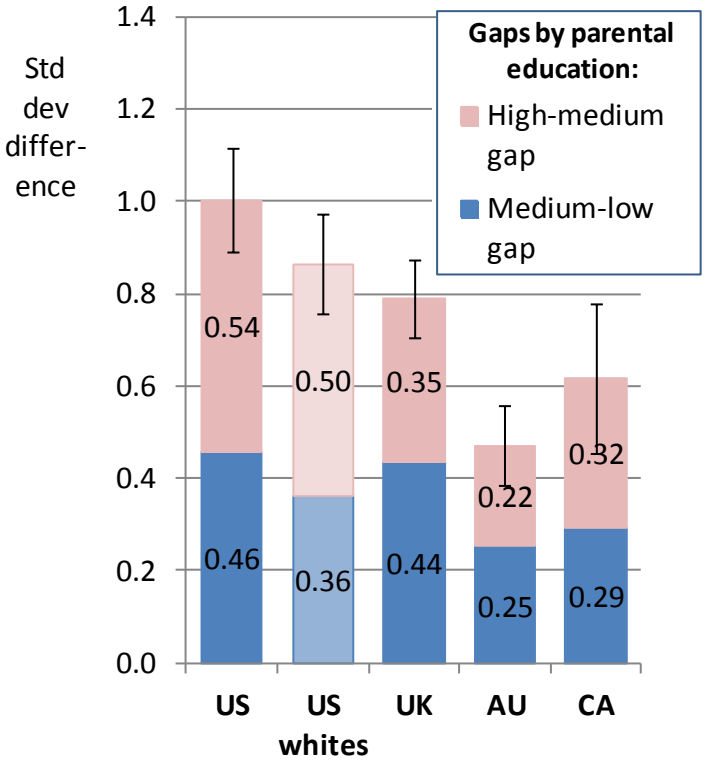
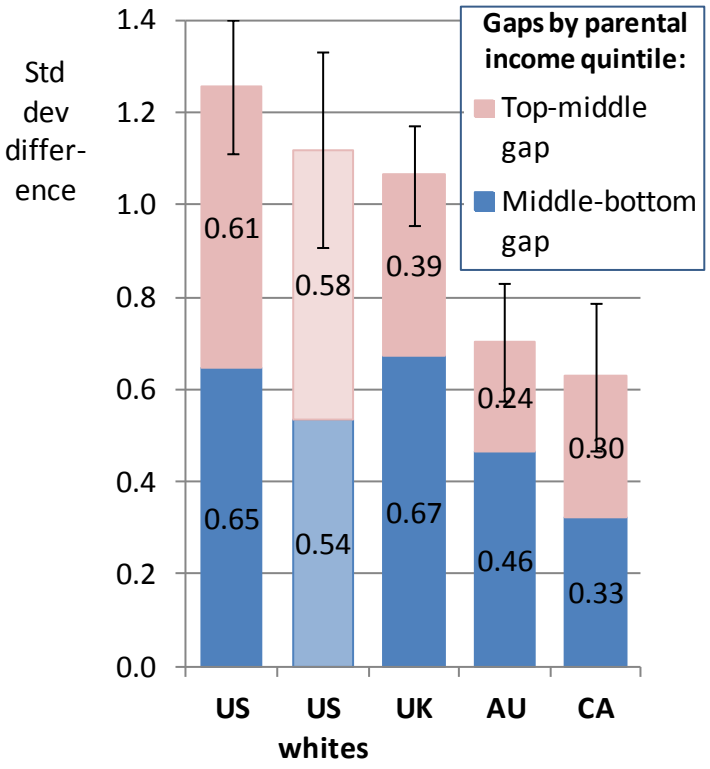


Figure A4.3 Gaps in language/reading at age 5, by income, including US overall and US non-Hispanic whites only



Section A4.6 Evidence on our four countries from international databases

A number of other studies have assessed the learning outcomes of children in multiple countries. Three studies in particular have sought to implement identical instruments in different countries to permit direct cross-national comparisons. These are the Program for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS). The PISA assesses students at around age 15, the TIMSS in 8th and 4th grade (around ages 14 and 10), and the PIRLS also in 4th grade.

Though the younger-age TIMSS and PIRLS data are not suitable for assessing SES gradients, they are able to provide information on the absolute levels and variances of achievement in our four countries. A number of rounds of these surveys have been undertaken, with the different cohorts summarised in the table below. The birth years of our US and Canadian cohorts best match those denoted as 'cohort 1' in the table, while our UK and Australian data best match those denoted as 'cohort 3'.

Table A4.6 International data sources for different cohorts

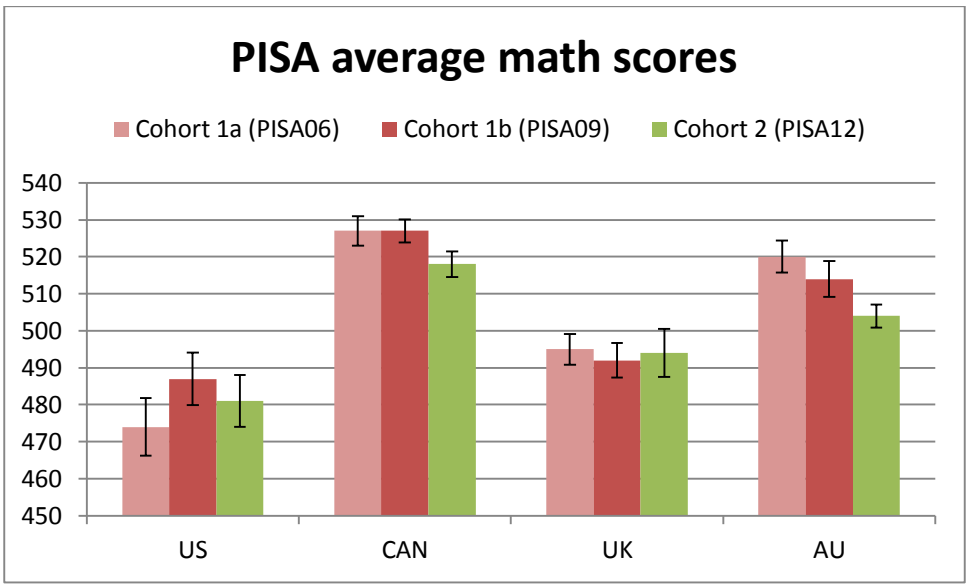
Cohort and birth years	4th grade	8th grade	Age 15
Cohort 1 (b. 1991-94)	TIMSS 2003 PIRLS 2001	TIMSS 2007	PISA 2006 (Science) PISA 2009 (Reading)
Cohort 2 (b. 1995-98)	TIMSS 2007 PIRLS 2006	TIMSS 2011	PISA 2012 (Math)
Cohort 3 (b. 1999-2002)	TIMSS 2011 PIRLS 2011		

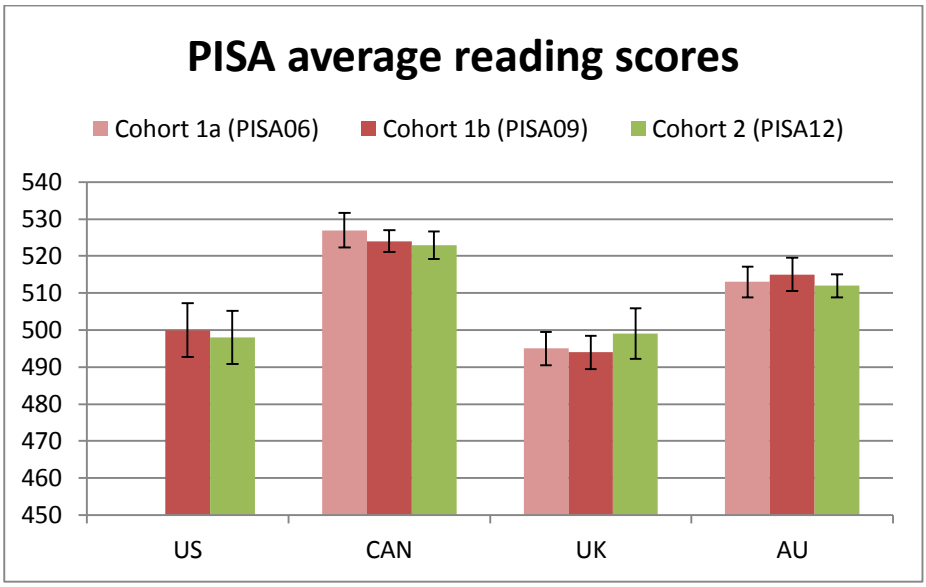
Evidence on *mean levels of achievement* in our four countries

Figure A4.4.a shows the average reading and math scores from several waves of the PISA study. Children in Canada and Australia score better on both measures than the US and UK, though these differences are relatively small in the context of the overall variance in cross-national achievement

found in the PISA study. These patterns mirror the differences in inequality that we find in our child cohorts.

Figure A4.4.a PISA (age 15)





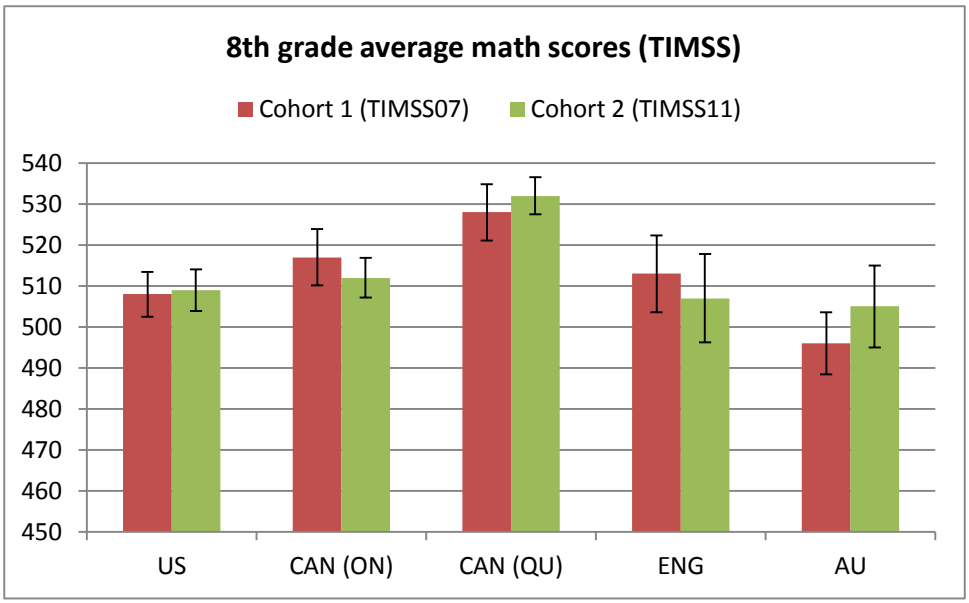
Notes: Cohorts 1a and 1b include children at the older and younger ranges of Cohort 1 (b. 1991-94) – the cohorts surveyed in our US and Canadian cohorts. The UK and Australian children are from Cohort 3, and are not yet old enough for us to observe them at this age.
PISA scores are normed so that OECD mean for a given cohort is 500 points.
US reading scores not available for Cohort 1a.
Range bars are 95% CIs.

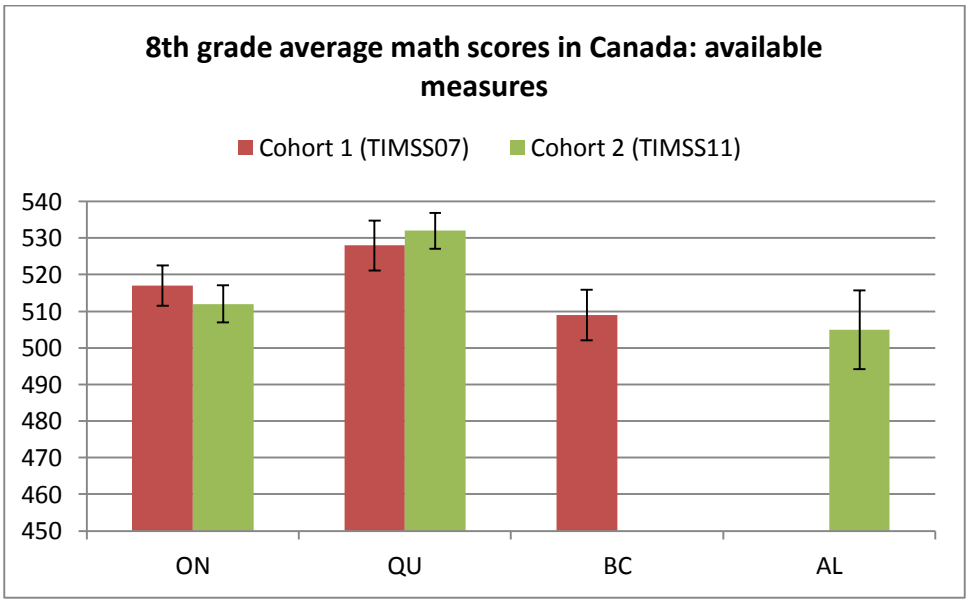
However, we do not find the same pattern in the TIMSS data (Figure A4.4.b). In theory the TIMSS results relate to the same cohorts as shown for PISA, and assess math at virtually the same age. But here the differences between the US, England and Australia are small and generally insignificant, and Australia is, if anything, the worst performer. This study doesn't have estimates for Canada as a whole. Ontario (the most populous state) doesn't stand out as different from the other three

countries, though students in Quebec have higher average scores. The bottom panel of the figure presents evidence for other provinces which suggests that Quebec is an outlier among Canadian provinces.

It is not obvious why the TIMSS presents different cross-national patterns to the PISA. The orientation of the tests do differ, with the TIMSS tending to measure math "knowledge", closer to the content of the school curriculum, while PISA is more about ability to apply that knowledge. This suggests, however, that one should be cautious about making strong statements about cross-national difference in ability *levels* on the basis of any single international survey.

Figure A4.4.b TIMSS 8th grade (~age 14)



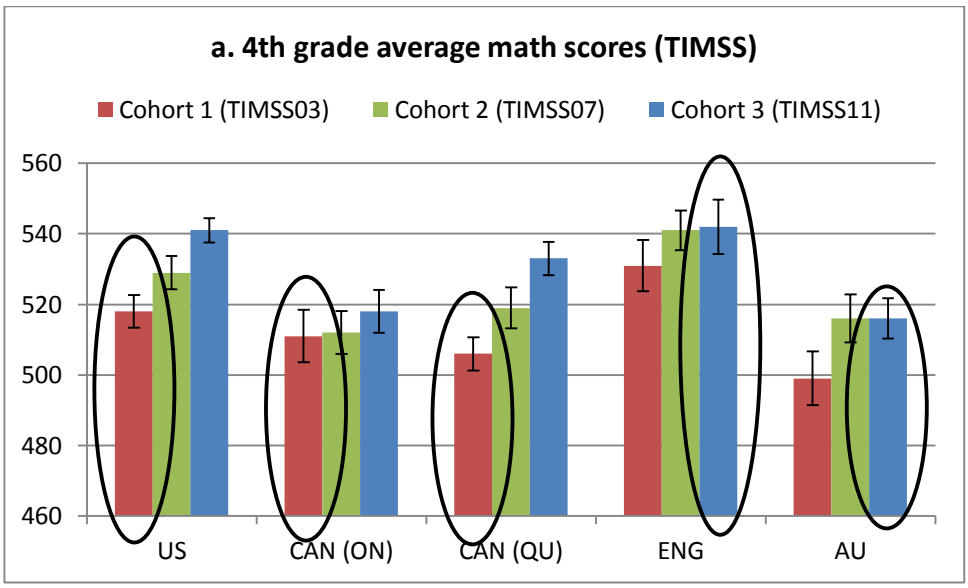


Notes. TIMSS scores are normed so that the mean was 500 *in 1995*, therefore the numbers don't have the same meaning as in PISA. Numbers are only available for selected Canadian provinces. The top figure plots the numbers for Ontario and Quebec (the two largest provinces); the bottom figure plots the numbers for all four available provinces (Ontario, Quebec, British Columbia, and Alberta). PIRLS reading scores are not available in 8th grade.

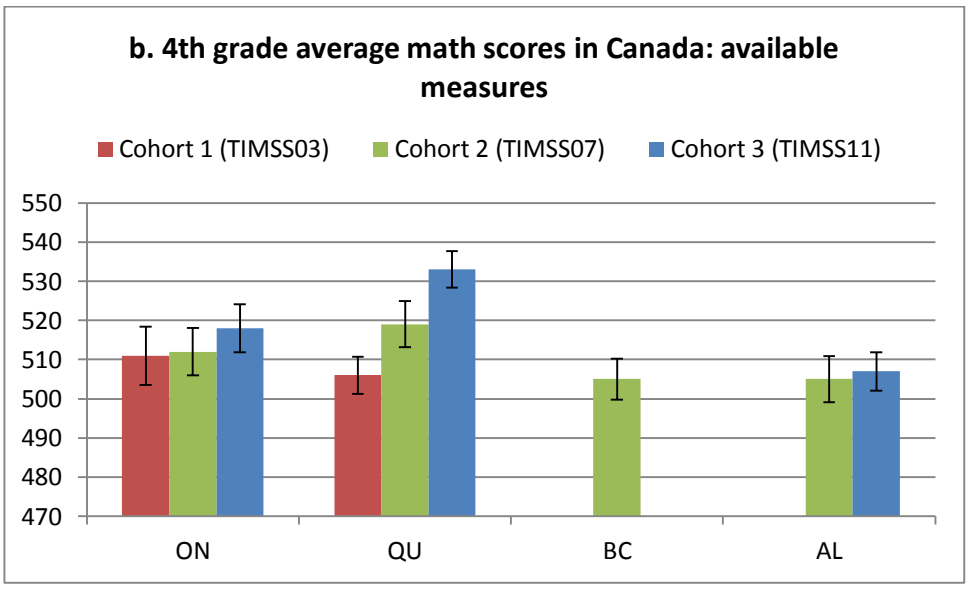
Canadian population by province: Ontario (38%); Quebec (24%); British Columbia (13%); Alberta (11%).

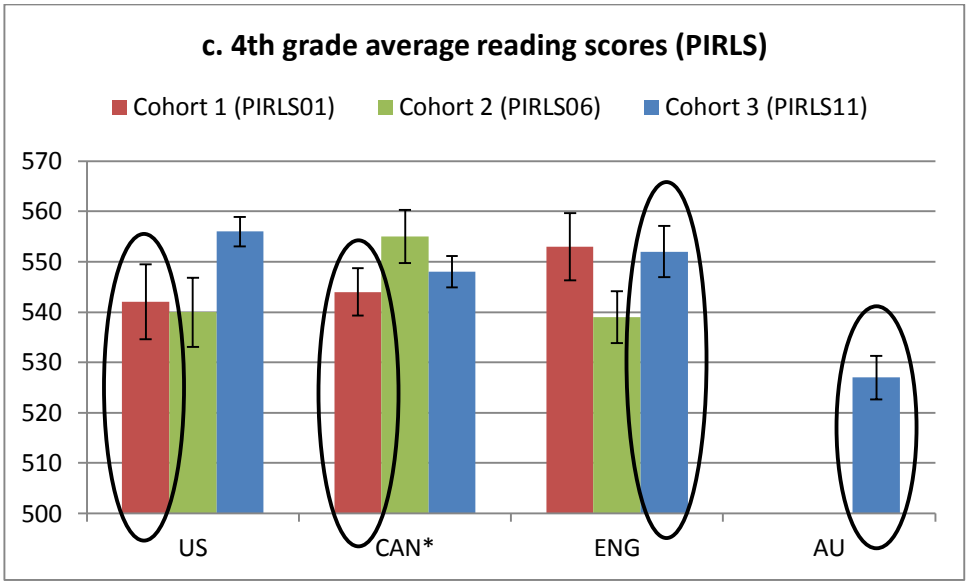
Further evidence on average 4th grade scores in our countries is presented below in Figures A4.5.a to A4.5.e. In general, average 4th grade scores are highly similar across the four countries. The only exceptions are that England has a higher average math score than the rest, and Australia has a lower average reading score than the rest. This pattern does depend somewhat on the fact that the timing of our cohorts are different. Scores have tended to improve for more recent cohorts in the US and Canada, so if we were able to use data on Cohort 3 for all our countries the US and Canada would look relatively better than they do here.

Figure A4.5 TIMSS/PIRLS 4th grade (~age 10)

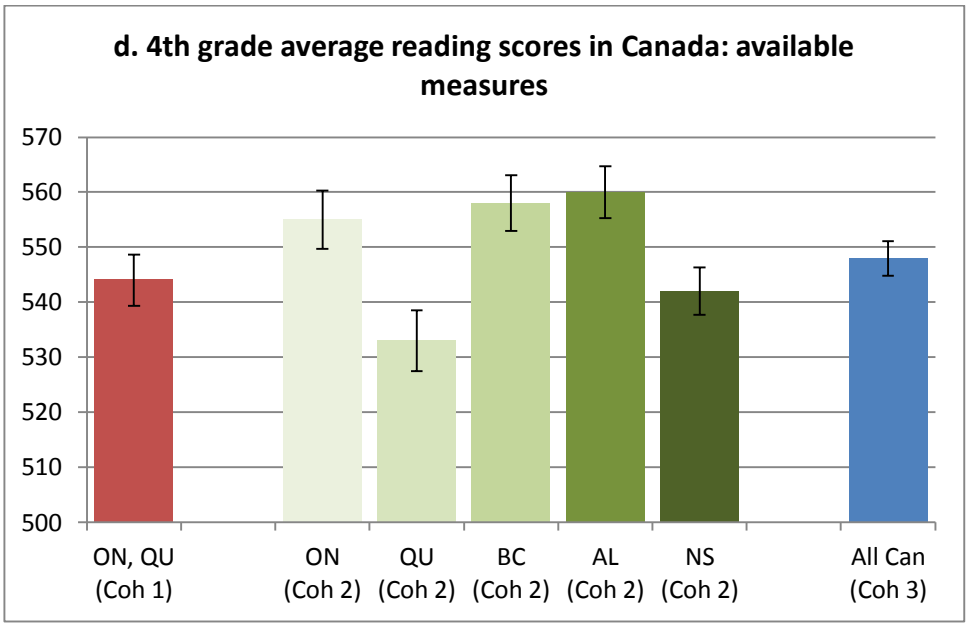


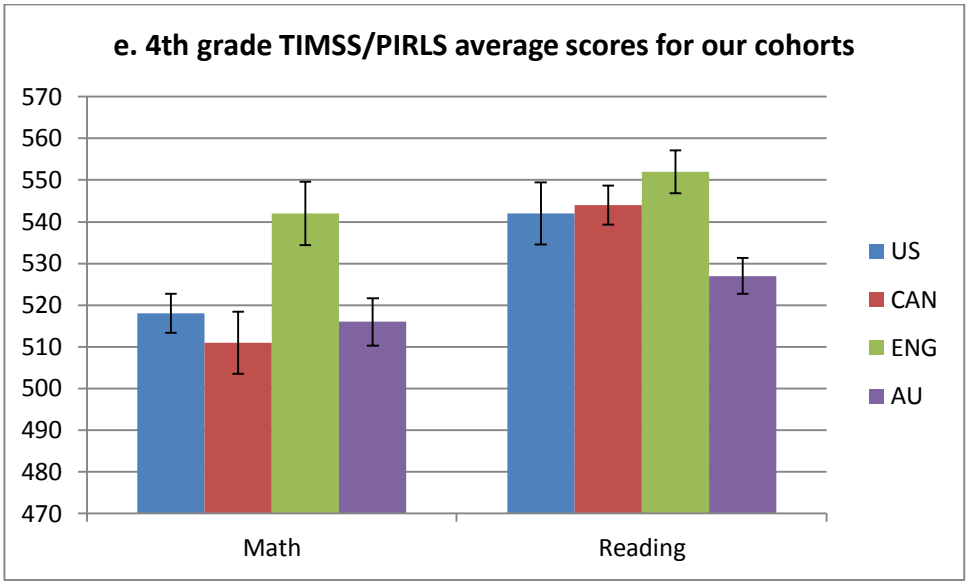
Circled bars relate to the results from the cohorts in our data (US/Can = Coh 1; AU/ENG = Coh 3).





Circled bars relate to the results from the cohorts in our data (US/Can = Coh 1; AU/ENG = Coh 3). The provinces represented by the Canadian numbers are different in each cohort: Cohort 1 = ON & QU combined; Cohort 2 = Ontario; Cohort 3 = All Canada. See Figure 8.





Numbers for Canada shown are for Ontario (Math) and Ontario and Quebec combined (Reading). US and Canadian numbers are for Cohort 1: TIMSS2003 and PIRLS2001; England and Australia numbers are for Cohort 3: TIMSS2011 and PIRLS2001.

Evidence on the *variance of achievement* in our four countries

Our key outcome measures are reported using standardized dependent variables, with the differences between SES groups reported in standard deviation units. This standardization allows us to compare across nations the amount of the variance in child outcomes that is associated with SES – a measure of the extent to which family background determines child outcomes. However, a given z-score gap between rich and poor children (e.g., one standard deviation) could be due to either a relatively large gap in average reading skills between the two groups combined with a large spread of scores within each group OR with a relatively small gap in average reading skills combined with a small spread of scores within each group. (This follows from the fact that the z-score gap is the gap relative to the overall standard deviation and that most of the overall variation comes from within-group variation).

For some purposes we might be more interested in the difference in average outcomes between groups – measured in some absolute way that is comparable across countries. Putting this another way, a difference of one standard deviation might mean a greater difference in terms of objective learning or skills in one country than in another. We cannot assess this directly with our source measures (since they have arbitrary scales) but we can look at the variances in these other cross-nationally comparable data sources to ascertain whether they find a greater real difference in outcomes corresponding to the same z-score difference.

We examine this issue using the 4th grade (age 10) PIRLS and TIMSS results since they are for children closest to our age range. In these studies the scores for all countries are scaled together in order to generate comparable scores and the scales are also linked across years (Foy et al, 2012). Since the main focus of this linking is to match mean outcomes across time and space, we should exercise caution in assuming that the linkage also matches variances appropriately (especially across years).

Apart from the methodological, age and cohort differences, these studies measure different skill sets than in our data, and so these patterns should only be interpreted as providing a general indication of the relative variances of our outcomes in different countries.

Table A4.7 Standard deviations at 4th grade in waves corresponding to our cohorts

Country	Survey	SD	95% LB	95% UB	Relative to US
Reading					
US	PIRLS01	83	[79.1,	86.9]	1
ENG	PIRLS11	82	[79.3,	84.7]	1.01
AU	PIRLS11	80	[77.5,	82.5]	1.04
ON & QU	PIRLS01	72	[70.0,	74.0]	1.15
Math					
US	TIMSS03	76	[74.0,	78.0]	1
ENG	TIMSS11	89	[85.7,	92.3]	0.85
AU	TIMSS11	86	[82.1,	89.9]	0.88
ON	TIMSS03	71	[67.1,	74.9]	1.07
QU	TIMSS03	65	[62.6,	67.4]	1.17

Source: Foy et al (2012)

Table A4.7 shows the standard deviations for the PIRLS and TIMSS waves most closely associated with our cohorts. Where the country standard deviation is higher, this means that a gap of one standard deviation implies a higher average gap in terms of a comparable learning outcome.

In reading, the Australian and English results are not significantly different from the US, while the standard deviation of scores in Ontario and in Quebec is lower. So, at face value, these results suggest that our general results of greater inequality in the US and least in Canada would not change (or would be reinforced in the case of Canada) if we were to put these scores in an absolute outcomes framework.

However, for math, Australian and England do have higher standard deviations than the US. Quebec has less variation than the other countries on both measures, but this is likely to be an underestimate of overall Canadian variance, as there is substantial inter-province variation (see Figure A4.5.b above).[3] PIRLS2011 gives a standard deviation for all Canada, though this is a more recent cohort than the one we examine. The US SD in that survey is 73 (smaller than the 83 for our cohort) and the all Canada SD is 69 – still implying less variation in Canada, but the gap is now smaller.

As an alternative approach (albeit one less closely linked to our data), Table A4.8 shows the standard deviations for our four countries from every PIRLS and TIMSS survey (where available). England and Australia consistently have larger variances than the US (from 5 to 15 per cent greater). Variances for the Canadian provinces tend to be slightly lower than for the US, implying the all Canada variance is probably very similar to that of the US.

[3] Similarly, we might expect the standard deviation for the UK to be slightly greater than for England.

If we were to assume that these within-cohort differences applied to our data (even though it is from different years) this would imply that the UK and AU gaps should be inflated by about 5 to 15 per cent. This would still leave both gaps smaller than the US gaps (e.g., in Figure 4.1 where the gap is over 20 per cent) but the UK gap might now not be significantly smaller than the US.

Note however, that, our conclusion above based on using the TIMSS and PIRLS surveys which do match our cohorts suggests that essentially no change should be made to the US vs UK/AU relativities to interpret them as comparable variations in absolute outcomes.

Table A4.8 Standard deviations in 4th grade in PIRLS and TIMSS

	SD	95% LB	95% UB	Ratio to US
Reading				
Cohort 1 (PIRLS01)				
US	83	79.1	86.9	1
ENG	87	83.7	90.3	0.95
AU				
ON/QU	72	70.0	74.0	1.15
Cohort 2 (PIRLS06)				
US	74	71.5	76.5	1
ENG	87	83.9	90.1	0.85
AU				
AL	68	65.6	70.4	1.09
BC	69	66.5	71.5	1.07
ON	71	68.5	73.5	1.04
QU	63	60.8	65.2	1.17
NS	76	73.1	78.9	0.97
Cohort 3 (PIRLS11)				
US	73	71.0	75.0	1
ENG	82	79.3	84.7	0.89
AU	80	77.5	82.5	0.91
All Can	69	67.2	70.8	1.06
Math				
Cohort 1 (TIMSS03)				
US	76	74.0	78.0	1
ENG	87	83.3	90.7	0.87
AU	81	76.9	85.1	0.94
ON	71	67.1	74.9	1.07
QU	65	62.6	67.4	1.17
Cohort 2 (TIMSS07)				
US	75	72.6	77.4	1.00
ENG	86	82.9	89.1	0.87
AU	83	79.1	86.9	0.90
AL	66	62.5	69.5	1.14
BC	71	68.1	73.9	1.06
ON	68	64.5	71.5	1.10
QU	67	64.8	69.2	1.12
Cohort 3 (TIMSS11)				
US	76	73.8	78.2	1.00
ENG	89	85.7	92.3	0.85
AU	86	82.1	89.9	0.88
AL	65	62.1	67.9	1.17
ON	73	70.1	75.9	1.04
QU	60	58.0	62.0	1.27

Section A4.7 SES gaps in child inputs

Table A4.9 SES gaps in child health at age 5, by parental education

	US	UK	AU	CA
Poor/fair health (%)				
All	3	4	3	2
HS or less	5	6	3	3
Some post-secondary	3	4	1	1
College degree or more 1	2	3	1	
Obesity (%)				
All	7	5	6	-
HS or less	9	7	7	-
Some post-secondary	8	5	5	-
College degree or more 5	3	5	-	

Note: Obesity is defined by applying the cut-offs recommended by the Childhood Obesity Working Group of the International Obesity Taskforce (Cole et al. (2000)) to measured body mass index.

Table A4.10 Employment patterns of mothers of 5 year olds

	US	UK	AU	CA
Not working (%)				
All	32	44	44	27
HS or less	37	61	52	42
Some post-secondary	28	36	41	25
College degree or more	29	28	36	20
Working 1-29 hours/week (%)				
All	15	36	36	24
HS or less	10	29	32	19
Some post-secondary	13	41	36	24
College degree or more	22	43	40	29
Working 30+ hours/week (%)				
All	53	19	20	49
HS or less	53	11	16	39
Working 30+ hours (%)	59	23	24	51
College degree or more	49	29	24	51

Table A4.11 Child care/early education arrangements for children in year prior to starting school

	US	UK	AU	CN
Preschool/center care (%)				
All	*69*	*93*	*94*	*57*
High school diploma or less	60	92	93	43
Some post-secondary	70	94	95	59
College degree or more	78	92	96	63
Other/informal care (%)				
All	*5*	*0*	*1*	*11*
High school diploma or less	5	0	1	10
Some post-secondary	6	0	0	10
College degree or more	5	1	1	15
Parental/relative care (%)				
All	*26*	*7*	*5*	*32*
High school diploma or less	35	8	7	47
Some post-secondary	24	6	4	31
College degree or more	17	7	3	23

Table A4.12 Educationally oriented items and activities for 5 year olds

	US	UK	AU	CA
More than 30 children's books in home (%)				
All	75	-	82	-
HS or less	55	-	76	-
Some post-secondary	83	-	84	-
College degree or more	93	-	88	-
More than 25 books (any kind) in home (%)				
All	-	-	-	91
HS or less	-	-	-	83
Some post-secondary	-	-	-	91
College degree or more	-	-	-	99
Number of children's books in home				
All	73	-	-	-
HS or less	48	-	-	-
Some post-secondary	75	-	-	-
College degree or more	103	-	-	-
Music lessons (%)				
All	8	-	5	-
HS or less	3	-	2	-
Some post-secondary	7	-	4	-
College degree or more	13	-	8	-
Music, art or any non-sport activity outside school hours (once a week or more; %)				
All	-	-	-	9
HS or less	-	-	-	3
Some post-secondary	-	-	-	8
College degree or more	-	-	-	15
Art classes/lessons (%)				
All	7	-	-	-
HS or less	4	-	-	-
Some post-secondary	6	-	-	-
College degree or more	13	-	-	-
Performing arts program (%)				
All	15	-	-	-
HS or less	9	-	-	-
Some post-secondary	16	-	-	-
College degree or more	22	-	-	-
Organized sports/gymnastics/dance (%)				
All	52	52	23	45
HS or less	30	34	18	26
Some post-secondary	52	58	23	43
College degree or more	78	76	31	65
Organized athletics activity (%)				

All	45	-	10	33
HS or less	26	-	7	17
Some post-secondary	45	-	11	31
College degree or more	69	-	12	50
Dance lessons (%)				
All	17	-	14	25
HS or less	8	-	11	12
Some post-secondary	16	-	13	24
College degree or more	29	-	20	38
(CN includes gymnastics/martial arts)				
Organized club or recreation program (%)				
All	13	-	-	16
HS or less	8	-	-	14
Some post-secondary	15	-	-	16
College degree or more	18	-	-	19

Note: Dashed line indicates data not available.

Table A4.13 Parent use of spanking at age 5

	US	UK	AU	CA
Children spanked in past week (%)				
All	27	-	-	-
High school or less	31	-	-	-
Some post-secondary	29	-	-	-
College degree or more	21	-	-	-
Child spanked once a month or more (%)				
All	-	12	-	-
High school or less	-	14	-	-
Some post-secondary	-	12	-	-
College degree or more	-	10	-	-
Parent uses physical punishment Sometimes/often/always (%)				
All	-	-	-	10
High school or less	-	-	-	14
Some post-secondary	-	-	-	10
College degree or more	-	-	-	7

Note: Dashed line indicates data not available.

Section A4.8. Input measure definitions

Table A4.14 below provides detailed information on the raw variables and derivation of the measures shown in the tables above.

Table A4.14 Parental inputs at age 5

	US	UK	Australia	Canada
Books in home	About how many children's books does {CHILD} have in your home now, including library books? Please only include books that are for children.	-	About how many books does child have in your home now, including any library books? (Include books owned by older brothers or sisters if age-appropriate for the study child) 0 None; 1 1-10; 2 11-20; 3 21-30; 4 More than 30	Which of the following materials do you currently have in your home? (Include electronic versions of materials listed below. Read categories to respondent. Mark all that apply.) 01 Daily newspapers 02 Magazines/weekly newspapers **03 More than 25 books** 04 A (multi-volume) encyclopedia 05 A dictionary
Activities question format	Outside of school hours in the past year, has {*child*} participated in ACTIVITY:	On average how many days a week does {*child*} ACTIVITY ...? 1 Five or more days a week 2 Four days a week 3 Three days a week 4 Two days a week 5 One day a week 6 Less often or not at all Coded: One day a week or more = 1; Less often or not at all = 0	In the last 6 months, has child regularly attended any special or extra cost activities that are not part of his/her normal child care, pre-school or school activities? (Regular means weekly or fortnightly activities, even if they lasted less than 6 months.)	In the past 12 months, outside of school hours, how often has {*child*}: ACTIVITY (If child is only active in one season, the respondent should give the frequency for that season, not try to average over the year. Read categories to respondent.) 1 Most days 2 A few times a week 3 About once a week 4 About once a month 5 Almost never

				Coded: About once a week or more = 1; About once a month or less = 0
Music lessons	Music lessons, for example, piano, instrumental music or singing lessons?	-	Musical instruments or singing	-
Music, art or any non-sport activity outside school hours	-	-	-	...taken lessons or instruction in music, art or other non-sport activities?
Art classes/lessons	Art classes or lessons, for example, painting, drawing, sculpturing?	-	-	-
Performing arts program	Organized performing arts programs?	-	-	-
Organized sports/gymnastics/dance	COMPOSITE VARIABLE: Coded 1 if Organized athletics activity OR Dance lessons = 1; coded 0 otherwise	... go to a club or class to do sport or any other physical activity like swimming, gymnastics, football, dancing etc?	COMPOSITE VARIABLE: Coded 1 if Organized athletics activity OR Dance lessons = 1; coded 0 otherwise	COMPOSITE VARIABLE: Coded 1 if Organized athletics activity OR Dance lessons = 1; coded 0 otherwise
Organized athletics activity	Organized athletic activities, like basketball, soccer, baseball, or gymnastics?	-	Team sport (athletics, football etc.)	...taken part in sports with a coach or instructor (except dance, gymnastics or martial arts)?
Dance lessons	Dance lessons?	-	Ballet or other dance	...taken lessons or instruction in other organized physical activities with a coach or instructor such as dance, gymnastics or martial arts?
Organized club or recreation program	Organized clubs or recreational programs, like scouts?	-	-	...taken part in any clubs, groups or community programs with leadership,

				such as Brownies, Cubs or church groups?
Spanking	About how many times, if any, have you spanked *{child}* in the past week? (open-ended question) Coded: Spanked one or more times = 1; Never spanked = 0	How often do you do the following when *{child}* is naughty: Smack *{him/her}* 1 Never 2 Rarely 3 Sometimes (about once a month) 4 Often (about once a week or more) 5 Daily 6 Cant say		Please tell me how often you, as *{his/her}* parent, do each of the following when *{child}* breaks the rules or does things that he / she is not supposed to: use physical punishment? (Read categories to respondent.) 1 Never 2 Rarely 3 Sometimes 4 Often 5 Always

Child care arrangements

United States. To measure child care arrangements the year before kindergarten, we used a set of questions that were assessed at the fall kindergarten survey about type of child care and number of hours per week spent in each type of child care. The original questions we used are as follows:

- Did {CHILD} attend Head Start the year before {he/she} started kindergarten?
- How many hours each week did {CHILD} go to the Head Start program?
- Did {CHILD} attend a day care center, nursery school, preschool or prekindergarten program on a regular basis the year before {he/she} started kindergarten?
- How many hours each week did {CHILD} go to that program?
- Did {CHILD} receive care from a relative on a regular basis the year before {he/she} started kindergarten?
- How many hours each week did {CHILD} receive care from {his/her} relative the year before {he/she} started kindergarten?
- Did {CHILD} receive care from a nonrelative on a regular basis the year before {he/she} started kindergarten?
- How many hours each week did {CHILD} receive care from a nonrelative the year before {he/she} started kindergarten?

Using these questions, we created an indicator for child care arrangements right before kindergarten entry by defining 3 groups in the following order: (1) a group of children who attended Head Start or other center-based care; (2) a group of children who did NOT receive any center-based care, but did receive care from nonrelatives at least 8 hours or more per week; (3) all remaining children those who received care from relatives only, or from non-relatives for less than 8 hours per week.

United Kingdom. The primary question from the MCS age 5 survey was: "(Including anything you've already told me about) Has [^Cohort child's name] ever been to any of the early education or childcare providers on this card?: 1 Nursery School/Nursery Class; 2 Playgroup; 3 Pre-school; 4 Childminder; 5 Day Nursery (including workplace/college creche); 6 None of these."

Note that because 99% of the MCS children were enrolled in compulsory schooling at the age 5 wave, this question relates to retrospective child care arrangements. If a child had attended a setting from categories 1 to 3, or 5, this was coded as centre-based care. If they had attended a childminder only, this was coded "other/informal care". Remaining children were coded to relative/parental care only.

Australia. Derived from the wave 1 question "Which [of these] does child go to for the most hours each week? (Choose program child is in for the most hours each week. If two of equal hours, choose the one child has been in longest) ". We define child care status only for those children who were not in school at the time of interview (around 83% of the sample). Categories "3 Pre-school program in a school; 4 Pre-school program at a non-school centre; 5 Mobile pre-school; 6 Day care centre where child has a pre-school program;" are categorised as formal child care. The categories "7 Day care centre where child does not have a pre-school program; 8 Day care centre, not sure about a pre-school program" are coded as non-formal child care, as is care by nannies, friends, neighbours, or relatives.

Canada. Child care arrangements were only coded for the 55% of the sample who were not in formal kindergarten at the time of the age 5 survey. Among the 45% remaining, children were classed as attending a center-based provider if they attended junior kindergarten, nursery school/pre-school or a daycare center (including at the workplace). Children who did not attend any of these, but were cared for by a non-relative in someone's home (either the child's or another's home), were classed as receiving other/informal care. If the child experienced neither of these two sorts of arrangements, they were placed in the parental/relative care only category.

See Section A3.4 for details of parental employment variables.

Appendix to Chapter 5

Section A5.1 Between- and within-school variation in achievement outcomes across countries

Zopluoglu (2012) reports the intra-class correlation coefficient (ICC) for reading, math and science skills in the PIRLS and TIMSS studies for 4th and 8th grade students. The ICC is a measure of the proportion of variance which is between rather than within schools. The results underlying the discussion in Chapter 5 are shown below.

Table A5.1 ICC coefficients

4th grade

	AU	CN	ENG	US
Reading 2001			0.18	0.27
Reading 2006			0.20	0.23
Math 1995	0.26	0.21	0.23	0.25
Math 2003	0.29		0.24	0.33
Math 2007	0.30		0.16	0.29
Science 1995	0.22	0.17	0.19	0.30
Science 2003	0.25		0.20	0.37
Science 2007	0.28		0.17	0.31

8th grade

	AU	CN	ENG	US
Math 1995	0.37	0.25	0.32	0.41
Math 1999	0.55	0.25	0.45	0.34
Math 2003	0.48		0.51	0.42
Math 2007	0.54		0.55	0.32
Science 1995	0.32	0.26	0.31	0.40
Science 1999	0.40	0.18	0.39	0.35
Science 2003	0.37		0.42	0.43
Science 2007	0.48		0.47	0.31

Source: Zopluoglu (2012)

Section A5.2 Achievement measures

United States

The achievement tests for reading and math skills were administered in all the selected study waves. Children were assessed at school, using CAPIs. Assessments used a set of "routing" items, so that children's responses determined whether they went on to subsequently receive more or less difficult items, and thus the same items were not given to all children. For this reason, the ECLS-K provided Item Response Theory (IRT) scores which reflect the predicted number and difficulty of items a child would answer correctly if the child was administered all questions. To construct IRT

scale scores, the IRT model first estimates individual ability on a test (known as theta) by combining characteristics of the items of the test with a child's pattern of responses. These reading and math theta scores are used as our achievement outcome measures since they have a clear advantage, which is that theta scores are on an absolute scale and thus their distribution is more symmetrical than that of IRT scores (LoGerfo, Nichols, & Reardon, 2006). Reading and math theta scores are not provided in the ECLS-K restricted-use data, but downloadable from the NCES website.

Reading. The reading assessment was developed by the ECLS-K, based mainly on the Reading Framework for the 1992 and 1994 National Assessment of Educational Progress (NAEP) fourth-grade test specifications (National Assessment Governing Board [NAGB], 1994), which define four types of reading comprehension skills—initial understanding; developing interpretation; personal reflection and response; and demonstrating a critical stance. The ECLS-K further added two additional types of reading skills to the NAEP framework, basic skills and vocabulary, since the framework starts with fourth grade and thus it needs to be modified to consider reading skills in the earliest grades (Rock & Pollack, 2002). Therefore, the ECLS-K reading test assessed six types of basic and comprehensive reading skills (Pollack, Atkins-Burnett, Najarian, & Rock, 2005; Rock & Pollack, 2002): 1) basic skills (print familiarity, letter recognition, beginning sounds, ending sounds, short vowels, long vowels, and rhyming words); 2) vocabulary (picture-spoken word matching and word recognition); 3) Initial understanding (providing an initial impression or global understanding, identifying the main point of a passage, and identifying the specific points); 4) developing interpretation (extending initial impressions, linking information across parts of the text, and focusing on specific information); 5) personal reflection and response (connecting knowledge from the text with children's own personal background knowledge); and 6) demonstrating a critical stance (asking about the adequacy of evidence, and asking kindergarten or first grade children about unrealistic stories to assess their notion of "real vs. imaginary" and third- and fifth-grade children about understanding of literary devices or the author's intention). The internal consistency across these items was 0.93 in fall kindergarten, 0.96 in spring first grade, 0.95 in spring third grade, and 0.93 in spring fifth grade (Pollack et al., 2005).

In the fall kindergarten and spring first-grade surveys, children who spoke a language other than English received the language screener (i.e., the OLDS); and children who did not pass the language screener and spoke Spanish received both Spanish-translated math assessments and OLDS, but did not receive English-version reading assessments (Tourangeau et al., 2000). We impute missing reading scores for children who did not pass the language screener by employing multiple imputation within a sample of children who received the language screener, separately at the fall kindergarten and spring first-grade waves. We create ten imputed datasets by using the ICE command in Stata (Royston, 2005) and a set of selected variables (i.e., reading and math theta scores, OLDS scores, household income, parental education, family type, mother's age at birth, number of siblings, child's gender, child's age, child's ethnicity, and child's low birth weight status), and then assign average predicted reading scores across the ten imputed datasets to the children who did not pass the language screener (about 1,170 and 330 children at the fall kindergarten and spring first-grade surveys, respectively).

Mathematics. The math assessment was developed by the ECLS-K, based on the Mathematics Framework for the 1996 NAEP (NAGB, 1996). To measure mathematical skills, the assessment was composed of the following content (Pollack et al., 2005; Rock & Pollack, 2002): 1) number sense,

properties, and operations (understanding of numbers, operations, and estimation, and application to real-world situations; understanding of numerical relationships as expressed in ratios, proportions, and percentages; understanding properties of numbers and operations; ability to generalize from numerical patterns; and verifying results); 2) measurement (choosing a measurement unit, comparing the unit to the measurement object, and reporting the results of a measurement task); 3) geometry and spatial sense (simple identification of geometric shapes and transformations and combinations of those shapes); 4) data analysis, statistics, and probability (skills of collecting, organizing, reading, and representing data); and 5) patterns, algebra, and functions (the ability to recognize, create, explain, generalize, and extend patterns and sequences; and the techniques of identifying solutions to equations with one or more missing pieces or variables). The internal consistency across these items was 0.92 in fall kindergarten, 0.94 in spring first grade, 0.95 in spring third grade, and 0.94 in spring fifth grade (Pollack et al., 2005).

United Kingdom

Language/reading. Three different scales were used at ages 5, 7 and 11, all taken from the British Ability Scales (BAS), a battery of individually administered tests of cognitive abilities and educational achievements.

The age 5 measure was the BAS Naming Vocabulary scale. This is a verbal scale for children aged 2 years 6 months to 7 years 11 months. It assesses the spoken vocabulary of young children. The test items consist of a booklet of coloured pictures of objects which the child is shown one at a time and asked to name. The scale measures expressive language ability, and successful performance depends on the child's previous development of a vocabulary of nouns. Picture recognition is also crucial; however, the pictures are large and brightly coloured and are unlikely to cause problems except for children with major visual impairments or with no experience of picture books. The items require the child to recall words from long-term memory rather than to recognise or understand the meaning of words or sentences.

The age 7 measure was the BAS Word Reading scale from the British Ability Scales: Second Edition (BAS 2) which assesses children's English reading ability. The child reads aloud a series of words presented on a card. The assessment consists of 90 words in total. The words are organised into 9 blocks of 10 words in ascending order of difficulty. The child is asked to read each word in a block out loud to the interviewer. The number of blocks of words the child is asked to attempt to read is dependent on the child's performance during the assessment. This assessment is designed to be used with children aged from 5 years to 17 years and 11 months. All of the children in MCS4 started at the first item, as this was the starting point for children of their age. A child's progression through the assessment is dependent on the number of words they read correctly. If a child makes 8 errors in a block of 10 words, then the assessment stops.

The age 11 measure was the Verbal Similarities scale from the British Ability Scales: Second Edition (BAS 2) which assesses children's verbal reasoning and verbal knowledge. The interviewer reads out three words to the child who must then say how the three things are similar or go together. This assessment is designed to be used with children aged from 5 years to 17 years and 11 months. All of the children in MCS5 start at the 16th item, as this is the starting point for children of their age. There are decision points after items 28 and 33 where the child's performance so far decides whether the test stops or continues to the next set of questions. The test stops at the decision point

unless the child has less than three failures on all items so far. In this case they are routed to the next set of questions. If the child has obtained less than three passes however, they are routed back to the previous starting point (e.g. item 8). After five consecutive failures the test is automatically stopped provided that at least three items have been passed prior to this, otherwise they are routed back to the previous starting point. If the child fails either of the first two items administered they are provided with teaching to help them to understand the concept of the test. If the child subsequently gives a correct answer to the same question it is acknowledged but they do not receive a point for that question.

Math. This test was adapted from the NFER Progress in Maths test which is aimed for 7-year-olds and was originally developed and nationally UK standardised in 2004. The whole test has a maximum raw score of 28. The national mean raw score in 2004 was 19.3 with a standard deviation of 5.3. The scores were nationally age standardised to a mean of 100 and SD of 15. The edition of this test used in the MCS is an adaptive version of the test created by Cres Fernandes of NFER. All children have to complete an initial test and based on their score they are routed to easier, medium or harder sections. The sections were devised to save administration time, as it means each child completes around half the original number of questions. An item response scaling method (Rasch) was used to scale the results of the easy, medium and hard subtest scores to the equivalent original raw scores.

Australia.

The age 5 and 9 cognitive assessments are shortened versions of the Peabody Picture Vocabulary Test - Third Edition (PPVT-III). The original PPVT-III test is documented in Dunn, Dunn & Dunn (1997) and the methods used to develop the shortened version in Rothman (2005). The shorter version uses 40 of the full 204 items. In exploratory analysis, the correlation between the full PPVT-III and the shortened version was 0.93 (for the same dataset used to develop the shortened version). This adaptation is based on work done in the United States for the Head Start Impact Study, with a number of changes for use in Australia.

Items consist of a stimulus word given orally by the examiner and four numbered picture plates, each with a simple black and white illustration. The child is asked to indicate which picture best represents the meaning of the stimulus word. An initial set of 20 items is administered to children, then one of two different sets of 10 items administered depending upon how many of the initial set the child answers correctly. Scores are combined into a single index using Rasch scoring methods.

The Age 11 assessment is the reading score from the NAPLAN test administered to year 5 children in Australian schools. The National Assessment Program – Literacy and Numeracy (NAPLAN) is documented at http://www.nap.edu.au/naplan/naplan.html.

The process for linking this data to the LSAC survey is documented in Daraganova et al (2013) – though it should be noted that linkage consent from additional families was been obtained after this document was written (and this additional data is used here). Of the balanced sample of 3,940 children who participated in the first four waves of the LSAC study, 89 per cent had valid values for the year 5 NAPLAN reading score.

The NAPLAN tests are intended to measure skills developed over time through the school curriculum rather than specific content. For the reading test, students are supplied with written material (eg an

8 page ‘reading magazine’) and then complete a pencil and paper multiple choice exam assessing their comprehension of the material (50 minutes duration). Student’s NAPLAN results are not part of the school assessment process, but parents do receive an individual report documenting their child’s results. The distribution of scores for each school is made publicly available on the My School website, alongside comparisons with national and state averages and results in other schools with similar socio-economic profiles.

Canada

Vocabulary. The Peabody Picture Vocabulary Test - Revised (PPVT-R) was administered to the NLSCY children age 4 to 6. The PPVT-R was designed to measure receptive or hearing vocabulary and in fact can be used for any group, up to adult. The test was developed by Lloyd and Leota Dunn, at the University of Hawaii, and has been widely used in large-scale data collections as well as assessments. A French adaptation of the PPVT-R was developed by the test's authors and Claudia M. Thériault at St. Thomas University in Fredericton, N. B. The French test is called the Échelle de vocabulaire en images Peabody (EVIP).

Verbal parental consent was required before the test was administered. If permission was granted, the interviewer then administered the test to the child in the home. The child looked at pictures on an easel and identified the picture which matched the word the interviewer read out. A total raw score was calculated for each child who completed the PPVT-R by computing correct responses.

Math. All NLSCY children in grade 2 or above were to complete a Mathematics Computation Test. The test that was administered was a shortened version of the Mathematics Computation Test of the standardized Canadian Achievement Tests, Second Edition (CAT/2). CAT/2 is a test series designed to measure achievement in basic skills. The Mathematics Computation Test measures a student's understanding of the operations of addition, subtraction and multiplication and division of whole numbers. The shortened test that was developed for the NLSCY was a 10 question test for grades 2 and 3 and a 15 question test for children in the higher grades.

Test scores for our combined cohort from the age 9 and 11 waves came from tests administered in the child’s home by the interviewer. At the age 7 wave, only half the sample were eligible for the math test, as they were required to be in Grade 2 at the time of the assessment. In addition, the older of our two sub-cohorts took the test in school rather than in the home at age 7. Cycle 3 was the last occasion on which this occurred, because of relatively low school-based response rates (56% for cohort).

Section A5.3 Additional results on achievement gaps at ages 7 and 9

Table A5.2 SES gaps at age 7, by parental education

	Reading age 7, by parental education				Math age 7, by parental education			
	US	UK	AU	CN	US	UK	AU	CN
HL gap	0.86 [A]	0.78 [A]	0.59 [U K]	0.69	0.89 [K C]	0.64 [U C]	.	0.29 [U K]
	0.05	*0.03*	*0.04*	*0.11*	*0.05*	*0.04*	.	*0.12*
HM gap	0.43	0.43	0.35	0.28	0.49 [K C]	0.34 [U]	.	0.17 [U]
	0.04	*0.03*	*0.05*	*0.08*	*0.04*	*0.04*	.	*0.11*
ML gap	0.43 [A]	0.34	0.25 [U]	0.42	0.40 [C]	0.31	.	0.12 [U]
	0.05	*0.03*	*0.05*	*0.10*	*0.05*	*0.04*	.	*0.11*

Note: Standard errors in italics. HL gap = High-low gap; HM gap = High-medium gap; ML gap = medium-low gap. Superscripts indicate that estimate is significantly different (at the 5% level) from the estimate for: Australia (A); Canada (C); UK (K); US (US).
Reading outcomes are: ECLS-K reading score theta (US); BAS Word reading ability score (UK); PPVT score (AU); PPVT-R (CN).
Math outcomes are: ECLS-K math score theta (US); NFER number skills (UK); CAT/2 Mathematical Operations Test (short version) (CN)

Table A5.3 SES gaps at age 9, by parental education

	Reading age 9, by parental education				Math age 9, by parental education			
	US	UK	AU	CN	US	UK	AU	CN
HL gap	0.98 [A]	.	0.61 [U]	.	0.92 [C]	.	.	0.36 [U]
	0.05	.	*0.04*	.	*0.05*	.	.	*0.09*
HM gap	0.56 [A]	.	0.31 [U]	.	0.53 [C]	.	.	0.26 [U]
	0.04	.	*0.04*	.	*0.05*	.	.	*0.07*
ML gap	0.42	.	0.30	.	0.38 [C]	.	.	0.10 [U]
	0.05	.	*0.04*	.	*0.06*	.	.	*0.08*

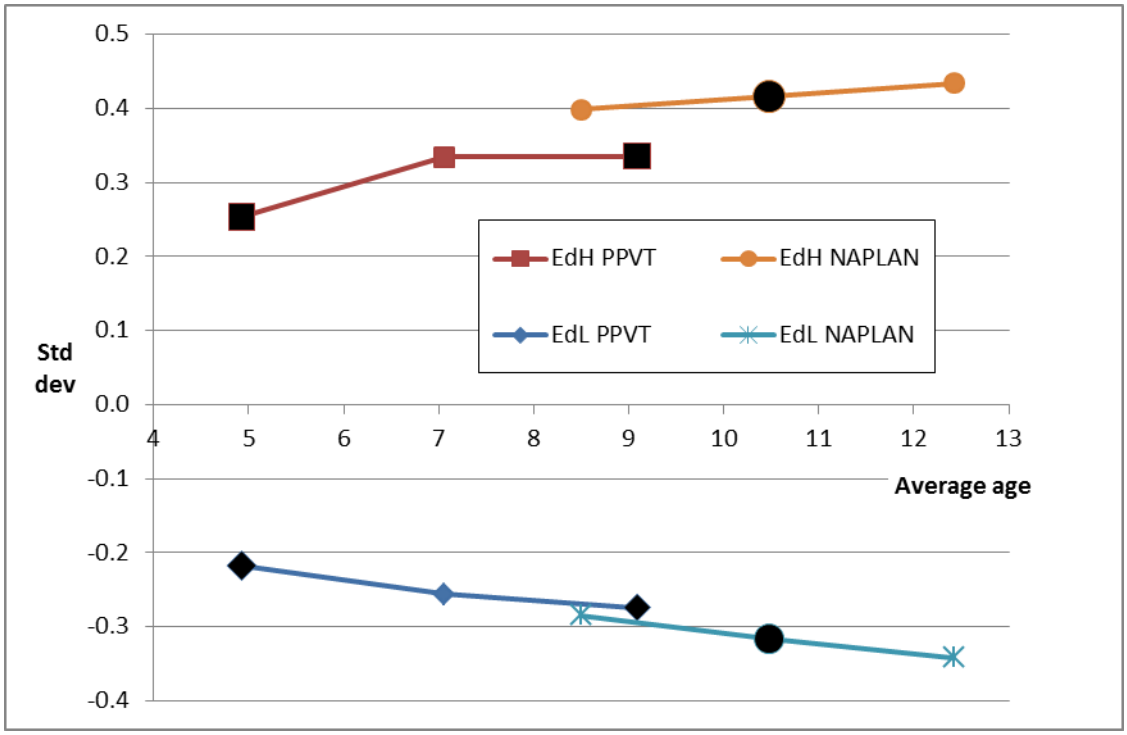
Note: Standard errors in italics. HL gap = High-low gap; HM gap = High-medium gap; ML gap = medium-low gap. Superscripts indicate that estimate is significantly different (at the 5% level) from the estimate for: Australia (A); Canada (C); UK (K); US (US).
Reading outcomes are: ECLS-K reading score theta (US); PPVT score (AU)
Math outcomes are: ECLS-K math score theta (US); CAT/2 Mathematical Operations Test (short version) (CN)

Section A5.4 Australian SES gradient trends

The measures of language ability collected in the Australian LSAC survey vary with age. In Chapter 5, we use the standardized results from the PPVT test conducted at around ages 5 and 9 and the results from the NAPLAN reading test at around age 11 (see Section A5.2).

The Australian mean scores for the top and bottom parental education groups at these ages are shown in Figure 5A.1 below, along with results for other ages. The points emphasised with large black dots correspond to the values shown in Figures 5.1, 5.2 and 5.3 in Chapter 5. (The gaps between the high and low dots in Figure 5A.1 are identical to the total height of the bars shown in Chapter 5).

Figure A5.1 Australia: PPVT and NAPLAN reading scores by SES (primary sample)



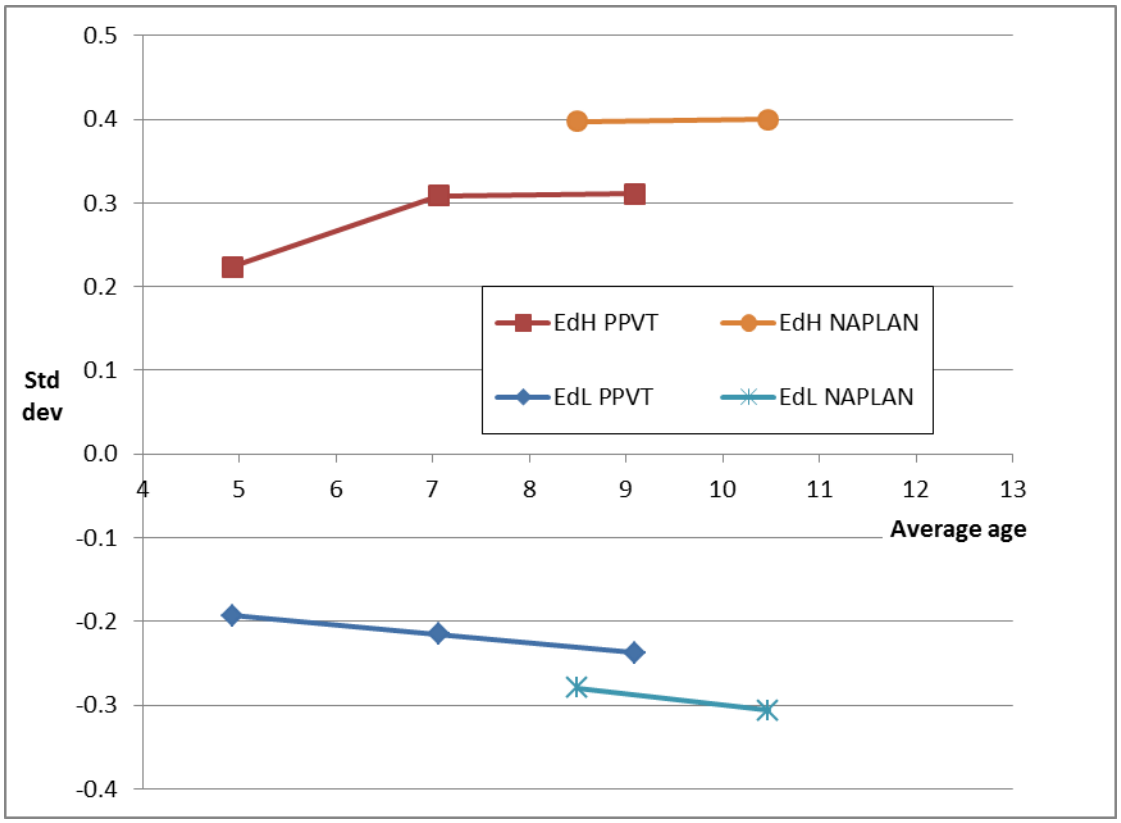
Note: EdH = High parental education level, EdL = Low parental education level. Sample of children with data at each wave (but not necessarily valid outcome scores). Approximate N with valid outcome scores (including the middle education group not shown) = 3,500, 3,800, 3,900 for the three PPVT waves, and 2,600, 3,500 and 3,200 for the three NAPLAN waves.

Though the PPVT test was not continued to older ages, and the NAPLAN test was not administered to young children, both outcome measures are available for children at around age 9. The overlapping scores presented in Figure 5A.1 suggest a possible break in continuity among children of highly educated parents at this age, but no break among the lower education group.

The results shown in Figure 5A.1 (and in the body of the book) are from a sample of children who were present in all waves of the survey (up to age 11). However, not all children had valid data recorded in every wave. This mainly affects the NAPLAN data at ages 8-9 and 12-13. A large fraction of the sample did not undertake the NAPLAN test at age 8-9 because they attended year 3 prior to the national test being introduced. The sample size for the last data point (age 12-13) is also smaller because some children had not yet reached year 7 when the administrative data were extracted.

Figure 5A.2 presents comparable results from a fully balanced sub-sample of children who completed the PPVT test at ages 5, 7 and 9 as well as the NAPLAN tests at ages 8 and 10 (years 3 and 5). The last NAPLAN data point is not included so as to maximise the sample size for this comparison.

Figure A5.2 Australia: PPVT and NAPLAN reading scores by SES (Balanced sample with outcome data for all waves to age 11)



Note: EdH = High parental education level, EdL = Low parental education level. Balanced sample of children with data at all waves. Approximate N (including the middle education group not shown) = 2,250.

For this balanced panel, there is clearer evidence of a break in continuity between the two measures. At around age 9, the top-bottom gap is 0.13 standard deviations greater for the NAPLAN than for the PPVT test (t=2.4).[4] This difference is slightly more than the increase in the gap between the ages 9 and 11 shown in Figures 5.2 and 5.3, and so we conclude that there is no evidence of a real increase in the SES gap in Australia between ages 9 and 11.

The two figures do however both show evidence of an increase in the SES gap between ages 5 and 7. This increase of 0.11 standard deviations in the balanced sample is just significant at the 5% level (though more clearly significant in a balanced sample of just these two waves).

Section A5.5 Input measure definitions

Below we provide details of the raw variables and definitions of the measures shown in Table A5.6 to A5.16. These are at age 11 and parent reported unless otherwise specified.

[4] Statistical tests take account of survey design.

	US	UK	Australia	Canada
Computer in home	Do you have a home computer that *{child}* uses?	Is there a computer, for example, a pc, MAC, laptop or netbook *{child}* can use at home?	Does study child have access to a computer at home? (Exclude handheld or TV plug-in computerised games e.g. Playstation, Wii, Xbox, Nintendo DS, PSP)	Does *{child}* use a computer: …at home? (age 9)
Books in home	About how many children's books does *{child}* have in your home now, including library books? Please only include books that are for children.	About how many books are there in your home? Please do not include magazines, newspapers or children's books. INTERVIEWER: THERE ARE USUALLY ABOUT 15 BOOKS PER FOOT OF SHELVING. 1 0-10 2 11-25 3 26-100 4 101-200 5 201-500 6 More than 500	About how many books does *{child}* have in your home now, including any library books? (Include books owned by older brothers or sisters if age-appropriate for the study child) 0 None; 1 1-10; 2 11-20; 3 21-30; 4 More than 30	Which of the following materials do you currently have in your home? (Include electronic versions of materials listed below) ...more than 25 books
Receives tutoring	Is *{child}* tutored on a regular basis, by someone other than you or a family member, in a specific subject, such as reading, math, science, or a foreign language? (3rd grade, age 9)	Some parents arrange for their children to have extra lessons or classes outside school in subjects they also do at school. Since the last interview on *{date of last interview}*, has *{child}* had any extra classes or lessons in English, Maths or Science?	In the last 12 months has the child received any additional help or tutoring from anyone outside the household?	During the previous school year, did *{child}* receive any additional help or tutoring? How often? Was this help or tutoring provided inside or outside of the school? 1 Inside of the school 2 Outside of the school 3 Both

	US	UK	Australia	Canada
				Coded 1 if outside school or both; coded 0 if inside school only.
Activities question format	Outside of school hours in the past year, has {*child*} participated in ACTIVITY:		In the last 12 months has child regularly participated in any of the following activities?('Regularly' means at least once a week, for three months or more e.g. a sports season)(Exclude activities done as part of the child's normal outside school hours care)	In the past 12 months, outside of school hours, how often has {*child*}: ACTIVITY (If child is only active in one season, the respondent should give the frequency for that season, not try to average over the year. Read categories to respondent.) 1 Most days 2 A few times a week 3 About once a week 4 About once a month 5 Almost never Coded: About once a week or more = 1; About once a month or less = 0
Participates in art/music/performance lessons	COMPOSITE VARIABLE: Coded 1 if any of the following = 1; coded 0 otherwise Music lessons, for example, piano, instrumental music or singing lessons;	-	Art, music or performance lessons (e.g. piano, dance, choir or drama)	...taken lessons or instruction in music, art or other non-sport activities? (age 9)

	US	UK	Australia	Canada
	Art classes or lessons, for example, painting, drawing, sculpturing; Organized performing arts programs			
Music lessons	Music lessons, for example, piano, instrumental music or singing lessons?	Does [^Cohort child's name] play a musical instrument? Do you pay for [^Cohort child's name] to have lessons for any instrument [^he/she] plays?	-	-
Organized club or recreation program	Organized clubs or recreational programs, like scouts?	-	Community group or club (e.g. scouts, guides or cultural group)	...taken part in any clubs, groups or community programs with leadership, such as Brownies, Cubs or church groups? (age 9)
Participates in organized physical activity (sport/ dance/gymnastics/martial arts)	COMPOSITE VARIABLE: Coded 1 if any of the following = 1; coded 0 otherwise Organized athletic activities, like basketball, soccer, baseball, or gymnastics; Dance lessons	How many days a week does *{child}* usually go to a club or class to do sport or any other physical activity like swimming, gymnastics, football, dancing etc? 1 Five or more days a week 2 Four days a week 3 Three days a week 4 Two days a week 5 One day a week 6 Less often 7 Not at all	COMPOSITE VARIABLE: Coded 1 if any of the following = 1; coded 0 otherwise Team sport (e.g. football, cricket or netball); Individual sport, coached or lessons (e.g. swimming, tennis, karate or gymnastics)	COMPOSITE VARIABLE: Coded 1 if any of the following = 1; coded 0 otherwise ...taken part in sports with a coach or instructor (except dance, gymnastics or martial arts); ...taken lessons or instruction in other organized physical activities with a coach or instructor

	US	UK	Australia	Canada
		Coded: One day a week or more = 1; Less often or not at all = 0		such as dance, gymnastics or martial arts? (age 9)
Organized athletics activity	Organized athletic activities, like basketball, soccer, baseball, or gymnastics?	-	-	-
Summer camp	Did *{child}* attend any day or overnight camps over the summer? Please do not include regular child care in this question, but only programs that are referred to as camp. (Fall 1st grade, age 6)	-	-	Did *{child}* attend an overnight camp last summer? Last summer, did *{child}* attend an day camp or recreational or skill-building activity that ran for half days or full days (eg music/reading/athletic program)? Coded 1 if attended either overnight or day camp; coded 0 otherwise
Summer activities	During the summer, did you or another family member take *{child}* to any of the following places? Art, science, or discovery museums; Zoos or aquariums (1st grade) (Fall 1st grade, age 6)	-	-	-

	US	UK	Australia	Canada
Spanking	If *{child}* got so angry that (he/she) hit you, what would you do? Would you spank (him/her)? Yes/no About how many times, if any, have you spanked *{child}* in the past week? (open-ended question) Coded: Spanked one or more times = 1; Never spanked = 0	-	-	Please tell me how often you, as his / her parent, do each of the following when *{child}* breaks the rules or does things that he / she is not supposed to: use physical punishment? 1 Never 2 Rarely 3 Sometimes 4 Often 5 Always
Parent-child interaction	In the past month, how often have you talked with *{child}* about {His/her} day at school? Not at all; A few times a month; A few times a week; Every day	-	-	How often do you and *{child}* talk about school work or behaviour in class? 1 Daily 2 A few times a week 3 Once a week 4 A few times a month 5 Once a month 6 Less than once a month 7 Rarely
	In the past month, how often have you talked with *{child}* about what {he/she} does with {his/her} friends? Not at all; A few times a month; A few times a week; Every day	How often do you talk to *{child}* about things that are important to [^him/her]? 1 Every day or almost every day 2 Several times a week 3 Once or twice a week 4 Once or twice a month	How often do you talk to *{child}* about what is going on in his/her life? 1 Always; 2 Almost always; 3 About half the time; 4 Almost never; 5 Never	How often do you and *{child}* talk about his / her school friends or activities? 1 Daily 2 A few times a week 3 Once a week 4 A few times a month 5 Once a month

	US	UK	Australia	Canada
		5 Less often than once a month 6 Not at all		6 Less than once a month 7 Rarely
Help with homework	Does *{child}* have someone who can help {him/her} with homework in math?	How often does anyone at home help *{child}* with [^his/her] homework? 1 Always 2 Usually 3 Sometimes 4 Never or almost never	During this school year, how often did someone in this household help the child with his/her homework? 1 5 or more days a week; 2 3 or 4 days a week; 3 1 or 2 days a week; 4 Less than once a week; 5 Never	How often do you check his / her homework or provide help with omework? 1 Never or rarely 2 Less than once a month 3 Once a month 4 A few times a month VS 5 Once a week 6 A few times a week 7 Daily
TV watching	On any given weekday, how many hours of television, videotapes, or DVDs on average does *{child}* watch **at home**? (We want you to include television shows, videotapes, and DVDs, but not games like NINTENDO.) How about… a. Before 8:00am? b. Between 3:00pm and dinner time? c. After dinner time? (open-ended questions) Coded as a continuous measure indicating the total hours of watching TV per day	On a normal week day during term time, how many hours does *{child}* spend watching television programmes or films? Please remember to include time spent watching programmes or films on a computer or mobile device as well as on a TV, DVD etc. Please also include time spent before school as well as time after school. 1 None 2 Less than an hour 3 1 hour to less than 2 hours	About how many hours on a typical weekday, would you say that child watches TV or videos at home? 1 Does not watch TV or videos; 2 Less than one hour; 3 1 up to 3 hours; 4 3 up to 5 hours; 5 5 or more hours	

	US	UK	Australia	Canada
		4 2 hours to less than 3 hours 5 3 hours to less than 5 hours 6 5 hours to less than 7 hours 7 7 hours or more		
Child absences from school	Please record the total number of absences for this child for the 2003-2004 school year. [Teacher report]	During this school year, has *{child}* ever been off school for a continuous period of 2 weeks or more, other than for school holidays? Please include any temporary or permanent exclusions.	Does the study child have frequent absences from school? (teacher report) (age 9) During the previous four weeks of school, how many days has study child been absent? (If school holidays have taken place during the past four weeks, exclude school holidays) (parent report) (age 9)	
Regular bedtimes	On weeknights during the school year, does *{child}* usually go to bed at about the same time each night, or does {his/her} bedtime vary a lot from night to night?	On weekdays during term-time, does *{child}* go to bed at a regular time? IF YES, PROBE: Is that sometimes, usually or always? 1 No, never or almost never 2 Yes, sometimes 3 Yes, usually		

	US	UK	Australia	Canada
		4 Yes, always Coded 1 if usually or always; coded 0 otherwise		
Family mealtimes	In a typical week, please tell me the number of days… a. At least some of the family eats breakfast together. b. Your family eats the evening meal together.			
Parental aspirations	How far in school do you expect *{child}* to go? Would you say you expect {him/her} … 1. To receive less than a high school diploma 2. To graduate from high school 3. To attend two or more years of college 4. To finish a four- or five-year college degree 5. To earn a master's degree or equivalent 6. To finish a Ph.D., MD, or other advanced degree	What would you like *{child}* to do when [^he/she] is 16 years of age? Would you like [^him/her] to... 1 ...continue in full-time education, such as school or college, 2 ...get an apprenticeship or other work-placed learning, 3 .. or get a job (with some part-time education or training)? 4 SPONTANEOUS: do something else Coded 1 for full-time education only; coded 0 otherwise	Looking ahead, how far do you think study child will go in his/her education? 1 Leave school before finishing secondary school; 2 Complete secondary school; 3 Complete a trade or vocational training course; 4 Go to university and complete a degree; 5 Obtain post-graduate qualifications at a university (e.g. Master degree or Doctoral degree)	How far do you hope *{child}* will go in school? 1 Primary/elementary school 2 Secondary or high school 3 Community college, CEGEP or nursing school 4 Trade, technical or vocational school, or business college 5 University 6 Post-secondary, unspecified 7 Other

	US	UK	Australia	Canada
		How likely or unlikely do you think it is that *{child}* will attend university? 1 Very likely 2 Fairly likely 3 Not very likely 4 Not at all likely Coded 1 if Very or Fairly likely; coded 0 otherwise		
Private/faith school	Is this a public school? (1=Public 2=Private) [Asked to school administrator] Is this private school … a. Catholic? Diocesan? Parish? Private order? b. Private, other religious affiliation? c. Private school accredited by NAIS? d. Other private? e. Special Education school—primarily serves children with disabilities? f. An Early Childhood Center (school or center includes	Do school fees have to be paid for [^Cohort child's name] to go to this school? INTERVIEWER: BY FEES, WE MEAN COMPULSORY FEES FOR THE CHILD TO ATTEND, NOT 'TOP-UP FEES' FOR EXTRA-CURRICULAR ACTIVITIES/ OTHER REASONS Some schools are primarily for children of a particular faith or religion. Is [^Cohort child's name]'s current school primarily for children of a particular faith or religion? 1 No / not a faith school 2 Christian (Church of England)	Does study child attend… 1 a government school; 2 a Catholic school; 3 an independent or private school; 4 not in school	What type of school is ^INFO.FNAME currently in? Is it a: 1 Public school? 2 Catholic school, publicly funded? 3 Private school?

	US	UK	Australia	Canada
	preschool and/or early elementary grades)?	3 Christian (Catholic) 4 Other Christian 5 Jewish 6 Islam/Muslim 7 Other		
Percent of school eligible for free school lunch	How many children in your school were (a) eligible for and (b) participating in the school lunch program as of October 2003? [School report]			
Percent of class limited English proficient	What percent of children in this school and in fifth grade are limited English proficient (LEP)? [School report]	How many children in this child's class come from homes where English is an additional language? [Teacher report]	Number of children from an English speaking background in the class [Teacher report]	
Teacher years of experience	Counting this school year, how many years have you been a school teacher, including part-time teaching? (reading/math teacher)	In total, how many years have you been teaching ?	How many years teaching experience do you have… altogether as a teacher?	
Teacher use of ability grouping	How often do you divide this reading/math class into instructional groups, based on achievement levels, for reading activities or lessons? (Never vs Less than once a week/Once or twice a week/Three or four times a week/Daily)	We are interested to know about class groupings in this child's year. Some schools group children from different classes in the same year by general ability and children are taught in these groups for most or all lessons. We refer to this as streaming.	How often do you organise your class in achievement level groupings for…? Reading? Maths? 1 Never; 2 Less than once a week; 3 Once or twice a week; 4 Three or more times per week;	

	US	UK	Australia	Canada
		Some schools group children from different classes in the same year by ability for certain subjects only and children may be taught in different ability groups for different subjects. We refer to this as setting. In this child's year, is there streaming? Which stream is this child in ? In this child's year are there sets for English? Math? Science? Which set is child in?	5 Daily	
Share with diagnosed disability	How many children in this mathematics class have the following characteristics? Have a diagnosed disability and need special services (%) [Teacher report]	How many children in this child's class have SEN statements? [Teacher report]	How many children in the class…? Have a diagnosed disability (e.g. intellectual, sensory, physical, Autistic Spectrum Disorder, developmental delay) 0 0; 1 1-5; 2 6-10; 3 11-20; 4 21-30; 5 31 or more [Teacher report]	
Share gifted and talented	How many children in this mathematics class have the following characteristics?			

	US	UK	Australia	Canada
	Are classified as Gifted and Talented (%) [Teacher report]			
Average class size	Average class size reported by: math teachers.	How many children are there on this child's class register? [Teacher report]	How many children are present in your class for the main educational program? [Teacher report]	
Grade retention	A binary indicator with a value of 1 if a child had ever been retained by the 5th grade survey, using indicators on whether or not to attend kindergarten first time and on the grade level of the 1st, 3rd, and 5th grade surveys			Has ^INFO.FNAME ever repeated a grade (including kindergarten)?

Table A5.4 Home resources reported by 4th grade students

	US	*UK*	*AU*	*CA*
More than 100 books in home (%)	28	36	41	35
Own room and internet connection in home (%)	64	73	74	77

Source: PIRLS 2011 International Results in Reading, Exhibit 4:2. Data for the UK refer to England only.

Section A5.6 Gender gaps

In this section we explore first whether average achievement outcomes differ with gender, and second whether the SES gap between children with high- and low-educated parents tends to be larger for boys or for girls. The standardized outcome scores used here are the same ones from our main balanced panel sample analyses from Chapters 3 to 5; that is, the mean and variance used to standardize the test scores are calculated from the pooled sample of boys and girls. The analyses presented below calculate the means and SES gaps of these standardized scores for separate samples by child gender, and test for differences between them.

In Table A5.5, below, the left-most panel investigates the first of these questions, by looking at differences in the *level* of mean outcomes by gender. For example, for the US, the first row shows that in Fall K girls scores on average 0.12 SD above the mean in reading, boys score 0.11SD below the mean, giving a gender gap of 0.23SD (G-B). This gap has a standard error of 0.05 (SE(G-B)), and is significant at the 1% level (Pr(G=B) is the test of the null hypothesis that the means are equal). The remaining 3 panels show the within-gender SES *gaps* for each outcome and test whether they are different. Again to illustrate using the first row, the average gap between the daughters of high and low educated parents is 1.04SD, and between the sons of high and low educated parents it is 0.95SD. The difference in the gaps, then, is 0.09SD (G-B), an estimate not statistically significantly different from zero. The two right panels focus on SES gaps at the top and bottom of the distribution respectively, and the p-values of differences that are significant at the 5% level are highlighted in bold.

Table A5.5 Gender differences in achievement outcomes and SES gaps

A. United States

	Outcome means					High-low SES gaps				High-medium SES gaps				Low-medium SES gaps			
Outcome	Girls	Boys	G-B	SE(G-B)	Pr(G=B)	Girls	Boys	G-B	Pr(G=B)	Girls	Boys	G-B	Pr(G=B)	Girls	Boys	G-B	Pr(G=B)
Reading																	
Fall K (5)	0.12	-0.11	0.23	0.05	**0.000**	1.04	0.95	0.09	0.382	0.59	0.50	0.08	0.402	-0.46	-0.45	-0.01	0.938
Spring K (6)	0.12	-0.11	0.23	0.05	**0.000**	0.90	0.84	0.07	0.484	0.39	0.40	-0.01	0.949	-0.51	-0.44	-0.07	0.398
1st grade (7)	0.13	-0.12	0.25	0.04	**0.000**	0.84	0.87	-0.02	0.818	0.40	0.46	-0.05	0.541	-0.44	-0.41	-0.03	0.765
3rd grade (9)	0.11	-0.11	0.22	0.05	**0.000**	0.98	0.98	0.00	0.965	0.50	0.63	-0.13	0.136	-0.48	-0.35	-0.13	0.178
5th grade (11)	0.08	-0.08	0.16	0.05	**0.001**	0.94	1.07	-0.13	0.173	0.48	0.65	-0.17	0.071	-0.46	-0.42	-0.04	0.714
8th grade (14)	0.12	-0.11	0.23	0.04	**0.000**	1.04	1.09	-0.05	0.641	0.61	0.62	-0.01	0.931	-0.43	-0.47	0.04	0.633
Math																	
Fall K (5)	0.04	-0.04	0.07	0.04	0.099	1.00	1.03	-0.04	0.684	0.58	0.51	0.07	0.461	-0.42	-0.52	0.11	0.230
Spring K (6)	0.03	-0.02	0.05	0.05	0.294	0.92	0.99	-0.07	0.500	0.43	0.51	-0.08	0.387	-0.49	-0.48	-0.01	0.912
1st grade (7)	-0.01	0.01	-0.02	0.05	0.727	0.87	0.92	-0.05	0.613	0.45	0.53	-0.08	0.354	-0.42	-0.39	-0.03	0.732
3rd grade (9)	-0.09	0.09	-0.18	0.05	**0.000**	0.90	0.95	-0.05	0.624	0.48	0.59	-0.11	0.256	-0.42	-0.36	-0.06	0.537
5th grade (11)	-0.10	0.10	-0.20	0.05	**0.000**	0.89	1.00	-0.11	0.282	0.53	0.58	-0.04	0.678	-0.36	-0.43	0.07	0.504
8th grade (14)	-0.03	0.03	-0.06	0.05	0.256	0.95	1.09	-0.14	0.187	0.58	0.57	0.02	0.870	-0.37	-0.53	0.16	0.108

B. United Kingdom

	Outcome means					High-low SES gaps				High-medium SES gaps				Low-medium SES gaps			
Outcome	Girls	Boys	G-B	SE(G-B)	Pr(G=B)	Girls	Boys	G-B	Pr(G=B)	Girls	Boys	G-B	Pr(G=B)	Girls	Boys	G-B	Pr(G=B)
Vocab age 5	0.02	-0.02	0.05	0.04	0.174	0.76	0.82	-0.05	0.459	0.37	0.34	0.03	0.603	-0.39	-0.48	0.08	0.209
Reading age 7	0.09	-0.08	0.17	0.03	**0.000**	0.69	0.87	-0.18	**0.004**	0.40	0.46	-0.06	0.284	-0.28	-0.40	0.12	**0.035**
Math age 7	0.00	0.00	-0.01	0.04	0.790	0.58	0.70	-0.12	0.082	0.33	0.34	-0.01	0.867	-0.25	-0.36	0.11	0.095
Verbal age 11	-0.04	0.04	-0.07	0.04	0.064	0.66	0.68	-0.03	0.689	0.36	0.32	0.03	0.569	-0.30	-0.36	0.06	0.382

C. Australia

	Outcome means					High-low SES gaps				High-medium SES gaps				Low-medium SES gaps			
	Girls	Boys	G-B	SE(G-B)	Pr(G=B)	Girls	Boys	G-B	Pr(G=B)	Girls	Boys	G-B	Pr(G=B)	Girls	Boys	G-B	Pr(G=B)
Vocab age 5	0.07	-0.06	0.13	0.04	**0.002**	0.49	0.46	0.03	0.697	0.22	0.21	0.01	0.933	-0.27	-0.24	-0.02	0.798
Vocab age 7	-0.04	0.04	-0.08	0.04	0.052	0.57	0.61	-0.04	0.634	0.34	0.35	-0.02	0.828	-0.24	-0.25	0.02	0.846
Vocab age 9	-0.03	0.03	-0.06	0.04	0.137	0.64	0.58	0.05	0.487	0.33	0.30	0.03	0.728	-0.30	-0.28	-0.02	0.818
Reading age 9	0.09	-0.09	0.18	0.04	**0.000**	0.61	0.75	-0.14	0.097	0.48	0.53	-0.04	0.667	-0.13	-0.22	0.10	0.336
Reading age 11	0.13	-0.13	0.26	0.04	**0.000**	0.71	0.76	-0.05	0.496	0.44	0.46	-0.01	0.864	-0.27	-0.30	0.04	0.663
Reading age 13	0.12	-0.12	0.25	0.04	**0.000**	0.80	0.76	0.04	0.617	0.52	0.39	0.14	0.130	-0.27	-0.37	0.10	0.266
Math age 9	0.01	-0.01	0.03	0.04	0.539	0.60	0.65	-0.05	0.528	0.45	0.43	0.02	0.862	-0.15	-0.22	0.07	0.472
Math age 11	-0.05	0.05	-0.11	0.04	**0.010**	0.61	0.76	-0.14	0.065	0.42	0.48	-0.06	0.469	-0.19	-0.27	0.08	0.343
Math age 13	-0.06	0.06	-0.11	0.04	**0.007**	0.75	0.80	-0.05	0.487	0.54	0.51	0.03	0.733	-0.21	-0.29	0.08	0.307
Common sample																	
Vocab age 5	0.09	-0.09	0.18	0.05	**0.000**	0.42	0.41	0.01	0.925	0.21	0.24	-0.03	0.766	-0.22	-0.17	-0.04	0.721
Vocab age 7	-0.02	0.02	-0.04	0.05	0.349	0.50	0.54	-0.04	0.670	0.39	0.44	-0.05	0.630	-0.11	-0.10	-0.01	0.897
Vocab age 9	-0.03	0.02	-0.05	0.05	0.331	0.53	0.56	-0.03	0.769	0.42	0.34	0.08	0.482	-0.11	-0.22	0.11	0.368
Reading age 9	0.08	-0.08	0.16	0.05	**0.001**	0.63	0.73	-0.11	0.262	0.49	0.57	-0.08	0.469	-0.14	-0.16	0.03	0.810
Reading age 11	0.12	-0.12	0.24	0.05	**0.000**	0.72	0.69	0.03	0.757	0.48	0.48	0.00	1.000	-0.24	-0.21	-0.03	0.787

D. Canada

	Outcome means					High-low SES gaps				High-medium SES gaps				Low-medium SES gaps			
Outcome	Girls	Boys	G-B	SE(G-B)	Pr(G=B)	Girls	Boys	G-B	Pr(G=B)	Girls	Boys	G-B	Pr(G=B)	Girls	Boys	G-B	Pr(G=B)
Vocab age 5	0.00	0.00	0.00	0.06	0.938	0.63	0.62	0.01	0.974	0.36	0.29	0.07	0.614	-0.26	-0.33	0.06	0.630
Vocab age 7	-0.03	0.03	-0.06	0.07	0.422	0.70	0.68	0.02	0.934	0.27	0.28	-0.02	0.911	-0.43	-0.40	-0.03	0.867
Math age 7	-0.10	0.11	-0.21	0.09	**0.017**	0.09	0.46	-0.37	0.137	0.05	0.25	-0.19	0.350	-0.04	-0.21	0.17	0.429
Math age 9	-0.02	0.02	-0.04	0.06	0.443	0.16	0.54	-0.38	**0.032**	0.12	0.38	-0.26	**0.043**	-0.05	-0.16	0.11	0.476
Math age 11	-0.03	0.03	-0.06	0.05	0.204	0.41	0.70	-0.29	0.063	0.20	0.52	-0.33	**0.009**	-0.22	-0.18	-0.04	0.756

Outcomes are age-adjusted standardized scores. Balanced panel samples.

Section A5.7 SES gaps in resources at age 11

Table A5.6 Out-of-school resources and activities at age 11

	US	UK	AU	CA
Computer in home that child uses (%)				
All	82	97	96	84
High school or less	66	94	94	71
Some post-secondary	87	98	96	85
College degree of more	96	99	99	92
More than 30 children's books in home (%)				
All	78	-	75	-
HS or less	65	-	68	-
Some post-secondary	82	-	76	-
College degree or more	89	-	83	-
More than 25 books (any kind) in home (%)				
All	-	72	-	92
HS or less	-	57	-	83
Some post-secondary	-	76	-	94
College degree or more	-	93	-	97
Number of children's books in home				
All	101	-	-	-
High school or less	67	-	-	-
Some post-secondary	102	-	-	-
College degree or more	144	-	-	-
Number of books in home (any kind)				
All	-	131	-	-
High school or less	-	78	-	-
Some post-secondary	-	126	-	-
College degree or more	-	228	-	-
Participates in art/music/performance lessons (%)				
All	44	-	44	26
High school or less	33	-	32	13
Some post-secondary	44	-	42	21
College degree or more	58	-	61	43
Participates in music lessons (%)				
All	28	39	-	-
High school or less	16	28	-	-
Some post-secondary	27	37	-	-
College degree or more	43	59	-	-
Participates in organized club/recreation program (%)				
All	27	-	12	29
High school or less	19	-	10	20
Some post-secondary	27	-	15	29
College degree or more	38	-	14	36

	US	UK	AU	CA
Participates in organized physical activity (sport/dance/gymnastics/martial arts) (%)				
All	67	74	76	72
High school or less	54	64	69	52
Some post-secondary	65	77	76	72
College degree or more	83	86	84	86
Participates in organized athletics (%)				
All	61	-	-	-
High school or less	48	-	-	-
Some post-secondary	60	-	-	-
College degree or more	78	-	-	-
Receives tutoring (%)				
All	15	21	16	6
High school or less	19	18	16	3
Some post-secondary	16	21	15	6
College degree or more	11	26	18	8

Note: Data on computer and books in home, lessons, clubs, and athletics for US refer to children in 5th grade. Data on computer in home, art/music/performance lessons, organized clubs, physical activities for Canada refer to children age 8/9. Data on tutoring for US refer to children in 3rd grade. Data for UK refer to children age 11.
"-" indicates data not available

Table A5.7 Summer enrichment activities

	US	UK	AU	CA
Child attended camp summer before 1st grade				
All	20	-	-	-
High school or less	8	-	-	-
Some post-secondary	21	-	-	-
College or more	34	-	-	-
Child attended camp summer before 5th grade				
All	-	-	-	51
High school or less	-	-	-	39
Some post-secondary	-	-	-	50
College or more	-	-	-	63
Child taken to a museum summer before 1st grade				
All	38	-	-	-
High school or less	22	-	-	-
Some post-secondary	38	-	-	-
College or more	56	-	-	-
Child taken to zoo or aquarium summer before 1st grade				
All	60	-	-	-
High school or less	51	-	-	-
Some post-secondary	62	-	-	-
College or more	68	-	-	-

"-" indicates data not available

Table A5.8 Parents' use of spanking at age 11

	US	UK	AU	CA
Would spank in a hypothetical situation (%)				
All	24	-	-	-
High school or less	28	-	-	-
Some post-secondary	26	-	-	-
College or more	16	-	-	-
Spanked in past week (%)				
All	12	-	-	-
High school of less	12	-	-	-
Some post-secondary	13	-	-	-
College or more	12	-	-	-
Parent uses physical punishment Sometimes/often/always (%)				
All	-	-	-	3
High school or less	-	-	-	4
Some post-secondary	-	-	-	3
College degree or more	-	-	-	1

"-" indicates data not available

Table A5.9 Family routines

	US	UK	AU	CA
Child watches 3 or more hours of TV on a weekday (%)				
All	28	16	20	17
High school or less	35	21	26	25
Some post-secondary	30	16	19	15
College degree or more	16	9	14	11
Days missed school in last school year (school report)				
All	6.7	-	-	-
High school or less	7.5	-	-	-
Some post-secondary	6.7	-	-	-
College degree or more	5.6	-	-	-
Child has frequent absences from school (teacher report) (%)				
All	-	-	5	-
High school or less	-	-	7	-
Some post-secondary	-	-	5	-
College degree or more	-	-	4	-

	US	UK	AU	CA
Days child absent from school in past four weeks (parent report)				
All	-	-	1.2	-
High school or less	-	-	1.4	-
Some post-secondary	-	-	1.2	-
College degree or more	-	-	1.0	-
Child been off school two weeks continuously this school year (parent report) (%)				
All	-	3.3	-	-
High school or less	-	4.4	-	-
Some post-secondary	-	3.3	-	-
College degree or more	-	1.4	-	-
Child usually has regular bedtimes on weekdays (%)				
All	92	90	-	-
High school or less	89	87	-	-
Some post-secondary	93	91	-	-
College degree or more	94	93	-	-
Days in a typical week family eats breakfast together				
All	3.4	-	-	-
High school or less	2.9	-	-	-
Some post-secondary	3.3	-	-	-
College degree or more	4.1	-	-	-
Days in a typical week family eats evening meal together				
All	5.5	-	-	-
High school or less	5.5	-	-	-
Some post-secondary	5.5	-	-	-
College degree or more	5.3	-	-	-

Note: Reports of child absences from school in AU for children age 8/9. Days missed school comes from school report in the US. TV and family meals for children age 8/9 for CA.
"-" indicates data not available

Table A5.10 Parent-child interaction

	US	UK	AU	CA
Parent talks with child about his/her day at school every day (%)				
All	83	-	-	-
High school or less	79	-	-	-
Some post-secondary	85	-	-	-
College degree or more	86	-	-	-
Parent talks with child about schoolwork or behavior in class every day (%)				
All	-	-	-	68
High school or less	-	-	-	64
Some post-secondary	-	-	-	68
College degree or more	-	-	-	72

	US	UK	AU	CA
Parent talks with child about what he/she does with friends every day (%)				
All	58	-	-	-
High school or less	57	-	-	-
Some post-secondary	60	-	-	-
College degree or more	56	-	-	-
Parent talks with child about school friends or activities every day (%)				
All	-	-	-	69
High school or less	-	-	-	64
Some post-secondary	-	-	-	70
College degree or more	-	-	-	71
Parent talks with child about things that are important to him/her every day (%)				
All	-	65	-	-
High school or less	-	61	-	-
Some post-secondary	-	67	-	-
College degree or more	-	69	-	-
Parent talks with child about what is going on in his/her life: Always (%)				
All	-	-	59	-
High school or less	-	-	57	-
Some post-secondary	-	-	59	-
College degree or more	-	-	60	-
Child has someone at home who can help with math homework (%)				
All	96	-	-	-
High school or less	93	-	-	-
Some post-secondary	97	-	-	-
College degree or more	98	-	-	-
Someone at home helps child with homework once a week or more (%)*				
All	-	90	81	90
High school or less	-	87	79	87
Some post-secondary	-	91	84	91
College degree or more	-	92	81	92

Note: Homework help question for UK uses categories Sometimes/Usually/Always (vs Never or almost never).
"-" indicates data not available

Table A5.11 Parental aspirations/expectations for child's future

	US	UK	AU	CA
Expect child to attain 4/5 year college degree or more (%)				
All	73	74*	64	-
High school or less	57	64	49	-
Some post-secondary	72	75	64	-

College degree or more	93	91	84	-
	US	UK	AU	CA
Expect child to attain any education beyond HS (%)				
All	89	-	83	-
High school or less	81	-	73	-
Some post-secondary	91	-	86	-
College degree or more	98	-	94	-
Hope child will attain any education beyond HS (%)				
All	-	88	-	92
High school or less	-	85	-	85
Some post-secondary	-	88	-	93
College degree or more	-	94	-	98

Note: For the UK, beyond HS means remaining in full-time education post-16 (i.e. post-compulsory education). AU question asked when children were age 8/9. *Expect college in UK means parent responded “very likely” or “fairly likely”. If “very likely” only is taken as expectation, numbers are: 34, 24, 31, 55.
“-“ indicates data not available

Table A5.12 Schools attended by children at age 11

	US	UK	AU	CA
Average class size				
All	23	27	26	-
HS or less	23	26	26	-
Some post-secondary	23	27	26	-
College degree or more	23	27	26	-
Private school (%)				
All	11	5	13	4
High school or less	4	1	7	1
Some post-secondary	9	3	13	3
College degree or more	20	14	22	10
Catholic school (%)				
All	-	-	21	17
High school or less	-	-	18	13
Some post-secondary	-	-	24	19
College degree or more	-	-	23	18
Religious/faith school (%)				
All	9	26	-	-
HS or less	3	22	-	-
Some post-secondary	8	27	-	-
College degree or more	16	32	-	-

Note: Class size data for UK refer to England and Wales only, and for the US refer to math class. All US religious/faith schools are private; in the UK they can be public or private. Catholic school is a separate category in AU and CA.
"-" indicates data not available

Table A5.13 Composition of schools/class at age 11

	US	UK	AU	CA
Students eligible for free school lunch (% of school)				
All	39	-	-	-
High school or less	51	-	-	-
Some post-secondary	38	-	-	-
College degree or more	23	-	-	-
Students with limited English proficiency (% of class)				
All	5	11	13	-
High school or less	9	15	13	-
Some post-secondary	4	9	13	-
College degree or more	2	7	13	-

Note: Data on limited English proficiency for UK refer to England and Wales only and refer to proportion of class with English as an additional language. Data for US refer to 5th grade.

"-" indicates data not available

Table A5.14 Teacher experience, age 11 classrooms

	US	UK	AU	CA
<5 years of experience (%)				
All	19	15	21	-
High school or less	22	15	22	-
Some post-secondary	19	16	21	-
College or more	14	13	20	-

Table A5.15 Ability grouping, special needs, gifted/talented, and class size at age 11

	US	UK	AU	CA
Teachers using ability groups for reading (%)				
All	67	47	94	-
High school or less	73	48	94	-
Some post-secondary	67	46	93	-
College degree or more	58	47	94	-
Teachers using ability groups for math (%)				
All	61	64	93	-

High school or less	65	64	92	-
Some post-secondary	61	63	92	-
College degree or more	53	66	93	-
	US	UK	AU	CA
Students with diagnosed disability (% of class)				
All	12	7	9	-
High school or less	15	8	9	-
Some post-secondary	11	6	9	-
College degree or more	10	5	8	-
Share of math class gifted and talented (%)				
All	8	-	-	-
High school or less	6	-	-	-
Some post-secondary	7	-	-	-
College degree or more	12	-	-	-

Note: Data on diagnosed disability pertain to math class for US. Data for UK refer to England and Wales only, and refer to children with a statement of special needs. UK ability grouping refers to "setting" – grouping children from different classes into ability groups for teaching of some/all lessons. Within-class ability grouping was not asked about at age 11. Data from age 7 suggest this form of ability grouping is also common in the UK (eg 85% were grouped within-class for math and 87% for English at that age).

"-" indicates data not available

Table A5.16 Retention in primary school

	US	UK	AU	CA
Share of children retained by 5th grade (%)				
All	19	-	5	5
High school or less	26	-	5	10
Some post-secondary	17	-	6	5
College degree or more	11	-	5	2

"-" indicates data not available

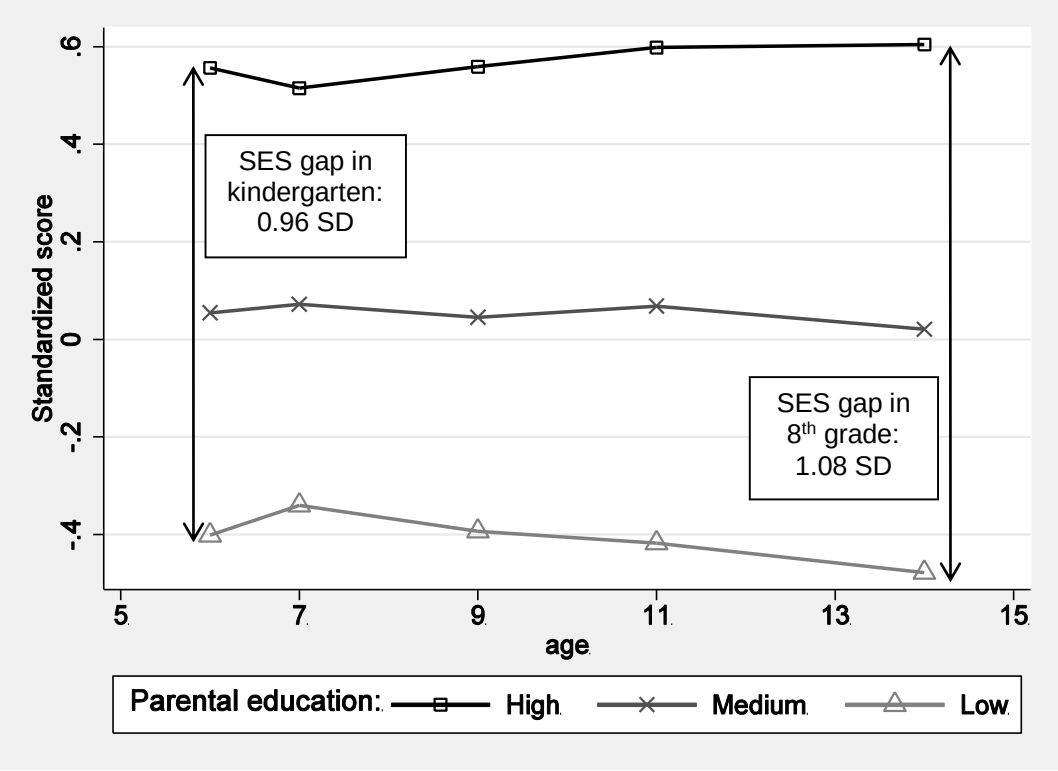
Appendix to Chapter 6

Section A6.1 SES-group means in the balanced and unbalanced samples

For the comparative analyses in Chapters 3 to 5, we use a balanced panel sample of children who were observed in all five waves between Fall Kindergarten (age 5) and 5th grade (age 11), and weight with longitudinal parent-level weights. In order to avoid throwing away data, for the trajectories analysis we use an unbalanced panel, using all available observations at each of the waves, with the proviso that a child must have a valid test score at both Fall and Spring Kindergarten. These estimates are weighted using cross-sectional child-level weights.

Mean reading scores in the unbalanced panel are plotted in Figure 6.1 of Chapter 6. Figure A6.1 below provides the equivalent chart for math test scores. Tables A6.1 and A6.2 compare mean estimates from the balanced and unbalanced panel samples for reading and math scores respectively. Differences in the estimates of the SES gaps between the two samples are very minor.

Figure A6.1 Mean math scores in the unbalanced panel between Spring Kindergarten and 8th grade, by parental education group



Note: The chart plots the average standardized math score of children from the three parental SES groups at: spring kindergarten, 1st, 3rd, 5th and 8th grade. Sample sizes are about 17,090; 14,190; 12,200; 9,630; and 8,020 respectively (all numbers rounded to nearest 10 in accordance with NCES reporting rule). All estimates are weighted using cross-sectional weights to be representative of the underlying national population.

Table A6.1 Mean reading scores in the balanced and unbalanced panels, by parental education group

	Balanced panel sample				Unbalanced panel sample			
		Parental education:				Parental education:		
	N	Low	Medium	High	N	Low	Medium	High
Fall K (5)	8,120	-0.46	0.00	0.54	17,170	-0.42	0.03	0.58
Spr K (6)	8,260	-0.42	0.06	0.46	17,170	-0.37	0.06	0.53
1st G (7)	8,190	-0.40	0.03	0.46	14,235	-0.33	0.08	0.49
3rd G (9)	8,020	-0.44	-0.02	0.55	12,161	-0.39	0.05	0.57
5th G (11)	7,970	-0.45	0.00	0.56	9,654	-0.42	0.09	0.61
8th G (14)	-	-	-	-	7,959	-0.47	0.02	0.61

Table A6.2. Mean math scores in the balanced and unbalanced panels, by parental education group

	Balanced panel sample				Unbalanced panel sample			
		Parental education:				Parental education:		
	N	Low	Medium	High	N	Low	Medium	High
Fall K (5)	8,150	-0.46	0.01	0.56	17,085	-0.44	0.05	0.60
Spr K (6)	8,230	-0.44	0.04	0.51	17,085	-0.40	0.05	0.56
1st G (7)	8,190	-0.40	0.01	0.49	14,190	-0.34	0.07	0.51
3rd G (9)	8,020	-0.40	-0.02	0.51	12,203	-0.39	0.05	0.56
5th G (11)	7,980	-0.41	-0.02	0.53	9,631	-0.42	0.07	0.60
8th G (14)	-	-	-	-	8,020	-0.48	0.02	0.60

Section A6.2 Common trajectories models

The common trajectories models depicted in Figures 6.3 and 6.4 predict a child's later test score purely on the basis of their score in Spring Kindergarten. Specifically, the models used for prediction are

$$Y_{it} = \alpha_0 + \alpha_1 Y_{i6} + \alpha_2 Y_{i6}^2 + \varepsilon_{it} \qquad (A6.1)$$

Where Y_{it}, $t = 7,9,11,14$, is the standardized test score of child i at a later measurement occasion (1st, 32rd, 5th or 8th grade), Y_{i6} is child i's score in Spring Kindergarten (age 6), and ε_{it} is an uncorrelated error term. The squared term, Y_{i6}^2, is included to allow for a non-linear relationship between initial and later test scores.

As is well known, measurement error in Y_{i6} will lead to attenuation bias in least squares estimates of α_1 and α_2, with the results that we would tend to underestimate the degree of persistence in a child's test score over time. To tackle this we use an instrumental variables (IV) approach, with the child's test score in Fall Kindergarten (measured at roughly age 5, six months prior to the Spring Kindergarten score) and its square serving as the instruments. The first stage equations are therefore:

$$Y_{i6} = \beta_0 + \beta_1 Y_{i5} + \beta_2 Y_{i5}^2 + \eta_{it} \qquad (A6.2a)$$

$$Y_{i6}^2 = \gamma_0 + \gamma_1 Y_{i5} + \gamma_2 Y_{i5}^2 + \nu_{it} \qquad (A6.2b)$$

Models were estimated using the `svy: regress` (for least squares estimates) and `svy: ivregress 2sls` commands (for the IV estimates) in Stata /MP 13.1. Estimates of the parameters from equation A6.1 are provided in Table A6.3 for reading scores, contrasting estimates from the ordinary least squares (OLS) and IV approaches. (Estimates of the first-stage parameters are shown in Table A6.4.) The inclusion of the quadratic terms makes interpretation of the raw coefficients difficult, and in Figures 6.2 and 6.3 of the chapter we predict from the IV models, calculating the later test scores expected for children with different initial scores in Spring Kindergarten. The predictions in Figure 6.3 plug in as initial scores the SES group mean scores shown in Table A6.1. Defining $Y_{6(L)}$, $Y_{6(M)}$ and $Y_{6(H)}$ as the mean Spring Kindergarten scores of children with low, medium and high educated parents respectively, the predicted score for a child in SES group ($G = \mathrm{L}, \mathrm{M}, \mathrm{H}$) at time t is:

$$\hat{Y}_{t(G)} = \hat{\alpha}_0 + \hat{\alpha}_1 Y_{6(G)} + \hat{\alpha}_2 Y_{6(G)}^2 \qquad (A6.3)$$

These predictions are shown for the OLS and IV models for reading at the bottom on Table A6.3. The corresponding results for math scores are provided in Tables A6.5 and A6.6, and Figure A6.2 provides the math version of Figure 6.3 (which shows results for reading) shown in the main chapter.

As expected, the OLS estimates considerably over-state the degree of convergence in test scores over time. For example, the counterfactual high/low SES gap in reading in 8th grade from the OLS model is 0.52 SD, compared with 0.64 SD from the IV model. Given that the raw gap in 8th grade is 1.09 SD, OLS common trajectories estimates predict a counterfactual gap that is 48% of that is observed in reality, whereas in fact the persistence of initial differences is greater. The IV estimates tell us that even if trajectories from a given initial score were not differentiated at all by SES, those initial differences would be associated with an 8th grade gap 59% of what observed in practice.

Table A6.3 Common trajectories model estimates for reading scores

	OLS				IV			
	1st grade (7)	3rd grade (9)	5th grade (11)	8th grade (14)	1st grade (7)	3rd grade (9)	5th grade (11)	8th grade (14)
Y_{i6}	0.772	0.688	0.672	0.583	0.871	0.844	0.824	0.731
	[0.011]*	[0.015]*	[0.017]*	[0.016]*	[0.015]*	[0.019]*	[0.024]*	[0.025]*
Y_{i6}^2	-0.050	-0.058	-0.052	-0.014	-0.084	-0.104	-0.099	-0.066
	[0.005]*	[0.007]*	[0.011]*	[0.010]	[0.008]*	[0.011]*	[0.016]*	[0.016]*
Constant	0.056	0.043	0.057	-0.0002	0.082	0.072	0.087	0.035
	[0.010]*	[0.017]*	[0.021]*	[0.028]	[0.010]*	[0.018]*	[0.023]*	[0.030]
N	14,240	12,160	9,650	7,960	14,240	12,160	9,650	7,960
R-squared	0.599	0.439	0.423	0.324	0.588	0.415	0.400	0.300
Predicted scores associated with SES group means at Spr K								
Low education ($Y_{6(L)} = -0.37$)	-0.24	-0.22	-0.20	-0.22	-0.25	-0.25	-0.23	-0.24
Med education ($Y_{6(M)} = 0.06$)	0.10	0.08	0.10	0.03	0.13	0.12	0.14	0.08
High education ($Y_{6(H)} = 0.53$)	0.45	0.39	0.40	0.30	0.52	0.49	0.49	0.40

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to the nearest 10 in accordance with NCES reporting rules.

Table A6.4 First-stage IV estimates for common trajectories models of reading scores

DV: Spr K (6) score	
Fall K (5) score	0.769
	[0.010]*
Fall K (5) score squared	0.009
	[0.00417]+
Constant	0.032
	[0.013]+
Joint F-test	F (2, 437) = 3763.20
R-squared	0.629
DV: Spr K (6) score squared	
Fall K (5) score	-0.026
	[0.025]
Fall K (5) score squared	0.649
	[0.016]*
Constant	0.307
	[0.015]*
Joint F-test	F(2, 437) = 1046.06
R-squared	0.494
Observations	14,240

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to nearest 10 in accordance with NCES reporting rules.

Table A6.5 Common trajectories model estimates for math scores

	OLS				IV			
	1st grade (7)	3rd grade (9)	5th grade (11)	8th grade (14)	1st grade (7)	3rd grade (9)	5th grade (11)	8th grade (14)
Y_{i6}	0.758	0.76	0.734	0.697	0.847	0.87	0.849	0.799
	[0.009]*	[0.010]*	[0.014]*	[0.015]*	[0.011]*	[0.013]*	[0.017]*	[0.018]*
Y_{i6}^2	-0.0574	0.00132	0.00844	0.00193	-0.0519	0.00793	0.00231	0.00557
	[0.006]*	[0.005]	[0.010]	[0.010]	[0.007]*	[0.009]	[0.012]	[0.014]
Constant	0.0666	-0.0128	-0.00809	-0.0299	0.0567	-0.0255	-0.0096	-0.0422
	[0.013]*	[0.015]	[0.021]	[0.021]	[0.013]~	[0.015]~	[0.022]	[0.021]+
N	14,190	12,200	9,630	8,020	14,190	12,200	9,630	8,020
R-squared	0.614	0.569	0.529	0.465	0.605	0.557	0.516	0.455
Predicted scores associated with SES group means at Spr K								
Low education ($Y_{6(L)} = -0.40$)	-0.25	-0.32	-0.30	-0.31	-0.29	-0.37	-0.35	-0.36
Med education ($Y_{6(M)} = 0.05$)	0.11	0.03	0.03	0.01	0.10	0.02	0.04	0.00
High education ($Y_{6(H)} = 0.56$)	0.47	0.41	0.40	0.36	0.51	0.46	0.46	0.40

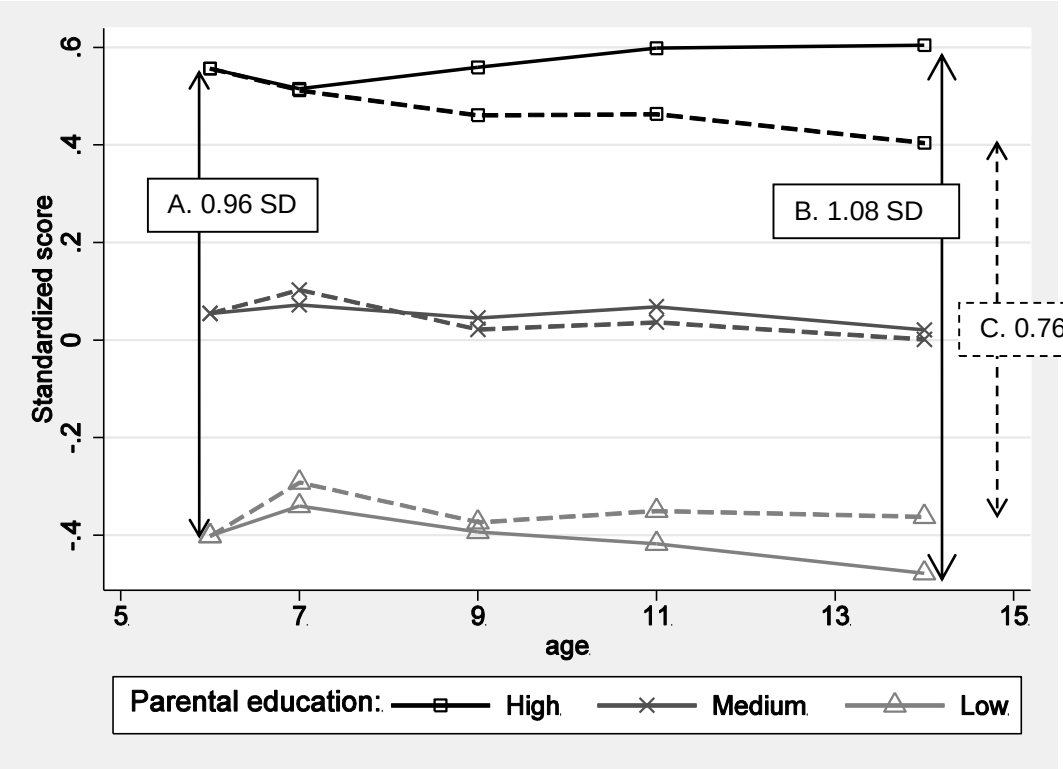
Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to the nearest 10 in accordance with NCES reporting rules.

Table A6.6 First-stage IV estimates for common trajectories models of math scores

DV: Spr K (6) score	
Fall K (5) score	0.811
	[0.00840]*
Fall K (5) score squared	-0.00281
	[0.00532]
Constant	0.0236
	[0.0113]+
Joint F-test	F(2, 437) = 5550.31
R-squared	0.672
DV: Spr K (6) score squared	
Fall K (5) score	-0.0382
	[0.0273]
Fall K (5) score squared	0.656
	[0.0177]*
Constant	0.324
	[0.0134]*
Joint F-test	F(2, 437) = 985.18
R-squared	0.437
Observations	14,190

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to the nearest 10 in accordance with NCES reporting rules.

Figure A6.2 **Mean math scores by parental education group, actual and predicted from a common trajectories model with initial Spring K scores set to education group means**



Note: Solid lines plot the average scores of children in each SES group at different ages. Distances A and B are the observed SES gaps in kindergarten and 8th grade respectively. Dashed lines trace the trajectories associated with the three average SES-group scores in kindergarten from a common trajectories model (i.e. one in which the outcome depends only on initial score and not on SES). Distance C shows the counterfactual SES gap in 8th grade predicted by the common trajectories model. Common trajectories models at ages 7, 9, 11 and 14 (1st, 3rd, 5th and 8th grades) are estimated using instrumental variables in the way described equations A6.1 and A6.2.

Section A6.3 Diverging trajectories models

To allow for diverging trajectories, the model in equation A6.1 is augmented with interaction terms that allow the intercept and the association with the linear and quadratic Spring Kindergarten test scores to differ with parental education group. In the specification below, the omitted education group is the High group. Main effects are supplemented with interactions with dummies for Low education group (L_i) and medium education group (M_i).

$$Y_{it} = \alpha_0 + \alpha_{0L}L_i + \alpha_{0M}M_i + \alpha_1 Y_{i6} + \alpha_{1L}(Y_{i6} \times L_i) + \alpha_{1M}(Y_{i6} \times M_i) + \alpha_2 Y_{i6}^2 + \alpha_{2L}\left(Y_{i6}^2 \times L_i\right) + \alpha_{2M}\left(Y_{i6}^2 \times M_i\right) + \varepsilon_{it} \qquad (A6.4)$$

There are now six first stage equations for the IV models as follows:

$$Y_{i6} = \beta_0 + \beta_1 Y_{i5} + \beta_2 Y_{i5}^2 + \eta_{it} \qquad (A6.5a)$$

$$(Y_{i6} \times L_i) = \beta_{0L} L_i + \beta_{1L}(Y_{i5} \times L_i) + \beta_{2L}\left(Y_{i5}^2 \times L_i\right) + \eta_{iLt} \qquad (A6.5b)$$

$$(Y_{i6} \times M_i) = \beta_{0M} M_i + \beta_{1M}(Y_{i5} \times M_i) + \beta_{2M}\left(Y_{i5}^2 \times M_i\right) + \eta_{iMt} \qquad (A6.5c)$$

$$Y_{i6}^2 = \gamma_0 + \gamma_1 Y_{i5} + \gamma_2 Y_{i5}^2 + \nu_{it} \qquad (A6.5d)$$

$$\left(Y_{i6}^2 \times L_i\right) = \gamma_{0L} L_i + \gamma_{1L}(Y_{i5} \times L_i) + \gamma_{2L}\left(Y_{i5}^2 \times L_i\right) + \nu_{iLt} \qquad (A6.5e)$$

$$\left(Y_{i6}^2 \times M_i\right) = \gamma_{0M} M_i + \gamma_{1M}(Y_{i5} \times M_i) + \gamma_{2M}\left(Y_{i5}^2 \times M_i\right) + \nu_{iMt} \qquad (A6.5f)$$

Output of the regression models is collected together in section A6.7. In order to illustrate the results, we calculate SES gaps for children with three specific values of the standardized Spring Kindergarten test score: High ability ($Y_{i6} = 1$); Mean ability ($Y_{i6} = 0$); and Low ability ($Y_{i6} = -1$). The SES gaps shown in the subsequent tables are therefore linear combinations of the parameters in equation A6.4 as follows.

Table A6.7 Deriving SES gaps at different values of the initial Spring Kindergarten test score

	High ability ($Y_{i6} = 1$)	Mean ability ($Y_{i6} = 0$)	Low ability ($Y_{i6} = -1$)
High-low SES gap at t	$-(\alpha_{0L} + \alpha_{1L} + \alpha_{2L})$	$-\alpha_{0L}$	$-(\alpha_{0L} - \alpha_{1L} + \alpha_{2L})$
High-medium SES gap at t	$-(\alpha_{0M} + \alpha_{1M} + \alpha_{2M})$	$-\alpha_{0M}$	$-(\alpha_{0M} - \alpha_{1M} + \alpha_{2M})$
Medium-low SES gap at t	$(\alpha_{0M} + \alpha_{1M} + \alpha_{2M}) - (\alpha_{0L} + \alpha_{1L} + \alpha_{2L})$	$\alpha_{0M} - \alpha_{0L}$	$(\alpha_{0M} - \alpha_{1M} + \alpha_{2M}) - (\alpha_{0L} - \alpha_{1L} + \alpha_{2L})$

By definition the SES gap for children with a given initial score is 0 at Spring Kindergarten. The estimates in the following tables show how scores between different SES groups are predicted to diverge over time. Gaps correspond the vertical distances between the lines for a given initial test score shown in Figures 6.5 and 6.6 of the main chapter. Table A6.8 shows results contrasting the OLS and IV estimates of the gaps for reading; Table A6.9 shows the equivalent for math.

Table A6.8 SES gaps in reading scores over time, by initial score in Spring Kindergarten – OLS and IV estimates

		OLS				IV				Ratio IV/OLS			
	Spr K (6)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
A. Spr K score: High (+1 SD)													
High-low SES gap	0	0.14*	0.40* M	0.49*	0.65*	0.07*	0.34*	0.41*	0.65* L	50%	85%	85%	101%
		(0.02)	(0.03)	(0.05)	(0.05)	(0.02)	(0.04)	(0.06)	(0.06)				
High-medium SES gap	0	0.09*	0.24*	0.29*	0.42*	0.06+	0.21*	0.27*	0.41* M	67%	85%	95%	100%
		(0.02)	(0.03)	(0.05)	(0.05)	(0.02)	(0.03)	(0.06)	(0.06)				
Medium-low SES gap	0	0.05+ L	0.15*	0.20*	0.23*	0.01	0.13*	0.14~	0.24*	22%	84%	69%	104%
		(0.02)	(0.04)	(0.06)	(0.06)	(0.02)	(0.04)	(0.07)	(0.06)				
B. Spr K score: Average (0 SD)													
High-low SES gap	0	0.15*	0.45* H	0.51*	0.69*	0.07*	0.33*	0.37*	0.57* L	42%	73%	73%	82%
		(0.02)	(0.03)	(0.05)	(0.05)	(0.02)	(0.04)	(0.05)	(0.06)				
High-medium SES gap	0	0.08*	0.25*	0.28*	0.38*	0.02	0.18*	0.19*	0.30* H	28%	72%	67%	78%
		(0.02)	(0.03)	(0.05)	(0.05)	(0.02)	(0.03)	(0.05)	(0.06)				
Medium-low SES gap	0	0.08* L	0.20*	0.23*	0.31*	0.04+	0.15*	0.19*	0.27* L	56%	74%	81%	86%
		(0.02)	(0.03)	(0.05)	(0.04)	(0.02)	(0.03)	(0.05)	(0.05)				
C. Spr K score: Low (-1 SD)													
High-low SES gap	0	0.23*	0.48*	0.52*	0.60*	0.10+	0.26*	0.31*	0.31* HM	44%	54%	60%	52%
		(0.04)	(0.05)	(0.09)	(0.08)	(0.05)	(0.08)	(0.11)	(0.10)				
High-medium SES gap	0	0.08~	0.29*	0.28*	0.37*	0.01	0.17+	0.16	0.22~	13%	58%	58%	61%
		(0.04)	(0.06)	(0.08)	(0.09)	(0.05)	(0.08)	(0.12)	(0.12)				
Medium-low SES gap	0	0.14* HM	0.19*	0.24*	0.24*	0.09*	0.09~	0.15~	0.09 M	62%	48%	63%	40%
		(0.03)	(0.04)	(0.06)	(0.05)	(0.03)	(0.05)	(0.08)	(0.07)				

Note. Standard errors in parentheses. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. Superscripts H, M and L indicate the estimate is significantly different to the gap for children with high, medium and low initial scores respectively (at the 5% level).

Table A6.9 SES gaps in math scores over time, by initial score in Spring Kindergarten – OLS and IV estimates

		OLS				IV				Ratio IV/OLS			
	Spr K (6)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
A. Spr K score: High (+1 SD)													
High-low SES gap	0	0.22* M	0.30* L	0.38* L	0.54* L	0.12*	0.19* L	0.30* L	0.44* L	56%	63%	79%	80%
		(0.03)	(0.03)	(0.05)	(0.06)	(0.03)	(0.04)	(0.06)	(0.07)				
High-medium SES gap	0	0.10*	0.21* ML	0.29* ML	0.36* L	0.05+	0.14*	0.22* L	0.29*	51%	65%	74%	80%
		(0.02)	(0.03)	(0.04)	(0.04)	(0.02)	(0.04)	(0.05)	(0.05)				
Medium-low SES gap	0	0.12* ML	0.09*	0.09+ M	0.18*	0.07+	0.05	0.09	0.15+	60%	56%	96%	81%
		(0.02)	(0.03)	(0.04)	(0.06)	(0.03)	(0.04)	(0.06)	(0.07)				
B. Spr K score: Average (0 SD)													
High-low SES gap	0	0.12* H	0.28* L	0.37* L	0.49* L	0.05+	0.18* L	0.27* L	0.41* L	43%	64%	73%	82%
		(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.03)	(0.05)	(0.05)				
High-medium SES gap	0	0.07*	0.14* HL	0.20* HL	0.29* L	0.04~	0.09*	0.13* L	0.25*	47%	64%	66%	85%
		(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.03)	(0.05)	(0.05)				
Medium-low SES gap	0	0.05* H	0.14*	0.18* H	0.21*	0.02	0.09*	0.14*	0.16*	38%	63%	80%	78%
		(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.03)	(0.05)	(0.05)				
C. Spr K score: Low (-1 SD)													
High-low SES gap	0	0.12*	0.13* HM	0.14~ HM	0.25* HM	0.00	0.00 HM	-0.03 HM	0.17+ HM	2%	0%	-25%	69%
		(0.04)	(0.04)	(0.07)	(0.07)	(0.06)	(0.06)	(0.08)	(0.08)				
High-medium SES gap	0	0.07~	0.03 HM	-0.04 HM	0.11 HM	0.00	-0.03	-0.07 HM	0.12	0%	-74%	197%	105%
		(0.04)	(0.04)	(0.08)	(0.08)	(0.06)	(0.06)	(0.10)	(0.09)				
Medium-low SES gap	0	0.05+ H	0.10*	0.18*	0.14*	0.00	0.03	0.04	0.06	6%	26%	22%	40%
		(0.02)	(0.03)	(0.05)	(0.05)	(0.03)	(0.04)	(0.06)	(0.07)				

Note. Standard errors in parentheses. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. Superscripts H, M and L indicate the estimate is significantly different to the gap for children with high, medium and low initial scores respectively (at the 5% level).

Section A6.4 School fixed effects estimates

We use school fixed effect models to test the contribution of within- and between school variation to the divergence in trajectories. This modifies the model in equations A6.4 and A6.5 by allowing for school-specific intercepts – the α_0, β_0 and $\gamma_0$parameters now take a school-specific sub-script, becoming α_{0s}, β_{0s} and γ_{0s}. School fixed effects control for all influences, both observed and unobserved, that are common to children in a particular school. Evidence of diverging trajectories conditional on school fixed effects, therefore, indicate different trajectories for children from different SES groups even after average school "quality" is equalized.

The school identifier belonging to a child relates throughout to the school attended in Fall Kindergarten. Changes in school after that time are not accounted for in the model. There are approximately 910 schools in the full kindergarten sample, with an average of about 10 pupils in each. To test for sensitivity of our results to the influence of children who change schools, we re-estimated the fixed effect models on the sub-sample of children who remained in the same school from Fall Kindergarten through 5th grade. (Since many children switch to a junior high school between 5th and 8th grades, we do not require in this sample that children be in the same school all the way to 8th grade.) This sub-sample contains approximately 730 schools, with an average of about 10 children per school. Since this is likely to be a non-random sub-set of all children, we also re-estimate the IV models without fixed effects on the more restricted sample for comparison.

Fixed effects estimates were calculated using the Stata command `xtivreg2`. This fixed effects IV estimator does not support the use of Stata's survey design command `svy`, so estimates are simply weighted using the cross-sectional weights. For comparability the IV models *without* fixed effects are also re-estimated without the `svy` command, using instead the `ivreg2` command with weights. Regression output can be found in section A6.7. Tables A6.10 and A6.11 show estimates of the SES gaps in reading from fixed effects IV models estimated on the full and non-mover restricted sub-sample respectively. Tables A6.12 and A6.13 show the equivalent estimates for math outcomes.

Table A6.10 SES gaps in reading scores over time, by initial score in Spring Kindergarten – IV estimates with and without school fixed effects, full sample

		IV estimates without FE				IV estimates with FE				Ratio FE/non-FE estimates			
	Spr K (6)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
A. Spr K score: High (+1 SD)													
High-low SES gap	0	0.07*	0.34*	0.41*	0.65* L	-0.01	0.17*	0.24*	0.41* L	-12%	51%	58%	63%
		(0.02)	(0.04)	(0.07)	(0.06)	(0.03)	(0.04)	(0.05)	(0.06)				
High-medium SES gap	0	0.06*	0.21*	0.27*	0.41* M	0.02	0.12*	0.19* M	0.25* M	32%	57%	69%	60%
		(0.02)	(0.03)	(0.05)	(0.05)	(0.02)	(0.03)	(0.04)	(0.05)				
Medium-low SES gap	0	0.01	0.13*	0.14~	0.24*	-0.03	0.05	0.05	0.17*	-227%	43%	37%	70%
		(0.03)	(0.04)	(0.07)	(0.07)	(0.03)	(0.04)	(0.05)	(0.06)				
B. Spr K score: Average (0 SD)													
High-low SES gap	0	0.07*	0.33*	0.37*	0.57* L	0.00	0.16*	0.22*	0.32* L	-6%	49%	60%	57%
		(0.02)	(0.03)	(0.05)	(0.05)	(0.03)	(0.03)	(0.04)	(0.05)				
High-medium SES gap	0	0.02	0.18*	0.19*	0.30* H	-0.02	0.09*	0.09+ H	0.13* H	-81%	47%	48%	45%
		(0.02)	(0.03)	(0.05)	(0.05)	(0.02)	(0.03)	(0.04)	(0.04)				
Medium-low SES gap	0	0.04~	0.15*	0.19*	0.27* L	0.01	0.08+	0.13*	0.19*	30%	52%	72%	70%
		(0.03)	(0.03)	(0.05)	(0.05)	(0.03)	(0.03)	(0.04)	(0.05)				
C. Spr K score: Low (-1 SD)													
High-low SES gap	0	0.10+	0.26*	0.31*	0.31* HM	0.05	0.13~	0.18+	0.12 HM	45%	49%	56%	38%
		(0.05)	(0.07)	(0.10)	(0.10)	(0.05)	(0.07)	(0.08)	(0.09)				
High-medium SES gap	0	0.01	0.17+	0.16	0.22~	-0.02	0.08	0.04	0.07	-173%	44%	25%	33%
		(0.05)	(0.07)	(0.11)	(0.12)	(0.05)	(0.07)	(0.09)	(0.10)				
Medium-low SES gap	0	0.09+	0.09~	0.15+	0.09 M	0.06~	0.05	0.13~	0.05	72%	57%	91%	50%
		(0.04)	(0.05)	(0.07)	(0.07)	(0.03)	(0.05)	(0.07)	(0.06)				

Note. Standard errors in parentheses. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. Superscripts H, M and L indicate the estimate is significantly different to the gap for children with high, medium and low initial scores respectively (at the 5% level).

Table A6.11 SES gaps in reading scores over time, by initial score in Spring Kindergarten – IV estimates with and without school fixed effects, restricted sample of children in the same school from Fall Kindergarten to 5th grade

		IV estimates without FE				IV estimates with FE				Ratio FE/non-FE estimates			
	Spr K (6)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
A. Spr K score: High (+1 SD)													
High-low SES gap	0	0.05~	0.31*	0.35*	0.54*	0.01	0.12*	0.17*	0.24*	24%	39%	48%	43%
		(0.03)	(0.04)	(0.05)	(0.05)	(0.03)	(0.04)	(0.05)	(0.06)				
High-medium SES gap	0	0.04	0.23*	0.25*	0.36*	0.02	0.12*	0.14*	0.16*	44%	50%	53%	46%
		(0.03)	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)	(0.05)				
Medium-low SES gap	0	0.01	0.08~	0.10+	0.19*	-0.01	0.01	0.04	0.07	-60%	8%	36%	39%
		(0.03)	(0.04)	(0.05)	(0.06)	(0.03)	(0.04)	(0.05)	(0.06)				
B. Spr K score: Average (0 SD)													
High-low SES gap	0	0.10*	0.31*	0.37*	0.57*	0.04	0.13*	0.21*	0.29*	40%	43%	56%	52%
		(0.03)	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)	(0.05)				
High-medium SES gap	0	0.04~	0.19*	0.21*	0.30*	-0.01	0.07+	0.10*	0.12*	-21%	37%	46%	40%
		(0.02)	(0.03)	(0.03)	(0.04)	(0.02)	(0.03)	(0.03)	(0.04)				
Medium-low SES gap	0	0.06+	0.12*	0.15*	0.27*	0.05~	0.06~	0.11*	0.18*	86%	51%	70%	65%
		(0.03)	(0.04)	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)	(0.05)				
C. Spr K score: Low (-1 SD)													
High-low SES gap	0	0.18*	0.28*	0.34*	0.48*	0.08	0.11	0.17+	0.27*	46%	40%	49%	55%
		(0.06)	(0.07)	(0.08)	(0.09)	(0.06)	(0.08)	(0.08)	(0.09)				
High-medium SES gap	0	0.12~	0.14~	0.23*	0.31*	0.05	0.02	0.11	0.17~	44%	18%	47%	55%
		(0.07)	(0.08)	(0.09)	(0.10)	(0.07)	(0.08)	(0.09)	(0.10)				
Medium-low SES gap	0	0.06	0.15*	0.11~	0.17*	0.03	0.09	0.06	0.10	48%	61%	56%	55%
		(0.05)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)				

Note. Standard errors in parentheses. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. Superscripts H, M and L indicate the estimate is significantly different to the gap for children with high, medium and low initial scores respectively (at the 5% level).

Table A6.12 SES gaps in math scores over time, by initial score in Spring Kindergarten – IV estimates with and without school fixed effects, full sample

		IV estimates without FE				IV estimates with FE				Ratio FE/non-FE estimates			
	Spr K (6)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
A. Spr K score: High (+1 SD)													
High-low SES gap	0	0.12* M	0.19* L	0.30* L	0.44* L	0.07+	0.13*	0.24* L	0.31*	53%	68%	78%	71%
		(0.03)	(0.04)	(0.05)	(0.07)	(0.03)	(0.04)	(0.05)	(0.05)				
High-medium SES gap	0	0.05+	0.14* L	0.22* ML	0.29*	0.03	0.08+	0.16* L	0.20*	50%	57%	75%	70%
		(0.02)	(0.03)	(0.04)	(0.04)	(0.02)	(0.03)	(0.04)	(0.04)				
Medium-low SES gap	0	0.07+	0.05	0.09	0.15+	0.04	0.05	0.07	0.11+	56%	96%	87%	72%
		(0.03)	(0.04)	(0.06)	(0.07)	(0.03)	(0.04)	(0.05)	(0.05)				
B. Spr K score: Average (0 SD)													
High-low SES gap	0	0.05+ H	0.18* L	0.27* L	0.41* L	0.05~	0.12* L	0.24* L	0.30* L	88%	66%	89%	75%
		(0.02)	(0.03)	(0.05)	(0.05)	(0.02)	(0.03)	(0.04)	(0.04)				
High-medium SES gap	0	0.04~	0.09* L	0.13* HL	0.25*	0.04	0.05~	0.11* L	0.15*	100%	53%	85%	62%
		(0.02)	(0.03)	(0.05)	(0.05)	(0.02)	(0.03)	(0.04)	(0.04)				
Medium-low SES gap	0	0.02	0.09*	0.14*	0.16*	0.01	0.07+	0.13*	0.15*	65%	80%	93%	94%
		(0.02)	(0.03)	(0.05)	(0.05)	(0.02)	(0.03)	(0.04)	(0.04)				
C. Spr K score: Low (-1 SD)													
High-low SES gap	0	0.00	0.00 HM	-0.03 HM	0.17+ HM	0.05	0.00 M	0.01 HM	0.13~ M	1633%	-	-26%	73%
		(0.05)	(0.06)	(0.08)	(0.08)	(0.05)	(0.06)	(0.07)	(0.07)				
High-medium SES gap	0	0.00	-0.03 HM	-0.07 HM	0.12	0.02	-0.02	-0.06 HM	0.03	-	88%	79%	24%
		(0.06)	(0.06)	(0.09)	(0.09)	(0.05)	(0.06)	(0.07)	(0.08)				
Medium-low SES gap	0	0.00	0.03	0.04	0.06	0.03	0.03	0.07	0.10~	900%	96%	169%	174%
		(0.03)	(0.03)	(0.06)	(0.06)	(0.03)	(0.03)	(0.04)	(0.06)				

Note. Standard errors in parentheses. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. Superscripts H, M and L indicate the estimate is significantly different to the gap for children with high, medium and low initial scores respectively (at the 5% level).

Table A6.13 SES gaps in math scores over time, by initial score in Spring Kindergarten – IV estimates with and without school fixed effects, restricted sample of children in the same school from Fall Kindergarten to 5th grade

		IV estimates without FE				IV estimates with FE				Ratio FE/non-FE estimates			
	Spr K (6)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
A. Spr K score: High (+1 SD)													
High-low SES gap	0	0.03	0.11*	0.17*	0.29* M	0.01	0.05 M	0.12* M	0.19* M	16%	42%	72%	65%
		(0.03)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)				
High-medium SES gap	0	0.03	0.10*	0.17*	0.26*	0.00	0.04	0.12*	0.16*	16%	43%	70%	61%
		(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)				
Medium-low SES gap	0	0.01	0.01	0.00	0.03 M	0.00	0.00 M	0.00	0.03 M	33%	27%	-100%	97%
		(0.03)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)				
B. Spr K score: Average (0 SD)													
High-low SES gap	0	0.05~	0.20*	0.27*	0.41* H	0.03 L	0.15* H	0.22* H	0.31* H	65%	74%	79%	77%
		(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)				
High-medium SES gap	0	0.05~	0.11*	0.20*	0.25*	0.02	0.03	0.13*	0.14*	46%	31%	64%	54%
		(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)				
Medium-low SES gap	0	0.00	0.09*	0.08+	0.15* H	0.01	0.11* H	0.09*	0.18* H	-267%	127%	116%	114%
		(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)				
C. Spr K score: Low (-1 SD)													
High-low SES gap	0	0.12~	0.22*	0.23*	0.33*	0.16+ M	0.20*	0.17+	0.24*	130%	90%	72%	72%
		(0.06)	(0.07)	(0.07)	(0.08)	(0.06)	(0.07)	(0.07)	(0.08)				
High-medium SES gap	0	0.12~	0.21*	0.20+	0.23*	0.11	0.15+	0.10	0.10	92%	71%	49%	44%
		(0.07)	(0.08)	(0.08)	(0.09)	(0.07)	(0.08)	(0.08)	(0.08)				
Medium-low SES gap	0	0.00	0.01	0.03	0.10~	0.05	0.05	0.07	0.14+	1300%	556%	215%	135%
		(0.04)	(0.05)	(0.05)	(0.06)	(0.04)	(0.05)	(0.05)	(0.06)				

Note. Standard errors in parentheses. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. Superscripts H, M and L indicate the estimate is significantly different to the gap for children with high, medium and low initial scores respectively (at the 5% level).

Section A6.5 Models with controls for child behavior

One hypothesis concerning the diverging trajectories is that low SES children have systematically poorer socio-emotional skills than higher SES children with identical achievement scores, and it is this that accounts for their poorer progress during the school years. We test this hypothesis in two ways.

Firstly, we add controls for four dimensions of socio-emotional skills measured in Spring Kindergarten, at the same time as the initial achievement measure, each interacted with parental education. These scores come from teacher reports of: conduct problems, inattention problems, internalizing problems and pro-social problems. Derivations of the measures are provided in Appendix section A4.3. All scores are standardized to mean zero, unit variance, and are such that higher scores indicate adverse behavioural outcomes.

Specifically, in the first set of models equations A6.4 and A6.5 are augmented with the following terms:

$$+\sum_{d=1}^{4}\left\{\delta_d X_{i6}^d + \delta_{dL}\left(X_{i6}^d \times L_i\right) + \delta_{dM}\left(X_{i6}^d \times M_i\right)\right\}$$

Where X_{i6}^d is the ith child's score on the dth domain of behavior measured in Spring Kindergarten.

Even if low SES children don't begin with systematically poorer behavior outcomes (conditional on their achievement scores), it is possible that they develop poor behavior over time, and it is the differential trajectory in behavior that leads to the differential trajectory in achievement. To test this we estimate models what control for behavior at time t, rather than in Spring Kindergarten, i.e. 5th grade reading achievement is predicted conditional on 5th grade behavior rather than initial behavior. Since teachers did not complete behavioral inventories in 8th grade, for outcomes at that age we use the nearest (5th grade) behavior scores as predictors. The terms added to the baseline models from equations A6.4 and A6.5 in this *contemporaneous specification* are:

$$+\sum_{d=1}^{4}\left\{\delta_d X_{it}^d + \delta_{dL}\left(X_{it}^d \times L_i\right) + \delta_{dM}\left(X_{it}^d \times M_i\right)\right\}$$

Observations with missing data on any of the behavior scores are dropped from the estimation. For comparability we show estimates from models that use these same, restricted, samples, but as in the baseline do not include any behavior controls. Regression output, including the coefficients on the behavior measures, can be found in section A6.7. Tables A6.14 and A6.15 show estimates of the SES gaps in reading from models including behavior measures in Spring Kindergarten and contemporaneous with the outcome respectively. Tables A6.16 and A6.17 show the equivalent estimates for math outcomes.

Table A6.14 SES gaps in reading scores over time, by initial score in Spring Kindergarten – IV estimates with and without controls for behaviour in Spring Kindergarten

		IV estimates without controls				IV estimates with controls				Ratio with/without controls			
	Spr K (6)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
A. Spr K score: High (+1 SD)													
High-low SES gap	0	0.06+	0.32*	0.38*	0.65* L	0.06+	0.32*	0.37*	0.62* L	92%	98%	97%	96%
		(0.02)	(0.04)	(0.07)	(0.06)	(0.03)	(0.04)	(0.07)	(0.07)				
High-medium SES gap	0	0.06+	0.21*	0.32* M	0.42* M	0.05~	0.19*	0.31* M	0.41* M	81%	93%	96%	97%
		(0.02)	(0.03)	(0.06)	(0.06)	(0.03)	(0.04)	(0.07)	(0.06)				
Medium-low SES gap	0	0.00 ML	0.12*	0.06 M	0.23*	0.01 M	0.12*	0.06 M	0.21*	300%	105%	103%	94%
		(0.03)	(0.04)	(0.08)	(0.07)	(0.03)	(0.05)	(0.08)	(0.07)				
B. Spr K score: Average (0 SD)													
High-low SES gap	0	0.09*	0.35*	0.42*	0.56* L	0.09*	0.34*	0.41*	0.55* L	97%	99%	99%	99%
		(0.03)	(0.04)	(0.06)	(0.07)	(0.03)	(0.04)	(0.06)	(0.07)				
High-medium SES gap	0	0.04	0.19*	0.18* H	0.26* H	0.03	0.17*	0.17* H	0.25* H	82%	93%	95%	95%
		(0.02)	(0.04)	(0.05)	(0.06)	(0.02)	(0.04)	(0.05)	(0.06)				
Medium-low SES gap	0	0.05+ H	0.16*	0.23* H	0.30*	0.06+ H	0.17*	0.24* H	0.31*	107%	105%	102%	102%
		(0.02)	(0.03)	(0.06)	(0.06)	(0.02)	(0.03)	(0.06)	(0.06)				
C. Spr K score: Low (-1 SD)													
High-low SES gap	0	0.15*	0.31*	0.30*	0.26+ HM	0.15*	0.31*	0.31*	0.27+HM	99%	100%	102%	106%
		(0.05)	(0.07)	(0.11)	(0.12)	(0.05)	(0.07)	(0.12)	(0.12)				
High-medium SES gap	0	0.05	0.20*	0.08	0.14	0.04	0.19*	0.09	0.14	94%	98%	106%	96%
		(0.05)	(0.07)	(0.13)	(0.14)	(0.05)	(0.07)	(0.14)	(0.13)				
Medium-low SES gap	0	0.10* H	0.12+	0.22+	0.12	0.11*	0.12+	0.22+	0.14	101%	103%	100%	117%
		(0.03)	(0.06)	(0.09)	(0.09)	(0.03)	(0.06)	(0.09)	(0.09)				

Note. Standard errors in parentheses. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. Superscripts H, M and L indicate the estimate is significantly different to the gap for children with high, medium and low initial scores respectively (at the 5% level).

Table A6.15 SES gaps in reading scores over time, by initial score in Spring Kindergarten – IV estimates with and without controls for behaviour contemporaneous with the outcome measure

		IV estimates without controls				IV estimates with controls				Ratio with/without controls			
	Spr K (6)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
A. Spr K score: High (+1 SD)													
High-low SES gap	0	0.04	0.32*	0.24*	0.49*	0.03	0.30*	0.18+	0.43*	65%	94%	74%	87%
		(0.04)	(0.05)	(0.07)	(0.08)	(0.04)	(0.05)	(0.07)	(0.08)				
High-medium SES gap	0	0.06	0.19*	0.20*	0.26*	0.06	0.17*	0.18+	0.23*	102%	89%	88%	87%
		(0.04)	(0.05)	(0.07)	(0.08)	(0.05)	(0.06)	(0.07)	(0.07)				
Medium-low SES gap	0	-0.02	0.13+	0.04	0.23+	-0.04	0.14+	0.00	0.20+	164%	102%	0%	87%
		(0.04)	(0.06)	(0.10)	(0.10)	(0.04)	(0.06)	(0.10)	(0.09)				
B. Spr K score: Average (0 SD)													
High-low SES gap	0	0.09+	0.32*	0.23*	0.51*	0.08+	0.28*	0.20*	0.47*	95%	89%	87%	91%
		(0.03)	(0.05)	(0.07)	(0.07)	(0.03)	(0.05)	(0.07)	(0.07)				
High-medium SES gap	0	0.05~	0.18*	0.10	0.22* L	0.05	0.15*	0.08	0.20* L	87%	85%	82%	89%
		(0.03)	(0.04)	(0.07)	(0.08)	(0.03)	(0.04)	(0.06)	(0.08)				
Medium-low SES gap	0	0.03	0.14*	0.13	0.29* L	0.04	0.13*	0.12	0.27* L	112%	93%	92%	93%
		(0.03)	(0.05)	(0.09)	(0.07)	(0.03)	(0.05)	(0.08)	(0.07)				
C. Spr K score: Low (-1 SD)													
High-low SES gap	0	0.20+	0.28*	0.29+	0.51*	0.20+	0.25*	0.30+	0.51*	103%	88%	104%	100%
		(0.09)	(0.09)	(0.14)	(0.12)	(0.09)	(0.09)	(0.14)	(0.12)				
High-medium SES gap	0	0.14	0.31+	0.25	0.58* M	0.13	0.27+	0.19	0.52* M	88%	88%	78%	89%
		(0.10)	(0.12)	(0.17)	(0.18)	(0.10)	(0.12)	(0.16)	(0.15)				
Medium-low SES gap	0	0.05	-0.02	0.04	-0.07 M	0.08	-0.02	0.10	-0.01 M	143%	86%	274%	13%
		(0.06)	(0.11)	(0.13)	(0.14)	(0.07)	(0.10)	(0.12)	(0.12)				

Note. Standard errors in parentheses. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. Superscripts H, M and L indicate the estimate is significantly different to the gap for children with high, medium and low initial scores respectively (at the 5% level).

Table A6.16 SES gaps in math scores over time, by initial score in Spring Kindergarten – IV estimates with and without controls for behaviour in Spring Kindergarten

		IV estimates without controls				IV estimates with controls				Ratio with/without controls			
	Spr K (6)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
A. Spr K score: High (+1 SD)													
High-low SES gap	0	0.11*	0.19* L	0.30* L	0.44* L	0.13*	0.20* L	0.31* L	0.45* L	116%	107%	104%	101%
		(0.04)	(0.04)	(0.06)	(0.07)	(0.04)	(0.04)	(0.07)	(0.08)				
High-medium SES gap	0	0.05+	0.14*	0.24* ML	0.29*	0.06+	0.16*	0.27* ML	0.30*	110%	116%	114%	100%
		(0.02)	(0.04)	(0.05)	(0.05)	(0.02)	(0.05)	(0.05)	(0.05)				
Medium-low SES gap	0	0.06	0.05	0.07	0.15~	0.07~	0.04	0.04	0.15~	121%	81%	65%	102%
		(0.04)	(0.04)	(0.07)	(0.08)	(0.04)	(0.04)	(0.07)	(0.08)				
B. Spr K score: Average (0 SD)													
High-low SES gap	0	0.08*	0.18* L	0.29* L	0.38* L	0.08*	0.18* L	0.30* L	0.39* L	104%	102%	102%	101%
		(0.03)	(0.03)	(0.05)	(0.06)	(0.03)	(0.03)	(0.05)	(0.06)				
High-medium SES gap	0	0.04~	0.07+	0.11+ HL	0.21*	0.04~	0.08*	0.11+ HL	0.21*	90%	103%	102%	98%
		(0.02)	(0.03)	(0.05)	(0.05)	(0.02)	(0.03)	(0.05)	(0.05)				
Medium-low SES gap	0	0.04	0.10*	0.18*	0.17*	0.04	0.11*	0.19*	0.18*	116%	102%	102%	105%
		(0.03)	(0.03)	(0.05)	(0.05)	(0.03)	(0.03)	(0.05)	(0.05)				
C. Spr K score: Low (-1 SD)													
High-low SES gap	0	0.07	0.02 HM	-0.04 HM	0.14 HM	0.06	0.01 HM	-0.04 HM	0.14 HM	77%	30%	100%	101%
		(0.05)	(0.06)	(0.09)	(0.09)	(0.05)	(0.06)	(0.09)	(0.09)				
High-medium SES gap	0	0.05	-0.01	-0.11 HM	0.06	0.03	-0.04	-0.14 HM	0.05	63%	345%	133%	73%
		(0.05)	(0.07)	(0.11)	(0.10)	(0.05)	(0.07)	(0.12)	(0.10)				
Medium-low SES gap	0	0.03	0.03	0.07	0.08	0.03	0.04	0.11	0.10	100%	147%	149%	123%
		(0.03)	(0.04)	(0.07)	(0.08)	(0.04)	(0.04)	(0.08)	(0.08)				

Note. Standard errors in parentheses. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. Superscripts H, M and L indicate the estimate is significantly different to the gap for children with high, medium and low initial scores respectively (at the 5% level).

Table A6.17 SES gaps in math scores over time, by initial score in Spring Kindergarten – IV estimates with and without controls for behaviour contemporaneous with the outcome measure

		IV estimates without controls				IV estimates with controls				Ratio with/without controls			
	Spr K (6)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
A. Spr K score: High (+1 SD)													
High-low SES gap	0	0.04	0.08	0.09	0.34*	0.07	0.08	0.07	0.30*	169%	101%	76%	88%
		(0.05)	(0.05)	(0.07)	(0.07)	(0.05)	(0.05)	(0.07)	(0.07)				
High-medium SES gap	0	0.04	0.07	0.01	0.19*	0.05	0.06	-0.03	0.14+	126%	98%	-215%	73%
		(0.03)	(0.05)	(0.06)	(0.06)	(0.04)	(0.05)	(0.06)	(0.07)				
Medium-low SES gap	0	0.00	0.01	0.07	0.15~	0.02	0.02	0.09	0.16~	1800%	121%	127%	106%
		(0.05)	(0.05)	(0.09)	(0.09)	(0.05)	(0.05)	(0.09)	(0.09)				
B. Spr K score: Average (0 SD)													
High-low SES gap	0	0.08+	0.17*	0.18+	0.33*	0.09+	0.17*	0.17*	0.31*	109%	102%	97%	95%
		(0.04)	(0.05)	(0.07)	(0.06)	(0.04)	(0.05)	(0.06)	(0.06)				
High-medium SES gap	0	0.08+	0.10+	-0.04	0.13~	0.07+	0.09+	-0.06	0.10~	99%	95%	133%	81%
		(0.03)	(0.04)	(0.09)	(0.07)	(0.03)	(0.04)	(0.08)	(0.06)				
Medium-low SES gap	0	0.00	0.07	0.22*	0.20*	0.01	0.08	0.23*	0.21*	300%	110%	105%	104%
		(0.04)	(0.06)	(0.08)	(0.07)	(0.04)	(0.05)	(0.08)	(0.07)				
C. Spr K score: Low (-1 SD)													
High-low SES gap	0	0.11~	0.23+	0.19	0.28*	0.09	0.22+	0.21~	0.32*	80%	96%	107%	112%
		(0.07)	(0.09)	(0.13)	(0.10)	(0.06)	(0.09)	(0.12)	(0.08)				
High-medium SES gap	0	0.14+	0.20+	0.20	0.23+	0.12~	0.18+	0.20	0.24+	85%	91%	97%	104%
		(0.06)	(0.09)	(0.15)	(0.11)	(0.06)	(0.09)	(0.14)	(0.11)				
Medium-low SES gap	0	-0.02	0.03	-0.01	0.06	-0.03	0.03	0.01	0.08	119%	131%	-64%	144%
		(0.06)	(0.06)	(0.10)	(0.09)	(0.06)	(0.06)	(0.10)	(0.09)				

Note. Standard errors in parentheses. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. Superscripts H, M and L indicate the estimate is significantly different to the gap for children with high, medium and low initial scores respectively (at the 5% level).

Section A6.6 Results for other countries

A6.6i. United Kingdom

Table A6.18 summarizes the outcome measures used for the trajectories analysis in the UK. The estimation sample is an unbalanced panel, which uses all available observations at ages 7 and 11, with the restriction that a child has valid test scores at both ages 3 and 5. Estimates are weighted using the relevant weights for each wave (DOVWT2 at age 7 and EOVWT2 at age 11), and survey design features are accounted for using Stata's svy command.

Table A6.18 Summary of outcome measures used in the UK trajectories analysis

Age	Outcome measure	N	Parental education group mean		
			Low	Medium	High
3	BAS Naming Vocabulary (instrument)	13,022	-0.32	0.08	0.34
5	BAS Naming Vocabulary (initial achievement)	13,022	-0.35	0.05	0.41
7	BAS Word Reading	11,263	-0.33	0.02	0.45
11	BAS Verbal Similarities	10,717	-0.29	0.05	0.40

Figure A6.3 plots the SES-group mean scores at the three analysis time points. The age 5 measure is used as the initial achievement measure (the age 3 measure provides the instrument).

Figure A6.3 Mean UK language/reading scores, by parental education group

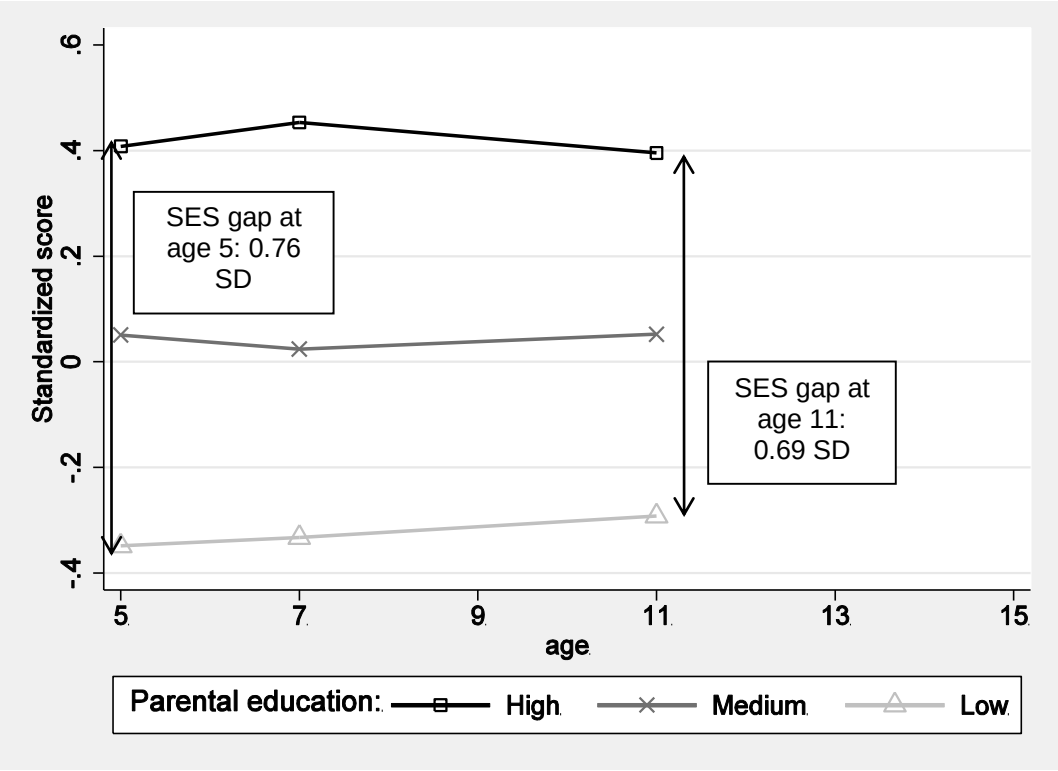
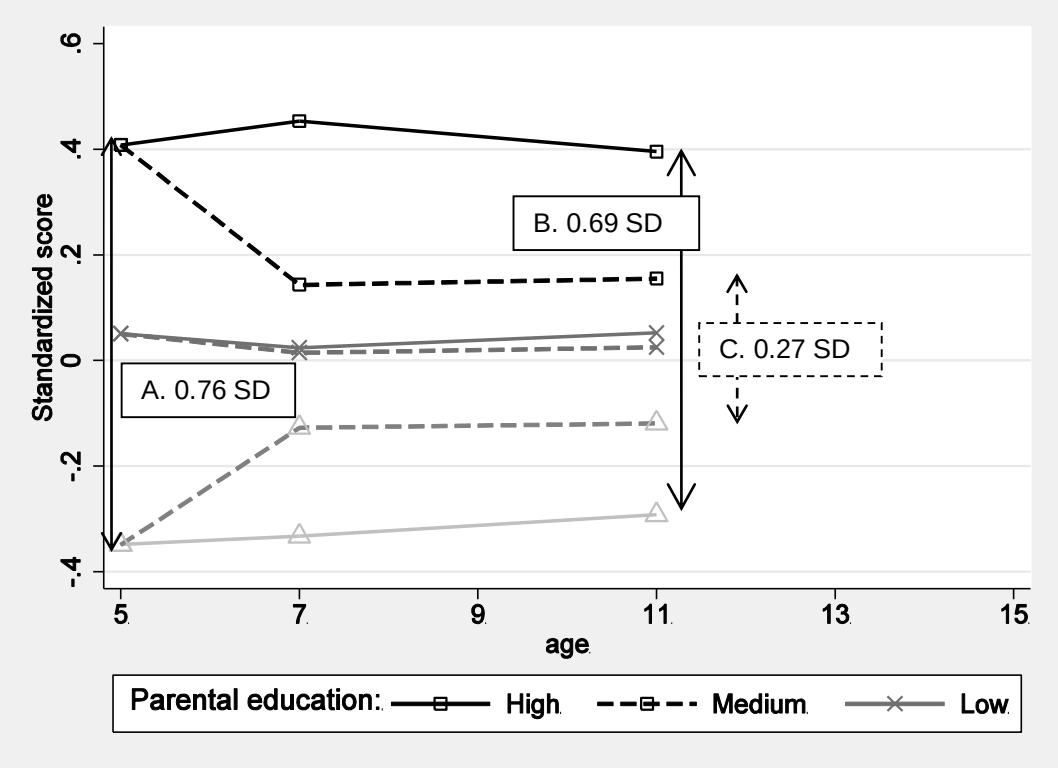


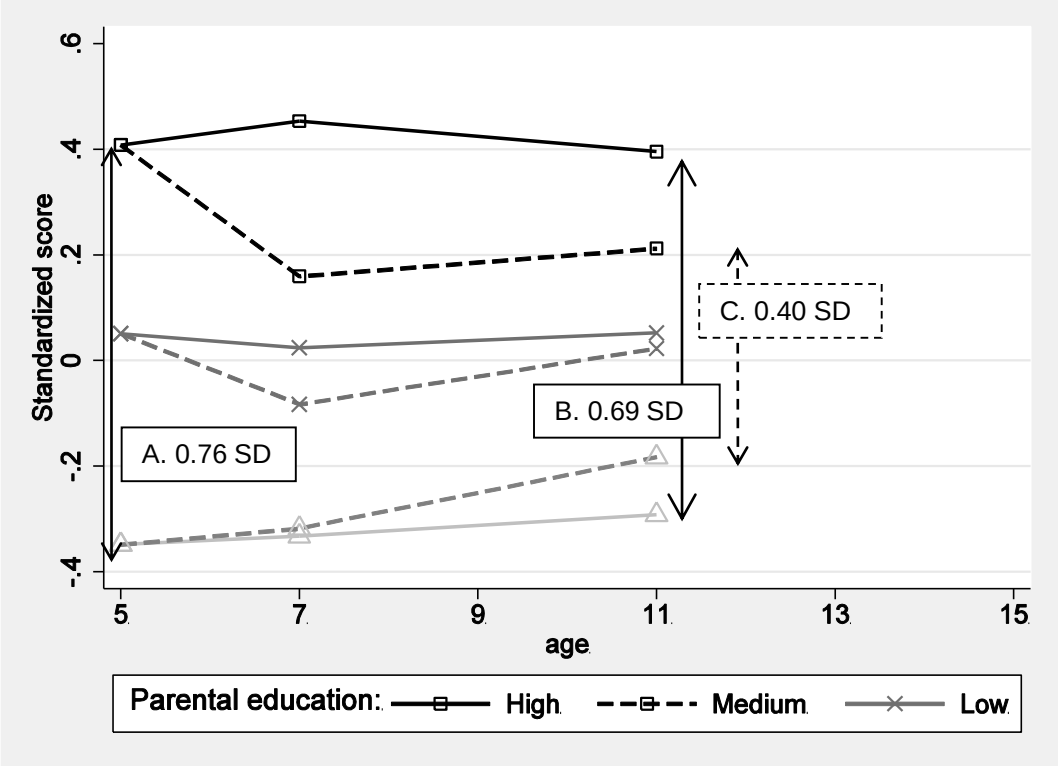
Figure A6.4 plots predictions from the counterfactual common trajectories analysis (see equations A6.1 to A6.3), first for the OLS estimates in panel a, and then for the IV estimates in panel b. Details of the underlying regression output and predicted scores are provided in Tables A6.19 and A6.20. The source of the results reported in Figure 6.7 of the main chapter are the age 11 IV results.

Figure A6.4 Mean UK language reading scores by parental education group, actual and predicted from a common trajectories model with initial age 5 scores set to education group means

a. OLS estimates



b. IV estimates



Note: Solid lines plot the average scores of children in each SES group at different ages. Distances A and B are the observed SES gaps at ages 5 and 11 respectively. Dashed lines trace the trajectories associated with the three average SES-group scores at age 5 from a common trajectories model (i.e. one in which the outcome depends only on initial score and not on SES). Distance C shows the counterfactual SES gap at age 11 predicted by the common trajectories model.

Table A6.19 Common trajectories model estimates for UK language/reading scores

	OLS		IV	
	Age 7	Age 11	Age 7	Age 11
Y_{i5}	0.357	0.362	0.624	0.521
	[0.0131]*	[0.0147]*	[0.0262]*	[0.0266]*
Y_{i5}^2	0.00538	0.00434	0.12	0.0196
	[0.00750]	[0.00983]	[0.0271]*	[0.0244]
Constant	-0.00312	0.00689	-0.115	-0.00373
	[0.0196]	[0.0233]	[0.0321]*	[0.0320]
N	11,263	10,717	11,263	10,717
R-squared	0.125	0.132	0.031	0.108
Predicted scores associated with SES group means at age 5				
Low education ($Y_{5(L)} = -0.35$)	-0.13	-0.12	-0.32	-0.18
Med education ($Y_{5(M)} = 0.05$)	0.01	0.03	-0.08	0.02
High education ($Y_{5(H)} = 0.41$)	0.14	0.16	0.16	0.21

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively.

Table A6.20 First-stage IV estimates for common trajectories models of UK language/reading scores

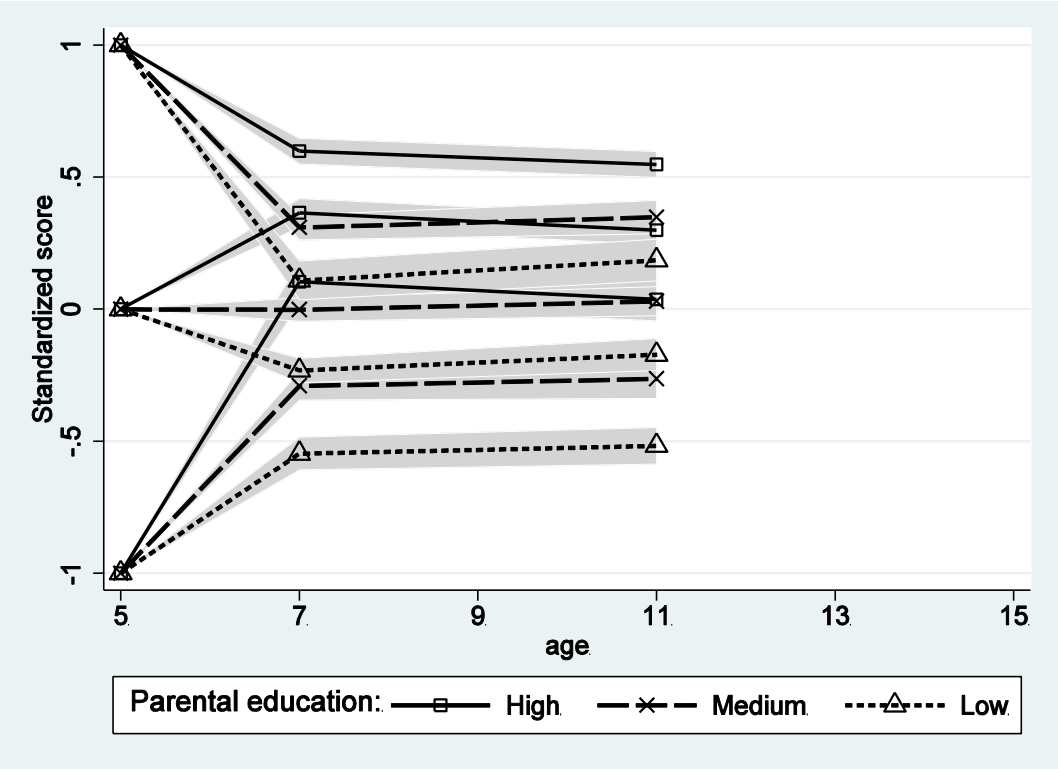
DV: Age 5 score	
Age 3 score	0.544
	[0.0128]*
Age 3 score squared	-0.0125
	[0.0063]+
Constant	0.0150
	[0.0160]
Joint F-test	F(2, 388) = 914.31
R-squared	0.308
DV: Age 5 score squared	
Age 3 score	-0.176
	[0.036]*
Age 3 score squared	0.300
	[0.023]*
Constant	0.688
	[0.026]*
Joint F-test	F(2, 388) = 83.05
R-squared	0.092
Observations	11,263

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively.

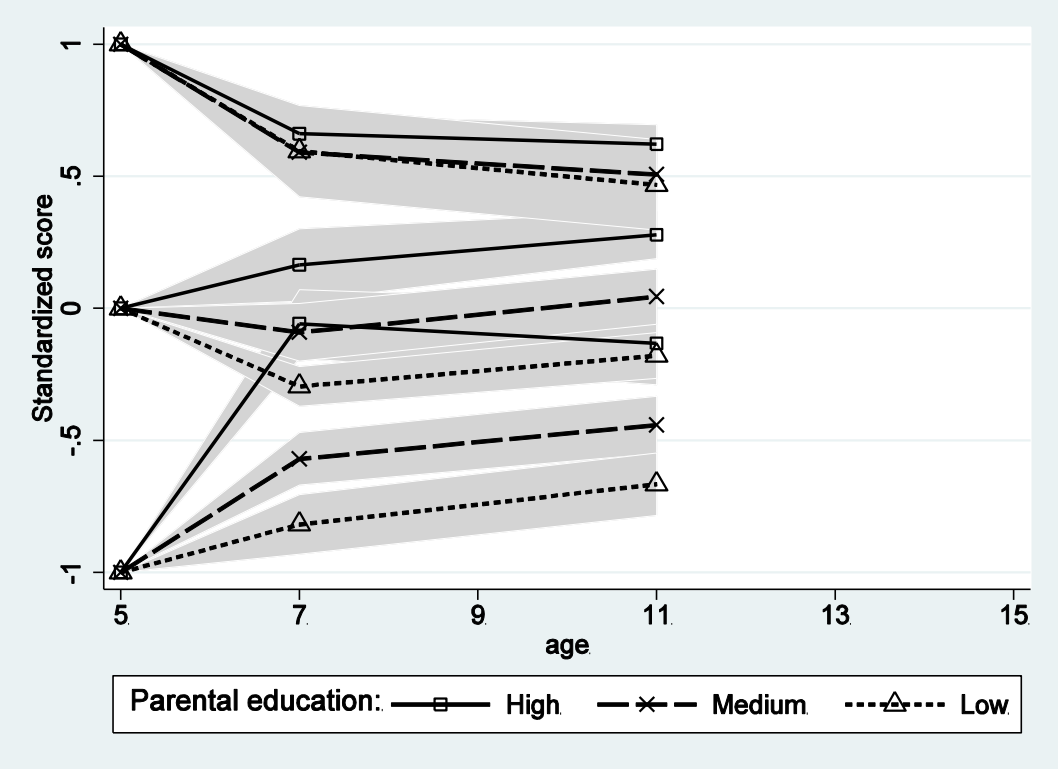
Figure A6.5 provides the results from an analysis of diverging trajectories (where initial age 5 achievement measures are interacted with SES, see equations A6.4 and A6.5). OLS results are shown in panel a, IV results in panel b. Estimates of the SES gaps at ages 7 and 11 (the vertical distances between the lines in Figure A6.5 are provided in Table A6.21, and the regression output in Tables A6.22 and A6.23.

Figure A6.5 Predicted UK language/reading scores at ages 7 and 11, for children with different combinations of initial achievement at age 5 and parental education

a. OLS estimates



b. IV estimates



Note. Lines depict the predicted scores at ages 7 and 11 of children with three specific reading test scores at age 5 (+1, 0, and -1 standard deviations above the mean respectively). The predicted scores associated with a given initial score were allowed to differ with SES. A quadratic relationship between age 5 and later test score was calculated separately for each group (with age 3 scores used as instruments to correct for measurement error in the IV models), and predictions were generated from these models. Shaded areas are 95% confidence intervals that indicate the precision with which we can predict later outcomes.

Table A6.21 SES gaps in UK language/reading scores over time, by initial score at age 5 – OLS and IV estimates

		OLS		IV		Ratio IV/OLS	
	Age 5	Age 7	Age 11	Age 7	Age 11	Age 7	Age 11
A. Spr K score: High (+1 SD)							
High-low SES gap	0	0.49* ML	0.36* ML	0.07 ML	0.15~ ML	13%	43%
		(0.04)	(0.04)	(0.09)	(0.09)		
High-medium SES gap	0	0.29* ML	0.20* M	0.07 L	0.11+	24%	57%
		(0.03)	(0.03)	(0.06)	(0.05)		
Medium-low SES gap	0	0.20*	0.16*	-0.01	0.04	-2%	24%
		(0.04)	(0.04)	(0.10)	(0.09)		
B. Spr K score: Average (0 SD)							
High-low SES gap	0	0.37* H	0.27* H	0.46* HL	0.46* H	126%	170%
		(0.03)	(0.03)	(0.08)	(0.06)		
High-medium SES gap	0	0.60* H	0.47* HL	0.26* L	0.23*	43%	50%
		(0.03)	(0.04)	(0.09)	(0.07)		
Medium-low SES gap	0	0.23*	0.20*	0.21*	0.22*	89%	110%
		(0.03)	(0.03)	(0.06)	(0.06)		
C. Spr K score: Low (-1 SD)							
High-low SES gap	0	0.65* H	0.55* HM	0.76* HM	0.53* H	117%	96%
		(0.05)	(0.05)	(0.09)	(0.10)		
High-medium SES gap	0	0.39* H	0.30*	0.51* HM	0.31*	130%	103%
		(0.05)	(0.05)	(0.08)	(0.09)		
Medium-low SES gap	0	0.26*	0.25*	0.25*	0.23*	97%	89%
		(0.04)	(0.04)	(0.07)	(0.07)		

Note. Standard errors in parentheses. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. Superscripts H, M and L indicate the estimate is significantly different to the gap for children with high, medium and low initial scores respectively (at the 5% level).

Table A6.22 Regression results: UK language/reading scores at ages 7 and 11 as a function of achievement at age 5 and SES

	OLS		IV	
	Age 7	Age 11	Age 7	Age 11
Constant	0.365	0.299	0.165	0.278
	[0.0280]*	[0.0273]*	[0.0698]+	[0.0460]*
Low parent education	-0.597	-0.471	-0.46	-0.457
	[0.0310]*	[0.0374]*	[0.0747]*	[0.0615]*
Medium parent education	-0.366	-0.269	-0.255	-0.234
	[0.0330]*	[0.0311]*	[0.0930]*	[0.0654]*
Age 5 score	0.247	0.255	0.36	0.377
	[0.0202]*	[0.0212]*	[0.0415]*	[0.0525]*
Age 5 score * Low ed	0.0803	0.0961	0.347	0.19
	[0.0317]+	[0.0297]*	[0.0809]*	[0.0864]+
Age 5 score * Medium ed	0.0526	0.0507	0.221	0.0971
	[0.0263]+	[0.0287]~	[0.0588]*	[0.0623]
Age 5 score squared	-0.0137	-0.00642	0.137	-0.0338
	[0.0103]	[0.00998]	[0.0613]+	[0.0526]
Age 5 score squared * Low ed	0.0267	0.0128	0.0472	0.113
	[0.0128]+	[0.0186]	[0.0743]	[0.0641]~
Age 5 score squared * Med ed	0.025	0.019	-0.0359	0.0223
	[0.0155]	[0.0190]	[0.0967]	[0.0721]
Observations	11,263	10,717	11,263	10,717
R-squared	0.17	0.161	0.064	0.135

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively.

Table A6.23 First-stage IV results for UK language/reading scores: Instrumenting age 5 scores (and their square) with age 3 scores (and their square), by parental education group

	Low education	Medium education	High education
DV: Age 5 score			
Age 3 score	0.528	0.477	0.486
	[0.0206]*	[0.0167]*	[0.0207]*
Age 3 score squared	-0.0167	0.00642	-0.0174
	[0.0117]	[0.0110]	[0.0112]
Constant	-0.156	0.0162	0.269
	[0.0219]*	[0.0186]	[0.0270]*
Joint F-test	F(2, 374) = 397.92	F(2, 385) = 468.21	F(2, 364) = 300.58
R-squared	0.313	0.241	0.227
DV: Age 5 score squared			
Age 3 score	-0.412	-0.0878	0.132
	[0.0551]*	[0.0374]+	[0.0494]*
Age 3 score squared	0.299	0.217	0.214
	[0.0366]*	[0.0297]*	[0.0345]*
Constant	0.659	0.637	0.772
	[0.0436]*	[0.0348]*	[0.0465]*
Joint F-test	F(2, 374) = 52.72	F(2, 385) = 26.72	F(2, 364) = 29.19
R-squared	0.149	0.05	0.046
Observations	3,976	4,246	3,041

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively.

A6.6ii. Australia

Table A6.24 summarizes the outcome measures used for the trajectories analysis in the Australia. The estimation sample is an unbalanced panel, which uses all available observations at ages 11 and 13, with the restriction that a child has valid test scores at both ages 5 and 7. Estimates are not weighted due to lack of weights designed specifically for use with the NAPLAN test outcomes used at ages 11 and 13. The age 7 measure is used as the initial achievement measure (the age 5 PPVT vocabulary score, not shown, provides the instrument).

Table A6.24 Summary of outcome measures used in the Australia trajectories analysis

Age	Outcome measure	N	Parental education group mean		
			Low	Medium	High
7	PPVT vocabulary (initial achievement)	3333	-0.27	-0.08	0.29
11	NAPLAN reading score	3333	-0.34	-0.08	0.35
13	NAPLAN reading score	3026	-0.37	-0.08	0.37

Details of the counterfactual common trajectories analysis (see equations A6.1 to A6.3) are provided in Table A6.25. The source of the results reported in Figure 6.7 of the main chapter are the age 11 IV results.

Table A6.25 Common trajectories model estimates for Australian language/reading scores

	IV	
	Age 11	Age 13
Y_{i7}	0.803	0.856
	[0.0367]*	[0.0363]*
Y_{i7}^2	0.0478	0.0660
	[0.0793]	[0.0795]
Constant	-0.0478	-0.0660
	[0.0855]	[0.0849]
N	3333	3026
Predicted scores associated with SES group means at age 7		
Low education ($Y_{7(L)} = -0.27$)	-0.26	-0.29
Med education ($Y_{7(M)} = -0.08$)	-0.11	-0.15
High education ($Y_{7(H)} = 0.29$)	0.19	0.19

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively.

Section A6.7 US regression output

This section collects together all the regression output that underlies the diverging trajectories estimates presented in the previous sections.

Table A6.26 Regression results: Reading scores between 1st and 8th grade as a function of achievement at in Spring Kindergarten and SES

	OLS				IV			
	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
Constant	0.141	0.297	0.337	0.365	0.114	0.25	0.283	0.321
	[0.0161]*	[0.0202]*	[0.0313]*	[0.0358]*	[0.0160]*	[0.0224]*	[0.0346]*	[0.0393]*
Low parent education	-0.154	-0.451	-0.508	-0.688	-0.0648	-0.33	-0.373	-0.565
	[0.0208]*	[0.0287]*	[0.0454]*	[0.0464]*	[0.0235]*	[0.0358]*	[0.0502]*	[0.0585]*
Medium parent education	-0.0761	-0.252	-0.279	-0.379	-0.0206	-0.182	-0.186	-0.297
	[0.0200]*	[0.0297]*	[0.0466]*	[0.0496]*	[0.0202]	[0.0329]*	[0.0495]*	[0.0567]*
Spr K (6) score	0.72	0.599	0.588	0.519	0.848	0.801	0.777	0.736
	[0.0211]*	[0.0280]*	[0.0504]*	[0.0411]*	[0.0250]*	[0.0415]*	[0.0610]*	[0.0617]*
Spr K score * Low ed	0.0455	0.0414	0.0159	-0.0235	0.0157	-0.0379	-0.0501	-0.17
	[0.0256]~	[0.0309]	[0.0572]	[0.0461]	[0.0303]	[0.0453]	[0.0688]	[0.0636]*
Spr K score * Medium ed	-0.00157	0.0252	-0.00251	-0.0257	-0.0232	-0.0186	-0.0551	-0.0968
	[0.0253]	[0.0339]	[0.0520]	[0.0490]	[0.0315]	[0.0471]	[0.0765]	[0.0779]
Spr K (6) score squared	-0.0327	-0.0514	-0.0462	-0.0339	-0.0713	-0.11	-0.0929	-0.102
	[0.00946]*	[0.0126]*	[0.0240]*	[0.0184]~	[0.0118]*	[0.0196]*	[0.0310]*	[0.0302]*
Spr K score squared * Low ed	-0.0264	0.0132	0.00593	0.0651	-0.0192	0.0323	0.0121	0.0804
	[0.0151]~	[0.0174]	[0.0308]	[0.0234]*	[0.0171]	[0.0287]	[0.0437]	[0.0397]+
Spr K score squared * Med ed	-0.00716	-0.0166	-0.00582	-0.0121	-0.0136	-0.00738	-0.0319	-0.0204
	[0.0137]	[0.0188]	[0.0278]	[0.0252]	[0.0162]	[0.0281]	[0.0419]	[0.0497]
Observations	14,240	12,160	9,650	7,960	14,240	12,160	9,650	7,960
R-squared	0.605	0.468	0.461	0.387	0.593	0.448	0.442	0.369

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to the nearest 10 in accordance with NCES reporting rules.

Table A6.27 First-stage IV results for reading scores: Instrumenting Spring Kindergarten scores (and their square) with Fall Kindergarten scores (and their square), by parental education group

	Low education	Medium education	High education
DV: Spr K (6) score			
Fall K (5) score	0.788	0.747	0.677
	[0.0167]*	[0.0181]*	[0.0157]*
Fall K (5) score squared	0.0166	0.00616	0.0378
	[0.0105]	[0.00881]	[0.00648]*
Constant	-0.0172	0.0473	0.0952
	[0.0182]	[0.0159]*	[0.0152]*
Joint F-test	F(2, 374) = 1129.02	F(2, 410) = 1284.19	F(2, 366) = 2543.55
R-squared	0.524	0.568	0.651
DV: Spr K (6) score squared			
Fall K (5) score	-0.0529	-0.0321	0.109
	[0.0396]	[0.0409]	[0.0311]*
Fall K (5) score squared	0.637	0.63	0.621
	[0.0341]*	[0.0329]*	[0.0200]*
Constant	0.362	0.303	0.221
	[0.0311]*	[0.0212]*	[0.0190]*
Joint F-test	F(2, 374) = 175.80	F(2, 410) = 227.45	F(2, 366) = 838.06
R-squared	0.323	0.44	0.66
Observations	4,860	4,680	4,700

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to the nearest 10 in accordance with NCES reporting rules.

Table A6.28 Regression results: Math scores between 1st and 8th grade as a function of achievement at in Spring Kindergarten and SES

	OLS				IV			
	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
Constant	0.138	0.139	0.196	0.241	0.0895	0.0706	0.135	0.186
	[0.0160]*	[0.0193]*	[0.0312]*	[0.0294]*	[0.0161]*	[0.0194]*	[0.0347]*	[0.0327]*
Low parent education	-0.12	-0.278	-0.374	-0.494	-0.0513	-0.178	-0.272	-0.405
	[0.0231]*	[0.0243]*	[0.0372]*	[0.0444]*	[0.0241]+	[0.0294]*	[0.0462]*	[0.0504]*
Medium parent education	-0.0742	-0.14	-0.195	-0.289	-0.0347	-0.0902	-0.128	-0.245
	[0.0186]*	[0.0233]*	[0.0430]*	[0.0410]*	[0.0196]~	[0.0261]*	[0.0494]*	[0.0469]*
Spr K (6) score	0.739	0.799	0.816	0.747	0.861	0.93	0.962	0.834
	[0.0224]*	[0.0205]*	[0.0362]*	[0.0353]*	[0.0314]*	[0.0335]*	[0.0432]*	[0.0451]*
Spr K score * Low ed	-0.048	-0.082	-0.122	-0.145	-0.0594	-0.0932	-0.168	-0.131
	[0.0280]~	[0.0257]*	[0.0466]*	[0.0460]*	[0.0383]	[0.0402]+	[0.0561]*	[0.0583]+
Spr K score * Medium ed	-0.0131	-0.0867	-0.165	-0.126	-0.0251	-0.0808	-0.145	-0.0874
	[0.0247]	[0.0268]*	[0.0463]*	[0.0466]*	[0.0351]	[0.0449]~	[0.0589]+	[0.0572]
Spr K (6) score squared	-0.0419	-0.0424	-0.0755	-0.0721	-0.0611	-0.0496	-0.0988	-0.0691
	[0.0117]*	[0.0120]*	[0.0166]*	[0.0177]*	[0.0159]*	[0.0182]*	[0.0236]*	[0.0277]+
Spr K score squared * Low ed	-0.0516	0.0629	0.115	0.0959	-0.0109	0.0842	0.138	0.1
	[0.0167]*	[0.0165]*	[0.0212]*	[0.0279]*	[0.0228]	[0.0256]*	[0.0372]*	[0.0412]+
Spr K score squared * Med ed	-0.0113	0.0189	0.067	0.0528	0.00995	0.0348	0.0557	0.0409
	[0.0141]	[0.0146]	[0.0238]*	[0.0283]~	[0.0210]	[0.0265]	[0.0404]	[0.0403]
Observations	14,190	12,200	9,630	8,020	14,190	12,200	9,630	8,020
R-squared	0.619	0.58	0.548	0.498	0.609	0.568	0.535	0.49

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to the nearest 10 in accordance with NCES reporting rules.

Table A6.29 First-stage IV results for math scores: Instrumenting Spring Kindergarten scores (and their square) with Fall Kindergarten scores (and their square), by parental education group

	Low education	Medium education	High education
DV: Spr K (6) score			
Fall K (5) score	0.812	0.782	0.771
	[0.0133]***	[0.0166]***	[0.0185]***
Fall K (5) score squared	0.00589	0.000352	-0.00083
	[0.0104]	[0.0111]	[0.00909]
Constant	-0.0366	0.0265	0.107
	[0.0153]**	[0.0160]*	[0.0128]***
Joint F-test	F(2, 375) = 2222.27	F(2, 411) = 1644.03	F(2, 366) = 1927.89
R-squared	0.598	0.613	0.644
DV: Spr K (6) score squared			
Fall K (5) score	-0.122	-0.029	0.119
	[0.0426]***	[0.0500]	[0.0469]**
Fall K (5) score squared	0.629	0.653	0.595
	[0.0343]***	[0.0438]***	[0.0307]***
Constant	0.353	0.272	0.316
	[0.0263]***	[0.0199]***	[0.0257]***
Joint F-test	F(2, 375) = 190.65	F(2, 411) = 212.85	F(2, 366) = 486.04
Observations	4,940	4,610	4,640
R-squared	0.393	0.363	0.538

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to the nearest 10 in accordance with NCES reporting rules.

Table A6.30 Regression results: IV estimates for reading scores with and without school fixed effects – full sample

	IV estimates, without FE				IV estimates with FE			
	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
Constant	0.114	0.25	0.283	0.321	-	-	-	-
	[0.0132]*	[0.0183]*	[0.0327]*	[0.0350]*				
Low parent education	-0.0648	-0.33	-0.373	-0.565	0.00416	-0.163	-0.223	-0.32
	[0.0245]*	[0.0316]*	[0.0514]*	[0.0536]*	[0.0257]	[0.0329]*	[0.0413]*	[0.0480]*
Medium parent education	-0.0206	-0.182	-0.186	-0.297	0.017	-0.0862	-0.0885	-0.133
	[0.0194]	[0.0281]*	[0.0451]*	[0.0494]*	[0.0198]	[0.0281]*	[0.0370]+	[0.0427]*
Spr K (6) score	0.848	0.801	0.777	0.736	0.867	0.83	0.8	0.719
	[0.0266]*	[0.0377]*	[0.0579]*	[0.0592]*	[0.0268]*	[0.0389]*	[0.0479]*	[0.0539]*
Spr K score * Low ed	0.0157	-0.0379	-0.0501	-0.17	0.0263	-0.0226	-0.0326	-0.147
	[0.0319]	[0.0437]	[0.0702]	[0.0676]+	[0.0313]	[0.0439]	[0.0554]	[0.0599]+
Spr K score * Medium ed	-0.0232	-0.0186	-0.0551	-0.0968	-0.0183	-0.021	-0.074	-0.0873
	[0.0319]	[0.0471]	[0.0724]	[0.0751]	[0.0319]	[0.0475]	[0.0609]	[0.0659]
Spr K (6) score squared	-0.0713	-0.11	-0.0929	-0.102	-0.0722	-0.118	-0.109	-0.0954
	[0.0117]*	[0.0176]*	[0.0310]*	[0.0293]*	[0.0120]*	[0.0179]*	[0.0250]*	[0.0257]*
Spr K score squared * Low ed	-0.0192	0.0323	0.0121	0.0804	-0.0229	0.0144	0.0153	0.054
	[0.0208]	[0.0265]	[0.0444]	[0.0397]+	[0.0219]	[0.0276]	[0.0349]	[0.0342]
Spr K score squared * Med ed	-0.0136	-0.00738	-0.0319	-0.0204	-0.0161	-0.0105	-0.0261	-0.0263
	[0.0164]	[0.0274]	[0.0406]	[0.0476]	[0.0172]	[0.0275]	[0.0355]	[0.0396]
Observations	14,240	12,160	9,650	7,960	14,220	12,150	9,630	7,920
R-squared	0.593	0.448	0.442	0.369	0.544	0.388	0.36	0.259
Number of schools					908	900	875	864

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to the nearest 10 in accordance with NCES reporting rules.

Table A6.31 Regression results: IV estimates for math scores with and without school fixed effects – full sample

	IV estimates, without FE				IV estimates with FE			
	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
Constant	0.0895	0.0706	0.135	0.186	-	-	-	-
	[0.0143]*	[0.0181]*	[0.0319]*	[0.0330]*				
Low parent education	-0.0513	-0.178	-0.272	-0.405	-0.0454	-0.117	-0.243	-0.302
	[0.0233]+	[0.0281]*	[0.0466]*	[0.0477]*	[0.0242]~	[0.0286]*	[0.0384]*	[0.0428]*
Medium parent education	-0.0347	-0.0902	-0.128	-0.245	-0.0346	-0.0479	-0.109	-0.151
	[0.0210]~	[0.0261]*	[0.0470]*	[0.0474]*	[0.0214]	[0.0266]~	[0.0384]*	[0.0407]*
Spr K (6) score	0.861	0.93	0.962	0.834	0.854	0.92	0.928	0.817
	[0.0301]*	[0.0317]*	[0.0432]*	[0.0419]*	[0.0266]*	[0.0311]*	[0.0386]*	[0.0388]*
Spr K score * Low ed	-0.0594	-0.0932	-0.168	-0.131	-0.00788	-0.0617	-0.114	-0.0905
	[0.0356]~	[0.0384]+	[0.0536]*	[0.0586]+	[0.0319]	[0.0376]	[0.0460]+	[0.0476]~
Spr K score * Medium ed	-0.0251	-0.0808	-0.145	-0.0874	-0.0015	-0.0502	-0.11	-0.0874
	[0.0337]	[0.0383]+	[0.0562]*	[0.0552]	[0.0304]	[0.0371]	[0.0450]+	[0.0500]~
Spr K (6) score squared	-0.0611	-0.0496	-0.0988	-0.0691	-0.061	-0.0274	-0.0879	-0.0551
	[0.0160]*	[0.0182]*	[0.0251]*	[0.0250]*	[0.0143]*	[0.0177]	[0.0258]*	[0.0259]+
Spr K score squared * Low ed	-0.0109	0.0842	0.138	0.1	-0.0114	0.0531	0.121	0.0845
	[0.0229]	[0.0255]*	[0.0376]*	[0.0384]*	[0.0206]	[0.0248]+	[0.0330]*	[0.0338]+
Spr K score squared * Med ed	0.00995	0.0348	0.0557	0.0409	0.0115	0.0197	0.0563	0.0352
	[0.0211]	[0.0259]	[0.0407]	[0.0423]	[0.0203]	[0.0256]	[0.0346]	[0.0391]
Observations	14,190	12,200	9,630	8,020	14,180	12,190	9,610	7,990
R-squared	0.609	0.568	0.535	0.49	0.559	0.523	0.486	0.418
Number of schools					908	900	877	863

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to the nearest 10 in accordance with NCES reporting rules.

Table A6.32 Regression results: IV estimates for reading scores with and without school fixed effects – restricted sample of children in the same school between Fall Kindergarten and 8th grade

	IV estimates, without FE				IV estimates with FE			
	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
Constant	0.152	0.273	0.308	0.357	-	-	-	-
	[0.0154]*	[0.0199]*	[0.0211]*	[0.0264]*				
Low parent education	-0.0984	-0.307	-0.365	-0.567	-0.0386	-0.131	-0.206	-0.293
	[0.0273]*	[0.0339]*	[0.0351]*	[0.0450]*	[0.0301]	[0.0370]*	[0.0382]*	[0.0451]*
Medium parent education	-0.0425	-0.187	-0.211	-0.295	0.00949	-0.0705	-0.0982	-0.117
	[0.0228]~	[0.0294]*	[0.0320]*	[0.0399]*	[0.0242]	[0.0311]+	[0.0332]*	[0.0407]*
Spr K (6) score	0.784	0.767	0.711	0.612	0.826	0.777	0.711	0.543
	[0.0364]*	[0.0395]*	[0.0442]*	[0.0525]*	[0.0355]*	[0.0420]*	[0.0473]*	[0.0507]*
Spr K score * Low ed	0.0636	-0.013	-0.00806	-0.0298	0.0345	-0.00325	-0.00198	0.016
	[0.0414]	[0.0473]	[0.0521]	[0.0618]	[0.0405]	[0.0500]	[0.0546]	[0.0593]
Spr K score * Medium ed	0.038	-0.0478	-0.0123	-0.0222	0.0172	-0.0454	-0.0139	0.00502
	[0.0446]	[0.0508]	[0.0557]	[0.0651]	[0.0434]	[0.0527]	[0.0569]	[0.0632]
Spr K (6) score squared	-0.0516	-0.102	-0.0719	-0.0523	-0.0577	-0.11	-0.082	-0.0411
	[0.0169]*	[0.0192]*	[0.0210]*	[0.0251]+	[0.0164]*	[0.0207]*	[0.0222]*	[0.0242]~
Spr K score squared * Low ed	-0.0156	0.00968	0.0218	0.0526	-0.00793	0.0138	0.039	0.041
	[0.0263]	[0.0313]	[0.0327]	[0.0403]	[0.0253]	[0.0316]	[0.0329]	[0.0371]
Spr K score squared * Med ed	-0.0362	0.00198	-0.0305	-0.0388	-0.0442	0.000827	-0.0225	-0.0498
	[0.0235]	[0.0297]	[0.0328]	[0.0399]	[0.0241]~	[0.0317]	[0.0335]	[0.0407]
Observations	6,630	6,570	6,610	5,290	6,600	6,540	6,580	5,250
R-squared	0.585	0.433	0.399	0.333	0.531	0.358	0.319	0.218
Number of schools					731	729	729	700

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to the nearest 10.

Table A6.33 Regression results: IV estimates for math scores with and without school fixed effects – restricted sample of children in the same school between Fall Kindergarten and 8th grade

	IV estimates, without FE				IV estimates with FE			
	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
Constant	0.0859	0.0896	0.169	0.19	-	-	-	-
	[0.0174]*	[0.0224]*	[0.0233]*	[0.0237]*				
Low parent education	-0.0487	-0.199	-0.274	-0.405	-0.0324	-0.148	-0.216	-0.311
	[0.0272]~	[0.0322]*	[0.0342]*	[0.0398]*	[0.0300]	[0.0341]*	[0.0348]*	[0.0408]*
Medium parent education	-0.0521	-0.109	-0.197	-0.25	-0.0241	-0.0336	-0.127	-0.135
	[0.0266]~	[0.0325]*	[0.0329]*	[0.0366]*	[0.0281]	[0.0325]	[0.0322]*	[0.0371]*
Spr K (6) score	0.811	0.809	0.8	0.743	0.794	0.8	0.813	0.716
	[0.0369]*	[0.0401]*	[0.0412]*	[0.0430]*	[0.0355]*	[0.0405]*	[0.0404]*	[0.0415]*
Spr K score * Low ed	0.0454	0.0549	0.0322	0.0199	0.077	0.077	0.0227	0.0241
	[0.0422]	[0.0470]	[0.0506]	[0.0537]	[0.0418]~	[0.0470]	[0.0471]	[0.0508]
Spr K score * Medium ed	0.0461	0.056	0.0144	-0.0152	0.052	0.0536	-0.0117	-0.0292
	[0.0434]	[0.0473]	[0.0487]	[0.0530]	[0.0426]	[0.0475]	[0.0477]	[0.0521]
Spr K (6) score squared	-0.0322	0.0149	-0.028	-0.0344	-0.0202	0.0206	-0.044	-0.0238
	[0.0220]	[0.0281]	[0.0250]	[0.0239]	[0.0206]	[0.0265]	[0.0231]~	[0.0227]
Spr K score squared * Low ed	-0.0279	0.0309	0.0728	0.0989	-0.0499	0.0238	0.0706	0.1
	[0.0291]	[0.0347]	[0.0354]+	[0.0351]*	[0.0273]~	[0.0338]	[0.0318]+	[0.0328]*
Spr K score squared * Med ed	-0.0195	-0.0486	0.0115	0.00899	-0.0314	-0.0642	0.0186	0.00581
	[0.0295]	[0.0367]	[0.0335]	[0.0363]	[0.0296]	[0.0355]*	[0.0321]	[0.0348]
Observations	6,620	6,590	6,610	5,330	6,600	6,560	6,580	5,280
R-squared	0.605	0.546	0.51	0.474	0.537	0.492	0.465	0.395
Number of schools					735	732	733	700

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to the nearest 10.

Table A6.34 Regression results: IV estimates for reading scores with and without controls for behavior in Spring Kindergarten

	IV estimates without controls				IV estimates with controls			
	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
Constant	0.128	0.262	0.28	0.306	0.119	0.249	0.274	0.297
	[0.0181]*	[0.0239]*	[0.0383]*	[0.0412]*	[0.0188]*	[0.0241]*	[0.0401]*	[0.0410]*
Low parent education	-0.0916	-0.348	-0.416	-0.559	-0.0891	-0.343	-0.41	-0.552
	[0.0267]*	[0.0355]*	[0.0558]*	[0.0664]*	[0.0265]*	[0.0347]*	[0.0562]*	[0.0685]*
Medium parent education	-0.0376	-0.187	-0.182	-0.259	-0.0313	-0.174	-0.173	-0.247
	[0.0234]	[0.0352]*	[0.0524]*	[0.0620]*	[0.0238]	[0.0350]*	[0.0526]*	[0.0593]*
Spr K (6) score	0.823	0.773	0.811	0.77	0.796	0.741	0.791	0.741
	[0.0258]*	[0.0354]*	[0.0647]*	[0.0664]*	[0.0263]*	[0.0390]*	[0.0673]*	[0.0676]*
Spr K score * Low ed	0.0452	-0.00662	-0.0401	-0.194	0.0465	-0.00258	-0.0309	-0.174
	[0.0322]	[0.0431]	[0.0724]	[0.0691]*	[0.0347]	[0.0495]	[0.0745]	[0.0740]+
Spr K score * Medium ed	-0.00549	-0.00575	-0.121	-0.139	-0.00139	-0.00057	-0.112	-0.135
	[0.0323]	[0.0457]	[0.0839]	[0.0848]	[0.0320]	[0.0493]	[0.0911]	[0.0834]
Spr K (6) score squared	-0.0615	-0.0968	-0.101	-0.113	-0.055	-0.0873	-0.0958	-0.107
	[0.0116]*	[0.0170]*	[0.0335]*	[0.0320]*	[0.0117]*	[0.0183]*	[0.0340]*	[0.0319]*
Spr K score squared * Low ed	-0.0143	0.0301	0.0732	0.108	-0.0137	0.029	0.071	0.105
	[0.0195]	[0.0297]	[0.0485]	[0.0425]+	[0.0198]	[0.0308]	[0.0485]	[0.0436]+
Spr K score squared * Med ed	-0.015	-0.0144	-0.0198	-0.0213	-0.0142	-0.0185	-0.0241	-0.0249
	[0.0171]	[0.0300]	[0.0468]	[0.0556]	[0.0168]	[0.0305]	[0.0480]	[0.0528]
Spr K Conduct problems	-	-	-	-	-0.00383	0.00638	-0.0518	-0.0707
					[0.0221]	[0.0261]	[0.0438]	[0.0540]
Spr K Conduct * Low ed	-	-	-	-	0.00818	0.00236	0.0432	0.0282
					[0.0356]	[0.0385]	[0.0579]	[0.0655]
Spr K Conduct * Med ed	-	-	-	-	0.0245	-0.0302	0.00747	-0.0353

	IV estimates without controls				IV estimates with controls			
	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
					[0.0300]	[0.0401]	[0.0665]	[0.0653]
Spr K Inattention problems	-	-	-	-	-0.0683	-0.0491	0.000789	-0.0585
					[0.0213]*	[0.0262]~	[0.0465]	[0.0519]
Spr K Inattention * Low ed	-	-	-	-	-0.00411	-0.0275	-0.00456	0.0695
					[0.0352]	[0.0401]	[0.0555]	[0.0672]
Spr K Inattention * Med ed	-	-	-	-	-0.0115	-0.0314	-0.0825	0.0194
					[0.0284]	[0.0372]	[0.0679]	[0.0725]
Spr K Internalizing problems	-	-	-	-	-0.00103	-0.00471	-0.0331	-0.00491
					[0.0126]	[0.0190]	[0.0295]	[0.0327]
Spr K Internalizing * Low ed	-	-	-	-	-0.00433	0.0143	0.0492	-0.00424
					[0.0194]	[0.0302]	[0.0416]	[0.0444]
Spr K Internalizing * Med ed	-	-	-	-	0.0127	0.0164	0.0721	0.0453
					[0.0191]	[0.0242]	[0.0413]~	[0.0471]
Spr K Prosocial problems	-	-	-	-	-0.00083	-0.0542	0.0155	0.0632
					[0.0179]	[0.0233]+	[0.0419]	[0.0393]
Spr K Prosocial * Low ed	-	-	-	-	0.00792	0.0363	-0.0601	-0.0816
					[0.0261]	[0.0343]	[0.0488]	[0.0557]
Spr K Prosocial * Med ed	-	-	-	-	0.00959	0.0852	0.0449	-0.065
					[0.0234]	[0.0407]+	[0.0620]	[0.0573]
Observations	11,800	10,070	8,030	6,630	11,800	10,070	8,030	6,630
R-squared	0.593	0.44	0.442	0.364	0.602	0.454	0.449	0.378

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to the nearest 10.

Table A6.35 Regression results: IV estimates for math scores with and without controls for behavior in Spring Kindergarten

	IV estimates without controls				IV estimates with controls			
	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
Constant	0.0985	0.0686	0.133	0.162	0.0937	0.0689	0.136	0.16
	[0.0167]*	[0.0206]*	[0.0351]*	[0.0344]*	[0.0165]*	[0.0206]*	[0.0351]*	[0.0339]*
Low parent education	-0.0782	-0.177	-0.292	-0.383	-0.081	-0.181	-0.298	-0.388
	[0.0255]*	[0.0300]*	[0.0452]*	[0.0563]*	[0.0264]*	[0.0309]*	[0.0456]*	[0.0570]*
Medium parent education	-0.0416	-0.0731	-0.11	-0.213	-0.0383	-0.075	-0.112	-0.209
	[0.0216]~	[0.0283]+	[0.0531]+	[0.0489]*	[0.0219]~	[0.0276]*	[0.0511]+	[0.0478]*
Spr K (6) score	0.832	0.93	0.984	0.849	0.824	0.935	0.987	0.839
	[0.0275]*	[0.0311]*	[0.0478]*	[0.0496]*	[0.0281]*	[0.0331]*	[0.0488]*	[0.0532]*
Spr K score * Low ed	-0.0173	-0.0826	-0.168	-0.15	-0.0346	-0.0955	-0.173	-0.152
	[0.0363]	[0.0378]+	[0.0616]*	[0.0654]+	[0.0375]	[0.0397]+	[0.0618]*	[0.0685]+
Spr K score * Medium ed	-0.00103	-0.0741	-0.171	-0.115	-0.0126	-0.0984	-0.206	-0.124
	[0.0321]	[0.0474]	[0.0696]+	[0.0644]~	[0.0334]	[0.0516]~	[0.0760]*	[0.0674]~
Spr K (6) score squared	-0.0442	-0.04	-0.108	-0.0672	-0.0418	-0.0408	-0.106	-0.0635
	[0.0142]*	[0.0167]+	[0.0259]*	[0.0317]+	[0.0143]*	[0.0168]+	[0.0259]*	[0.0313]+
Spr K score squared * Low ed	-0.0126	0.075	0.16	0.0902	-0.00941	0.0788	0.16	0.0923
	[0.0238]	[0.0266]*	[0.0404]*	[0.0491]~	[0.0248]	[0.0275]*	[0.0396]*	[0.0488]~
Spr K score squared * Med ed	-0.00834	0.00985	0.0458	0.0345	-0.00557	0.0143	0.0485	0.0384
	[0.0207]	[0.0324]	[0.0491]	[0.0470]	[0.0210]	[0.0321]	[0.0506]	[0.0474]
Spr K Conduct problems	-	-	-	-	-0.0263	-0.0175	-0.00125	-0.0608
					[0.0221]	[0.0240]	[0.0396]	[0.0388]
Spr K Conduct * Low ed	-	-	-	-	0.0477	0.0402	-0.0197	0.0742
					[0.0305]	[0.0353]	[0.0519]	[0.0551]
Spr K Conduct * Med ed	-	-	-	-	0.0469	0.0239	-0.0661	-0.0338

	IV estimates without controls				IV estimates with controls			
	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
					[0.0271]~	[0.0327]	[0.0582]	[0.0521]
Spr K Inattention problems	-	-	-	-	0.0134	0.0483	0.0304	-0.0158
					[0.0196]	[0.0236]+	[0.0451]	[0.0370]
Spr K Inattention * Low ed	-	-	-	-	-0.0793	-0.0514	-0.0121	-0.00017
					[0.0305]*	[0.0350]	[0.0643]	[0.0552]
Spr K Inattention * Med ed	-	-	-	-	-0.0855	-0.0944	-0.0632	0.00584
					[0.0251]*	[0.0329]*	[0.0603]	[0.0524]
Spr K Internalizing problems	-	-	-	-	-0.00821	0.00148	-0.0356	-0.00678
					[0.0136]	[0.0172]	[0.0269]	[0.0283]
Spr K Internalizing * Low ed	-	-	-	-	-0.0333	-0.034	0.0173	-0.0172
					[0.0194]~	[0.0237]	[0.0358]	[0.0369]
Spr K Internalizing * Med ed	-	-	-	-	0.0202	0.0012	0.0201	0.0161
					[0.0169]	[0.0221]	[0.0357]	[0.0412]
Spr K Prosocial problems	-	-	-	-	-0.0216	-0.0395	0.0264	0.068
					[0.0189]	[0.0226]~	[0.0311]	[0.0400]~
Spr K Prosocial * Low ed	-	-	-	-	0.0459	0.0229	-0.0372	-0.0839
					[0.0245]~	[0.0278]	[0.0449]	[0.0542]
Spr K Prosocial * Med ed	-	-	-	-	0.035	0.0342	-0.0285	-0.0542
					[0.0230]	[0.0298]	[0.0493]	[0.0534]
Observations	11,740	10,090	7,980	6,650	11,740	10,090	7,980	6,654
R-squared	0.606	0.566	0.544	0.492	0.612	0.569	0.55	0.498

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to the nearest 10.

Table A6.36 Regression results: IV estimates for reading scores with and without controls for behavior contemporaneous with the outcome

	IV estimates without controls				IV estimates with controls			
	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
Constant	0.144	0.297	0.237	0.328	0.131	0.268	0.215	0.293
	[0.0216]*	[0.0258]*	[0.0431]*	[0.0522]*	[0.0216]*	[0.0272]*	[0.0406]*	[0.0508]*
Low parent education	-0.0857	-0.318	-0.228	-0.51	-0.0825	-0.282	-0.199	-0.465
	[0.0344]+	[0.0456]*	[0.0717]*	[0.0720]*	[0.0344]+	[0.0451]*	[0.0681]*	[0.0702]*
Medium parent education	-0.0516	-0.18	-0.0974	-0.224	-0.0448	-0.153	-0.08	-0.199
	[0.0276]~	[0.0359]*	[0.0645]	[0.0784]*	[0.0288]	[0.0392]*	[0.0607]	[0.0748]*
Spr K (6) score	0.754	0.741	0.697	0.557	0.727	0.714	0.648	0.505
	[0.0444]*	[0.0486]*	[0.0721]*	[0.0638]*	[0.0470]*	[0.0483]*	[0.0713]*	[0.0628]*
Spr K score * Low ed	0.078	-0.0192	0.0238	0.00613	0.0878	-0.0272	0.0605	0.0382
	[0.0564]	[0.0591]	[0.0838]	[0.0802]	[0.0618]	[0.0628]	[0.0855]	[0.0802]
Spr K score * Medium ed	0.0408	0.0576	0.0233	0.156	0.0316	0.0496	0.00822	0.143
	[0.0633]	[0.0782]	[0.103]	[0.113]	[0.0661]	[0.0761]	[0.0960]	[0.0927]
Spr K (6) score squared	-0.0281	-0.0814	-0.0196	-0.0129	-0.0223	-0.0745	-0.018	-0.00703
	[0.0199]	[0.0241]*	[0.0381]	[0.0333]	[0.0204]	[0.0243]*	[0.0368]	[0.0337]
Spr K score squared * Low ed	-0.0325	0.0162	-0.0329	0.00981	-0.0318	0.00647	-0.0377	-0.0032
	[0.0300]	[0.0368]	[0.0531]	[0.0429]	[0.0315]	[0.0387]	[0.0508]	[0.0435]
Spr K score squared * Med ed	-0.0512	-0.067	-0.126	-0.197	-0.0496	-0.0649	-0.105	-0.173
	[0.0310]~	[0.0477]	[0.0594]+	[0.0785]+	[0.0317]	[0.0466]	[0.0529]+	[0.0670]+
Conduct problems (time t)	-	-	-	-	-0.0108	0.0195	-0.0477	-0.0359
					[0.0234]	[0.0338]	[0.0596]	[0.0728]
Conduct * Low ed	-	-	-	-	0.0946	0.00161	0.0329	0.0596
					[0.0491]~	[0.0539]	[0.0850]	[0.0958]
Conduct * Med ed	-	-	-	-	0.0479	0.0118	0.0299	0.0543

	IV estimates without controls				IV estimates with controls			
	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
					[0.0374]	[0.0485]	[0.0749]	[0.0873]
Inattention problems (time t)	-	-	-	-	-0.0651	-0.104	-0.116	-0.166
					[0.0246]*	[0.0336]*	[0.0493]+	[0.0529]*
Inattention * Low ed	-	-	-	-	-0.00793	0.0294	0.0713	0.0638
					[0.0369]	[0.0593]	[0.0791]	[0.0784]
Inattention * Med ed	-	-	-	-	-0.0169	-0.00459	-0.0192	0.0201
					[0.0400]	[0.0550]	[0.0642]	[0.0774]
Internalizing problems (time t)	-	-	-	-	-0.0205	-0.0177	-0.0835	-0.067
					[0.0157]	[0.0224]	[0.0367]+	[0.0384]~
Internalizing * Low ed	-	-	-	-	0.0281	0.0167	0.0808	0.0251
					[0.0294]	[0.0337]	[0.0716]	[0.0569]
Internalizing * Med ed	-	-	-	-	0.00393	0.0324	0.0383	0.0466
					[0.0249]	[0.0340]	[0.0535]	[0.0587]
Prosocial problems (time t)	-	-	-	-	0.0305	-0.0161	0.0704	0.0537
					[0.0197]	[0.0280]	[0.0494]	[0.0512]
Prosocial * Low ed	-	-	-	-	-0.0823	-0.0706	-0.0559	-0.0565
					[0.0364]+	[0.0508]	[0.0761]	[0.0757]
Prosocial * Med ed	-	-	-	-	-0.0507	0.00344	-0.00185	-0.0694
					[0.0315]	[0.0420]	[0.0605]	[0.0738]
Observations	4,470	4,450	4,470	3,660	4,470	4,450	4,470	3,660
R-squared	0.586	0.414	0.385	0.331	0.597	0.434	0.411	0.363

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to the nearest 10.

Table A6.37 Regression results: IV estimates for math scores with and without controls for behavior contemporaneous with the outcome

	IV estimates without controls				IV estimates with controls			
	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
Constant	0.116	0.0939	0.0906	0.151	0.113	0.0909	0.0865	0.13
	[0.0210]*	[0.0332]*	[0.0549]~	[0.0399]*	[0.0212]*	[0.0331]*	[0.0523]~	[0.0378]*
Low parent education	-0.0794	-0.167	-0.177	-0.326	-0.0858	-0.17	-0.172	-0.309
	[0.0365]+	[0.0504]*	[0.0684]+	[0.0643]*	[0.0366]+	[0.0497]*	[0.0631]*	[0.0627]*
Medium parent education	-0.0754	-0.0965	0.0416	-0.128	-0.0741	-0.0917	0.0558	-0.104
	[0.0307]+	[0.0433]+	[0.0845]	[0.0658]~	[0.0308]+	[0.0425]+	[0.0804]	[0.0626]~
Spr K (6) score	0.796	0.823	0.833	0.782	0.789	0.815	0.796	0.722
	[0.0338]*	[0.0470]*	[0.0617]*	[0.0505]*	[0.0345]*	[0.0473]*	[0.0575]*	[0.0443]*
Spr K score * Low ed	0.0375	0.0736	0.0533	-0.027	0.0123	0.0679	0.0701	0.0113
	[0.0462]	[0.0570]	[0.0779]	[0.0629]	[0.0481]	[0.0553]	[0.0746]	[0.0580]
Spr K score * Medium ed	0.0485	0.0674	0.0953	0.0175	0.0336	0.0593	0.113	0.048
	[0.0414]	[0.0587]	[0.0833]	[0.0726]	[0.0436]	[0.0603]	[0.0835]	[0.0718]
Spr K (6) score squared	-0.0276	0.0115	-0.00617	-0.0145	-0.0268	0.00869	-0.00588	-0.00236
	[0.0211]	[0.0320]	[0.0331]	[0.0283]	[0.0208]	[0.0319]	[0.0319]	[0.0247]
Spr K score squared * Low ed	0.00281	0.0147	0.0376	0.0156	0.00743	0.0217	0.0375	0.00251
	[0.0353]	[0.0430]	[0.0544]	[0.0521]	[0.0351]	[0.0426]	[0.0506]	[0.0490]
Spr K score squared * Med ed	-0.0112	-0.036	-0.15	-0.0814	-0.00775	-0.0313	-0.141	-0.0835
	[0.0311]	[0.0449]	[0.0760]+	[0.0563]	[0.0309]	[0.0446]	[0.0714]+	[0.0520]
Conduct problems (time t)	-	-	-	-	-0.0197	-0.0124	0.0251	-0.0872
					[0.0260]	[0.0418]	[0.0777]	[0.0659]
Conduct * Low ed	-	-	-	-	0.0977	0.0965	0.0133	0.106
					[0.0363]*	[0.0573]~	[0.0924]	[0.0896]
Conduct * Med ed	-	-	-	-	0.064	0.0159	-0.185	0.0182

	IV estimates without controls				IV estimates with controls			
	1st G (7)	3rd G (9)	5th G (11)	8th G (14)	1st G (7)	3rd G (9)	5th G (11)	8th G (14)
					[0.0366]~	[0.0513]	[0.0963]~	[0.0912]
Inattention problems (time t)	-	-	-	-	-0.0242	-0.032	-0.0982	-0.122
					[0.0271]	[0.0422]	[0.0541]~	[0.0468]*
Inattention * Low ed	-	-	-	-	-0.071	-0.00885	0.0385	0.101
					[0.0409]~	[0.0540]	[0.0655]	[0.0680]
Inattention * Med ed	-	-	-	-	-0.0149	-0.00784	0.0333	0.0309
					[0.0367]	[0.0542]	[0.0823]	[0.0800]
Internalizing problems (time t)	-	-	-	-	-0.00804	-0.0323	-0.0472	-0.0226
					[0.0186]	[0.0316]	[0.0348]	[0.0307]
Internalizing * Low ed	-	-	-	-	-0.0263	0.00745	-0.0242	-0.0621
					[0.0272]	[0.0413]	[0.0500]	[0.0509]
Internalizing * Med ed	-	-	-	-	-0.0349	0.0384	0.0761	0.018
					[0.0217]	[0.0405]	[0.0511]	[0.0480]
Prosocial problems (time t)	-	-	-	-	0.0357	0.0431	0.0284	0.0809
					[0.0237]	[0.0293]	[0.0525]	[0.0432]~
Prosocial * Low ed	-	-	-	-	-0.0452	-0.0953	0.0138	-0.104
					[0.0349]	[0.0524]~	[0.0715]	[0.0769]
Prosocial * Med ed	-	-	-	-	-0.029	-0.0705	0.15	-0.00432
					[0.0390]	[0.0456]	[0.0835]~	[0.0712]
Observations	4,450	4,440	4,440	3,680	4,450	4,440	4,440	3,680
R-squared	0.589	0.517	0.497	0.471	0.597	0.524	0.514	0.489

Note. Standard errors in brackets. *, +, and ~ indicate significance at the 1%, 5% and 10% levels respectively. All sample sizes rounded to the nearest 10.

Appendix to Chapter 7

Section A7.1 Out-of-school resources in 4th grade

Table A7.1 Out-of-school challenges reported by 4th grade teachers

	US	*UK*	*AU*	*CA*
Instruction is limited some or a lot by:				
Students lacking basic nutrition (%)	40	23	27	33
Students lacking sufficient sleep (%)	76	63	67	67
Disruptive students (%)	16	9	14	18
Uninterested students (%)	11	3	5	6

Source: PIRLS 2011 International Results in Reading, Exhibit 8:10 and 8:11. Data for the UK refer to England only.

Table A7.2 Resources reported by 4th grade students

	US	*UK*	*AU*	*CA*
More than 100 books in home (%)	28	36	41	35
Own room and internet connection in home (%)	64	73	74	77

Source: PIRLS 2011 International Results in Reading, Exhibit 4:2. Data for the UK refer to England only.

Section A7.2 Results from national testing in England

Table A7.3 Proportion of pupils achieving expected National Curriculum level

	English			Math	
	Age 7 reading (KS1) 2008	Age 7 writing (KS1) 2008	Age 11 English (KS2) 2012	Age 7 math (KS1) 2008	Age 11 math (KS2) 2012
All pupils	84	80	85	90	84
FSM	69	64	74	79	73
Non-FSM/Other pupils	87	84	88	92	87
Gap	**18**	**20**	**14**	**13**	**14**

Notes:
Expected level is Level 2 at age 7 (Key Stage 1) and Level 4 at age 11 (Key Stage 2).
Key Stage 1 results from 2008, Key Stage 2 results from 2012 (pupils assessed in May of Year 2/6).
"Other pupils" includes those with unknown FSM status at age 11 but not at age 7.
Key Stage 1 assessments are teacher assessed. Key Stage 2 English combines results from a reading test and a writing teacher assessment. Key Stage 2 math is a test.
Key Stage 2 coverage is England, State-funded schools (including Academies and CTCs). Key Stage 1 coverage is England. Roughly 535,000 pupils.
Source: Dept for Children, Schools and Families (2008), Dept for Education (2012). Published national statistics from the National Pupil Database.

References

Australian Curriculum, Assessment and Reporting Authority (2014). National Assessment Program – Literacy and Numeracy 2013: Technical Report, ACARA, Sydney.

Australian Institute of Family Studies (AIFS) (2013). *Longitudinal Study of Australian Children Data User Guide-August 2013.* Melbourne, AU.

Bradbury, Bruce, Miles Corak, Jane Waldfogel, and Elizabeth Washbrook (2012). "Inequality during the Early Years: Child Outcomes and Readiness to Learn in Australia, Canada, United Kingdom, and United States." In John Ermisch, Markus Jantti, and Timothy Smeeding (eds). *From Parents to Children: The Intergenerational Transmission of Advantage.* New York: Russell Sage Foundation.

Cole, Tim J., Mary C. Bellizzi, Katherine M. Flegal, and William H. Dietz (2000). "Establishing a Standard Definition for Child Overweight and Obesity Worldwide: International Survey." *British Medical Journal* 320(7244): 1240-1243.

Cusack, Benedict and Ryan Defina (2013). *Wave 5 Weighting and Non-Response*. Melbourne, AU: Australian Institute of Family Studies.

Daraganova, Galina, Ben Edwards, and Mark Sipthorp (2013). *Using National Assessment Program-Literacy and Numeracy (NAPLAN) Data in the Longitudinal Study of Australian Children (LSAC)*. Melbourne, AU: Australian Institute of Family Studies.

Dunn, Lloyd M., Leota M. Dunn, & Douglas M. Dunn (1997). *The Peabody Picture Vocabulary Test, 3rd Edition*. Circle Pines, MN: American Guidance Service Inc Publishing.

Foy, Pierre, Bradley Brossman, and Joseph Galia (2012). *Scaling the TIMSS and PIRLS 2011 Achievement Data*. Available at: http://timssandpirls.bc.edu/methods/pdf/TP11_Scaling_Achievement.pdf

Gallop, Kathryn et al. (2013). *Millennium Cohort Study Fifth Sweep (MCS5) Teacher Survey: Technical Report*. London, UK: Ipsos MORI.

Gonzales, Patrick, Juan Carlos Guzmán, Lisette Partelow, Erin Pahlke, Leslie Jocelyn, David Kastberg, and Trevor Williams (2004). *Highlights from the Trends in International Mathematics and Science Study (TIMSS) 2003* (NCES 2005-005). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hansen, Kirstine (2014). *Millennium Cohort Study: A Guide to the Datasets, 8th Edition*. London, UK: Centre for Longitudinal Studies.

Hotz, V. Joseph and John Karl Scholz (2001). *The Earned Income Tax Credit*. Available at: http://www.nber.org/papers/w8078.pdf

Kessler, Ronald C. et al. (2002). "Short Screening Scales to Monitor Population Prevalences and Trends in Non-Specific Psychological Distress." *Psychological Medicine* 32(6): 959-976.

Kessler, Ronald C. et al. (2003). "Screening for Serious Mental Illness in the General Population." *Archives of General Psychiatry* 60(2):184-189.

Letourneau, Nicole Lyn, L.; Cara B. Fedick, Douglas Willms, Cindy Lee Dennis, Kathy Hegadoren, and Miriam J. Stewart (2006). "Longitudinal Study of Postpartum Depression, Maternal-Child Relationships and Children's Behaviour to 8 Years of Age." In Dorothy M. Devore (ed). *Parent-Child Relations: New Research*. New York, NY: Nova Science.

LoGerfo, Laura, Austin Nichols, and Sean F. Reardon (2006). *Achievement Gains in Elementary and High School*. Washington, DC: Urban Institute.

National Assessment Governing Board (1994). *Reading Framework for the 1992 and 1994 National Assessment of Educational Progress*. Washington, DC: Government Printing Office.

National Assessment Governing Board (1996). *Mathematics Framework for the 1996 National Assessment of Educational Progress*. Washington, DC: Government Printing Office.

Ogle, Laurence, Anindita Sen, Erin Pahlke, Leslie Jocelyn, David Kastberg, Stephen Roey, and Trevor Williams (2003). *International Comparisons in Fourth-Grade Reading Literacy: Findings from the Progress in International Reading Literacy Study (PIRLS) of 2001* (NCES 2003-073). Washington, DC: U.S. Department of Education, NCES.

Pollack, Judith M, Sally Atkins-Burnett, Michelle Najarian, and Donald A. Rock (2005). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K), Psychometric Report for the Fifth Grade*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Radloff, Lenore Sawyer (1977). "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population." *Applied Psychological Measurement, 1,* 385-401.

Reardon, Sean F. (2011) Online Appendix 5. The Widening Academic-Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations. In the online appendix to Greg J. Duncan and Richard J. Murnane (eds), *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances*. Russell Sage Foundation. Available at: https://www.russellsage.org/sites/all/files/duncan_murnane_online_appendix.pdf

Rock, Donald A. and Judith M. Pollack (2002). *Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K), Psychometric Report for Kindergarten Though First Grade*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Rothman, Sheldon (2005). *Report on Adapted PPVT-III and Who Am I?* Melbourne , AU: Australian Institute of Family Studies, Melbourne.

Royston, Patrick (2005). "Multiple Imputation of Missing Values: Update of ICE." *The Stata Journal, 5,* 527-536.

Statistics Canada (2005). *Microdata User Guide. National Longitudinal Survey of Children and Youth. Cycle 6. September 2004 to June 2005*. Available at: http://www23.statcan.gc.ca/imdb-bmdi/document/4450_D4_T9_V7-eng.pdf

Tax Policy Center (2013), *Historical EITC Parameters*. Washington, DC: Urban Institute and Brookings Institution. Available at http://www.taxpolicycenter.org/taxfacts/displayafact.cfm?Docid=36

Tourangeau, Karen et al. (2000). *ECLS-K Restricted-Use Base Year Data Files and Electronic Codebook*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Tourangeau, Karen et al. (2002). *User's Manual for the ECLS-K First-grade Restricted-Use Date Files and Electronic Codebook*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Tourangeau, Karen, Christine Nord, Thand Le, Alberto G. Sorongon, Michelle Najarian, and Elvira Germino Hausken (2009). *Combined User's Manual for the ECLS-K Eight-grade and K-8 Full Sample Data Files and Electronic Codebooks*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Vidmar, Suzanna I., Tim J. Cole, and Huiqi Pan (2013). "Standardizing Anthropometric Measures in Children and Adolescents with Functions for Egen: Update" *Stata Journal, 13*(2), 366-378.

Zopluoglu, Cengiz (2012), "A Cross-National Comparison of Intra-Class Correlation Coefficients in Educational Achievement Outcomes." *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, Yaz* 3(1): 242-278.