DICE: Technical Appendix

UNDER CONSTRUCTION

This version: Saturday, 07 May 2022

Contact in case of queries:
Dr Liz Washbrook
University of Bristol
Liz.Washbrook@bristol.ac.uk

Contents

1. Introduction to the DICE Project

DICE The Development of Inequalities in Child Educational Achievement: A Six Country Study is a project funded under Round 5 of the Open Research Area (ORA) for the Social Sciences scheme (original funding period: January 2019 – December 2021).

The Round 5 ORA scheme was run by the Agence Nationale de la Recherche (ANR, France), the Deutsche Forschungsgemeinschaft (DFG, Germany), the Economic and Social Research Council (ESRC, United Kingdom) and the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO, The Netherlands).

The Principal Investigator of the DICE project is Dr. Liz Washbrook (University of Bristol) and the DICE team is a collaboration of researchers based at the following institutions:

- Institut National d'Etudes Démographiques, INED (France)
- University of Bamberg (Germany)
- University of Leipzig (Germany)
- Erasmus University Rotterdam (Netherlands)
- University of Bristol (United Kingdom)
- Columbia University (USA)
- Keio University (Japan)
- Kochi University (Japan)
- Teikyo University (Japan)
- Japan Women's University (Japan)

Aims and scope

The aim of the Development of Inequalities in Child Educational Achievement (DICE) project is to advance our understanding of disparities in child development by parental education status, a key indicator of parental socioeconomic status (SES). Inequalities in child development by parental SES are a direct impediment to social mobility and life chances and hence are a concern across OECD countries.

In order to investigate inequalities in child development, the DICE project leverages rich cohort and administrative data from six countries - France, Germany, Japan, Netherlands, the United Kingdom, and the United States – embedding them in a harmonised framework. These six countries are similar enough to form valid comparisons, but also sufficiently different to allow us to learn about the role of context in the development of inequalities.

The novel feature of this project is to move beyond cross-sectional and single country snapshots by studying the question of how inequalities develop over time (ages 3 to 16), what factors may influence inequalities and how national context may strengthen or buffer these processes.

Because different aspects of development matter and may be differentially affected by SES and by country contexts, child development is conceptualised broadly, to include cognitive, social/emotional and health outcomes, thus recognising the interplay of multiple spheres of development in childhood.

The specific aims of the project are:

1) to provide new evidence on the extent and sources of inequalities in early childhood and at the start of school, in particular on the role of parenting/home environment and preschool
2) to describe trajectories of child development, and identify factors such as parental involvement and parent-school interactions that reduce/increase inequalities as children move through primary school
3) to describe inequalities at the start of secondary school and analyse trajectories of development, and factors such as school type and tracking that reduce/increase inequalities as children move through secondary school

2. DICE datasets

The DICE project analyses inequalities in child development across six disparate country contexts, relying on an unprecedented array of extremely rich cohort and administrative data (listed in Table 2.1). The project consists of six sets of analyses, which build sequentially to advance our understanding of inequalities in child development across the six countries.

- Analysis 1: benchmarking of inequalities across the six countries with existing cross-sectional international large-scale assessment (ILSA) data from PISA, PIRLS and TIMSS (around ages 10 and 15).
- Analysis 2: Inequalities at age 3-4 in the DICE micro-datasets
- Analyses 3 and 4: Inequalities from ages 5-7 to 9-11 in the DICE micro-datasets
- Analyses 5 and 6: Inequalities from ages 9-11 to 14-16 in the DICE micro-datasets

Different analyses draw on different datasets, depending on the birth cohort years and dates of survey follow-up, as shown in Table 2.1.

Table 2.1. Datasets, birth cohort years and years of data collection of DICE surveys

Country	Birth cohorts	Year(s) of data collection
France		
ELFE birth cohort	2011	2011-2017
DEPP panel primary	2005	2011-2016
DEPP panel secondary	1996	2007-2013
Germany		
NEPS SC1	2012	2012-2019
NEPS SC2	2005-2006	2011-2019 (1st Grade: 2013)
NEPS SC3	1998-2000	2010-2017
Japan		
JCPS (KHPS & JHPS)	2003-2012	2010-2018
LSN21	2001	2001-2017
Netherlands		
Generation R	2002-2006	2002 - ongoing
UK		
MCS	2000-2002	2001, 2004, 2006, 2008, 2012, 2015, 2018
US		
ECLS-B	2001	2001-2008
ECLS-K 2011	2004-2005	2010-2016
ECLS-K 1998	1992-1993	1998-2007

Note. Acronyms stand for: ELFE: French Longitudinal Study of Children; DEPP: La Direction de l'évaluation, de la prospective et de la performance; NEPS: National Educational Panel Study; JCPS: Japan Child Panel Survey; LSN21: Longitudinal Study of Newborns; MCS: Millennium Cohort Study; ECLS-B: Early Childhood Longitudinal Study, Birth Cohort; ECLS-K: Early Childhood Longitudinal Study, Kindergarten Cohort.

2.1. France

ELFE

The Etude Longitudinale Française depuis l'Enfance/French Longitudinal Study since Childhood (ELFE) is a national longitudinal cohort study with recruitment representative of live births in France in 2011 (see Charles at al. (2020) for further details).

At baseline, the ELFE consisted of all children born in 2011: (1) at a chosen hospital (349 hospitals randomly chosen, of which 320 participated, out of 544 hospitals with a maternity ward in continental France – i.e. excluding Overseas Territories -, stratified by the maternity ward size); (2) during an inclusion period (four inclusion periods: in April (4 days); June-July (6 days); September-October (7 days); November-December (8 days)). Excluded were babies at less than 33 weeks gestation (to avoid overlap with a concurrent survey of very premature births); mothers less than 18 years at birth and not able to sign a consent form in one of 5 available languages (due to consent issues; very few mothers in France are under 18); triplets and higher order births.

Maternity ward served as a primary sampling unit (PSU). Maternity ward was stratified by ward size and birth period (4 to 8 days each, in March, June, September and November). No strata is oversampled. All births within both strata (born at a selected hospital during a selected inclusion period) were targeted.

Baseline wave consisted of short face-to-face questionnaire at inclusion in the maternity ward shortly after birth (Wave 1, n=18,329); first main interview, by telephone, with both parents roughly 2 months after birth (Wave 2, n=16,567). At Wave 1, 51% of targeted births agreed to participate. There were no additions to sample after Wave 1.

Follow-ups were conducted at: Wave 3 (age 1 year); Wave 4 (age 2); Wave 5 (age 3.5); Wave 6 (age 5.5). Wave 6 is planned for 2021 (age 10). Response rates were Wave 2: 87%; Wave 3: 77%; Wave 4: 71%; Wave 5: 64% (Denominator = 18,041 households with complete maternity questionnaire). 59% participated in all of the first four waves (n=10,688).

Systematic exclusions to follow-up (i.e. other than refusal/unproductive at wave t-1) were child death; family emigrated outside continental France; parent asks to no longer be contacted; ELFE study team decision to exclude families after too many refusals/ unproductive issues.

DEPP panel primary

Under construction

DEPP panel secondary

Under construction

2.2. Germany

DICE project data for Germany came from the German National Educational Panel Study (NEPS). The NEPS is a multi-cohort study aimed at providing data on the development of a range of skills throughout the life course of cohort members. From 2008 to 2013, NEPS data was collected as part of the Framework Programme for the Promotion of Empirical Educational research funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung and Forschung, BMBF); since January 2014, the NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network. Below a brief review of the main characteristics of the relevant NEPS cohorts is presented, further details can be found at https://www.neps-data.de/.

NEPS-SC1

The target population of Starting Cohort 1 is defined as all children born in Germany from February 2012 to July 2012 and their families. At the start of the panel survey, the target children had to be at least six months old, but not older than eight months, in order to ensure a valid measurement of infant development. This means that the time window for direct measurements with the newborns was fixed exactly according to the age of the child.

Access to this population was via a register-based sample of addresses available at the municipal level. The random sample is based on a two-stage disproportional stratified sampling strategy with: municipalities as primary sampling units, proportionally stratified according to a classification of urbanization (BIK scale) and addresses of newborns as secondary sampling units, disproportionally stratified with more addresses in bigger municipalities.

The selection of 84 municipalities at the first stage was based on the distribution of births in the first half of 2009 according to the German Microcensus in three explicit strata (less than 50,000 inhabitants; 50,000 to 500,000 inhabitants; 500,000 and more inhabitants), whereby municipalities having less than ten births were excluded. At the second stage, addresses were then randomly selected from the municipalities' register data via systematic interval sampling, divided into two tranches (births from February to April; births from May to July1). In the end, a gross sample size of 8,483 addresses out of 90 sampling points in 84 municipalities turned out to be sufficient to achieve the planned sample size of approximately 3,000 newborns. With 3,481 participants in the first survey wave of Starting Cohort 1, the realized sample size has clearly exceeded this target, corresponding to a response rate of 41 percent (FDZ-LIfBi Data Manual 2018: 12f.).

In Wave 2, parent interviews were conducted with all parents from Wave 1 who gave their consent to be contacted again, but only a subsample of children was asked to take part in the direct measurements. A random sample of 34 municipalities has been drawn from the initial 84 municipalities for this purpose. Wave 2 was conducted when children were 12-17 month-old (April-December 2013). At that time, 2,849 parents were interviewed, 1,407 children took part in direct measurements. Wave 3 was conducted when children were 25-27 months-old (April-December 2014). In Wave 3, all panel respondents—children and parents—were

Since the response rate in tranche 1 was unexpectedly high, those target persons born in July were not used, provided the exact month of birth was known.

invited to be surveyed. 2,609 parents were interviewed at home; data on direct measures is available for 1,921 children. At wave 4, children were 37-39 months old (April-December 2015). 2,478 parents were interviewed and data on competence tests is available for 2,324 children. Wave 5 was conducted when children were 48 month-old (April-September 2016). 2,381 parents were interviewed and data on competence tests is available for 2,138 children. Wave 6 was conducted when children were five years old (March-August 2017). At that time, 2,209 parents were interviewed and competencies of 2,087 children were tested. In Waves 2-6, educators/childminders were also interviewed. In addition, institution managers were interviewed in Waves 4-6.

NEPS-SC2

A brief overview of the NEPS – Starting Cohort 2 (NEPS-SC) is presented below. For more information, see Skope, Pink, and Bela (2012).

Population represented by sample at baseline and sample design. The NEPS-SC2 was designed to collect data on two underlying population, kindergarten and elementary school children. The design of these two samples was coordinated, thus large parts of the kindergarten sample are also present in the elementary school sample. For the kindergarten sample, the target population of the NEPS-SC2 was all children that in 2010/2011 attended day-care facilities (kindergartens) and were thus expected to enter primary school in the school year 2012/2013.

The sample was drawn using an indirect two-stage approach, with institutions – i.e., kindergartens – drawn in a first step and children in a second. Since no complete list of kindergartens existed in Germany, elementary schools with a Grade 1 were sampled on a nationwide and representative basis through size-proportional random selection. Then, each sampled school provided a list of linked kindergartens: kindergartens and elementary schools are connected in that children transfer from one to the other and so schools were asked to provide a list of kindergarten from which children had entered their school in the 2009/2010 year. In a second sampling stage, a set of these kindergartens was randomly drawn from each of those lists. All four-year-old children at the sampled kindergartens were then invited to participate in the study. This indirect sampling strategy allowed to simultaneously provide random samples for kindergarten and elementary school children.

Baseline cohort and follow-up waves. At baseline, 2,996 kindergarten children and their parents in 279 Kindergartens agreed to participate in the study, which corresponded to a response rate of 56.2%. At the end, 2,741 children and 2,340 parents as well as 831 educators and 237 principals that actually took part in the study. Yearly follow-up with these children were then carried out in Waves 2-9. Table 2.2 below reports, for each wave, the number of children interviewed and tested, the number of parents to whom phone interviews were administered, and the number of educators/class teachers and school principals of the target students that took part in the study. It is important to note that Wave 3 (and then all subsequent waves) included a refreshment sample. For this follow-up the originally sampled kindergarten students had now moved to Grade 1 in elementary schools. Thus, this sample was integrated by inviting all 1st graders at the originally sampled elementary schools to participate in the study.

Table 2.2: Baseline and follow-up waves of NEPS-SC2

	WAVE 1	**WAVE 2**	**WAVE 3**	**WAVE 4**	**WAVE 5**
Interview period: start	Jan 2011	Jan 2012	Feb 2013	Nov 2013	Nov 2014
Interview period: end	Dec 2011	May 2012	Oct 2013	May 2014	Jun 2015
Age	4	5	6	7	8
Grade	K	K	1	2	3
Children	2,949	2,727	6,734	6,333	5,800
Parents	2,349	2,111	6,935	6,201	5,299
Educators/Teachers: assessment sheet	2,741	2,497	5,974	785	693
Educators/ Teachers: questionnaire	831	975	874	747	664
Heads of Kindergartner/Principals	237	220	324	303	279

	WAVE 6	**WAVE 7**	**WAVE 8**	**WAVE 9**
Interview period: start	Sep 2015	Sep 2016	Sep 2017	Oct 2018
Inteview period: end	Jun 2016	Apr 2017	Jan 2018	Jun 2019
Age	9	10	11	12
Grade	4	5	6	7
Children	5,418	4,220	4,164	4,088
Parents	4,873	4,356	-	3,277
Educators/Teachers: assessment sheet	636	-	-	-
Educators/ Teachers: questionnaire	636	-	-	-
Heads of Kindergartner/Principals	292	-	-	-

NOTE: this tables is based on the latest study overview report by the Leibniz Institute for Educational Trajectories (LIfBi, 2020)

NEPS-SC3

Under construction

2.3 Japan

JCPS

Under construction

LSN21

The Longitudinal Survey of Newborns in the twenty-first Century (LSB21) series covers children born between January 10 and 17, 2001, and between July 10 and 17, 2001, nationwide. Japan's Ministry of Health, Labour and Welfare sampled the subjects based on the live birth forms from the Vital Statistics Data (survey). In the case of twins and triplets, both siblings were surveyed individually. There is also another cohort, the 2nd LNS21 (Babies born in 2010) for 2011. Date of first survey was August 1, 2001 for babies born between January 10 and 17, 2001 and February 1, 2002 for babies born between July 10 and 17, 2001 (the babies are 6 months old infants). Response rate at Wave 1 was 87.7%. The years of a cohort of children born on May 10th through May 24th in 2010 and to compare with those babies born in 2001 with help of parents and to capture key information used for formulating measures against the declining birth rate. Children whose birth registrations were not submitted to Japanese municipalities are not included. In addition, when the address of birth registration is outside of Japan (full address is not required to be written), the questionnaire cannot be sent to the incomplete address. Therefore, those children are also not included from the first wave.

Follow-up waves were collected annually at ages 6 month to 16 year-old. Response rates were 87.7%.(Wave 1), 93.5%(Wave 2), 91.3%(Wave 3),92.7%(Wave 4), 91.4%(Wave 5), 91.3%(Wave 6), 90.6%(Wave 7), 92.0%(Wave 8), 93.0%(Wave 9), 92.3%(Wave 10), 91.3%(Wave 11), 91.5%(Wave 12), 89.5%(Wave 13), 89.8%(Wave 14), 91.7%(Wave 15).

2.4 Netherlands

The DICE project used one longitudinal survey for the Netherlands, the Generation R study (Gen-R). The Gen-R is a prospective cohort study following cohort members from foetal stage to young adulthood. The Gen-R Study is conducted by the Erasmus Medical Center (MC) in close collaboration with the Erasmus University Rotterdam, the Municipal Health Service Rotterdam area, the Rotterdam Homecare Foundation, and the Stichting Trombosedienst & Artsenlaboratorium Rijnmond (STAR). The study is made possible by long-term financial support from Erasmus MC, Netherlands Organization for health Research and Development (ZonMw) and the Dutch Ministry of Health, Welfare and Sport. Below a brief review of the main characteristics of Gen-R is presented, further details can be found at https://generationr.nl/.

Gen-R

A brief overview of Gen-R is presented below. For more information on the study design and initial cohort profile, see Jaddoe et al. (2006); for subsequent cohort profile updates see: Jaddoe et al. (2008; 2010; 2012) and Koojman et al. (2017).

Population represented by sample at baseline and sample design. Gen-R focused on the multi-ethnic urban population of Rotterdam, the second largest city in the Netherlands. The target population was that of expecting mothers living in the municipality of Rotterdam, including the suburban areas of the city, with an expected delivery date between April 2002 and January 2006. Enrolment of eligible mothers – and their partners - in the study was possible until the birth of the child/children. All eligible pregnant women who had visited a midwife or obstetrician in Rotterdam were contacted by phone or, if not possible, by home visits. Eligible pregnant women were contacted again when they had their scheduled first ultrasound. Pre-birth maternal assessments were carried out at multiple times during pregnancy whilst fathers were only assessed once. Once the children were born, they formed a prenatally recruited birth cohort and were then followed into young adulthood.

Baseline cohort and follow-up waves. At baseline 9,778 expecting mothers in the municipality of Rotterdam were enrolled in the Gen-R study, response rate = 61%. These mothers gave birth to 9,749 live born children, the cohort members. The Gen-R study was then designed to have two stages of follow-up, one for the pre-school period assessing children living in Rotterdam several times during their first two years of life, and one for the childhood and adolescent period, articulated into 3 waves (age 6, 10, and 13) and focusing on all 9.449 cohort members) - see Figure 2.1 below.

Of the original 9,749 cohort members, 1,166 lived outside the defined study area at birth and were therefore not approached for the postnatal follow-up studies for the preschool period. Of the remaining 8,583 children, 690 (8%) parents did not give consent, or their children died or were lost to follow-up. Thus, 7,893 children took part in the study focusing on the preschool period; the logistics of the study was embedded in the municipal routine childcare system. Nine postnatal follow-ups were carried out at 2 months, 3 months, 6 months, 12 months, 18 months, 24 months, 30 months, 36 months, 48 months.

At age 6 (early school age), all 9,278 children from the original cohort of 9,749 children were invited to participate in follow-up studies (471 children not invited due to being dead, withdrawn from the study, or lost to follow-up). This invitation was independent of their home address and participation in the preschool period. In total, 8,305 children (90% of those who were invited and 85% of the original cohort) participated in the study at this age, of whom 6,690 visited the research centre were assessments were carried out. For the next follow-up phase at the age of 10 years (mid childhood period) 730 children of the 9,278 could not be invited (again, due to being dead, having withdrawn from the study, or being lost to follow-up). In total, 7,393 children (86% of those who were invited and 76% of the original cohort) participated in the study in mid-childhood, of whom 5,862 visited the research centre. Finally, for the next follow-up phase at age 13 (early adolescence period), 7,968 children and their parents were invited (580 had either withdrawn from the study or were lost to follow-up).

Figure 2.1:Generation-R: design and cohort update

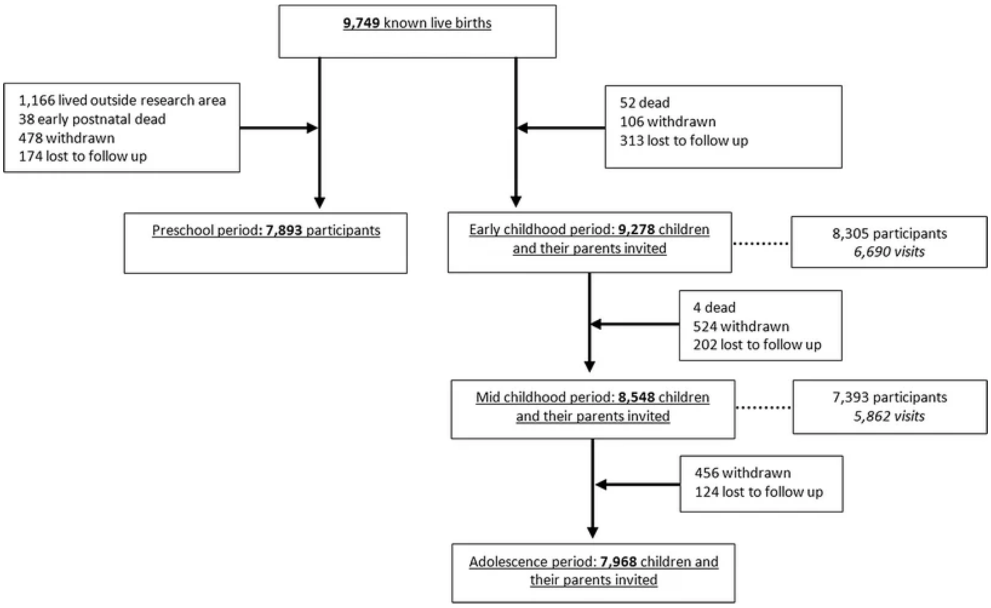

Figure taken from Kooijman et al (2017)

2.5. United Kingdom

The DICE project used one longitudinal survey for the United Kingdom, the Millennium Cohort Study (MCS). The MCS is a longitudinal cohort study following children in England, Scotland, Wales, and Northern Ireland through childhood and adolescence into young adulthood. The MCS was initiated and has been core funded by the Economic and Social Research Council (ESRC), and the consortium of Government Departments; it is conducted by the Centre for Longitudinal Studies (CLS). Below a brief review of the main characteristics of the MCS is presented, further details can be found at [CLS | Millennium Cohort Study (ucl.ac.uk)](https://cls.ucl.ac.uk).

MCS

A brief overview of the MCS is presented below. For more information on the study design and cohort profile, see CLS (2020a; 2020b; 2020c).

Population represented by sample at baseline and sample design. The Millennium Cohort Study (MCS) population at baseline consisted of all children born between 1st September 2000 and 11th January 2002[2], alive and living in the UK at age 9 months, eligible for (universal) Child Benefit (CB). Excluded were those ineligible for CB due to temporary or uncertain residency status (e.g., members of foreign armed forces, asylum seekers) and Sensitive Child Benefit cases (e.g., child taken into care by 9 months).

The primary sampling unit (PSU) used was electoral ward. In total 9 strata were created, three for England, and two each for Scotland, Wales, and Northern Ireland. In these latter three countries, disadvantaged electoral wards were first identified as those that fell in the poorest 25 percent of wards using the Child Poverty Index in orde to form, respectively, the *Scotland – Disadvantaged*, the *Wales-Disadvantaged*, and the *Northern Ireland – Disadvantaged* strata. The remaining electoral wards in each country formed then the corresponding *Scotland – Advantaged*, *Wales-Advantaged*, and *Northern Ireland – Advantaged* strata. In England, an *England – Ethnic* stratum was first identified by selecting electoral wards that in the 1991 Census had a proportion of ethnic minorities of at least 30 percent. Of the remaining wards, those that fell into the poorest 25 percent of wards according to the Child Poverty Index were placed in the *England – Disadvantaged* stratum, and those that did not were placed in the *England – Advantaged* stratum. Note that the indicators used for stratification were area-level measures. Thus, the MCS was designed to better identify those from a disadvantaged or a minority ethnic background when they lived in areas where people shared that same background.

The MCS sample was then randomly selected within each stratum to produce a disproportionately stratified cluster sample. In each selected sample ward, the Department for Work and Pensions (DWP) used the CB register to get in contact with all eligible families inviting CB recipients to opt-out if they did not want to be included in the survey.

Baseline cohort and follow-up waves. At baseline, 18,552 productive families took part in the study, including 18,818 cohort children (due to multiple births: twins, triplets, and two singletons) which were surveyed at approximately 9 months of age, with a response rate around 90%. At Wave 2 (age 3) 692 'new families' were added to the sample of wave 1 productive families. These so-called 'new-families' were families that although eligible at baseline, did not participate in MCS1; they were identified through the DWP and contacted. Thus, the MCS sample of families that took part in the study (ever) is 19,244. Five subsequent waves took

---

[2] Slightly different windows of study enrolment were used in the four UK countries. England and Wales: 1st September 2000 and 31st August 2001. Scotland and Northern Ireland: 24th November 2000 and 11th January 2002.

place: MCS3 (age 5), MCS4 (age 7), MCS5 (age 11), MCS6 (age 14), and MCS7 (age 14) following up MCS cohort members that were still alive and living in the UK and that had not been classified as 'permanent refusals' or 'permanently untraced' (unless they opted back into the study or new address details for them were found) by the CLS who was responsible for carrying out the fieldwork for the study. Table 2.3 reports, for each MCS wave, information such as, the year in which fieldwork took place, the number of productive families (and cohort members), and the survey response rate.

Table 1.3: Baseline and follow-up waves of MCS

	Fieldwork	**Issued sample**	**N. of productive families**	**Survey Response rate**	**Productive families as % of the 19,244 families that were ever in the sample**	**N. of cohort members**
MCS1	2001	20,646	18,552	90%	96%	18,818
MCS2	2004	19,870	15,590	78%	81%	15,808
MCS3	2006	18,528	15,246	82%	79%	15,459
MCS4	2008	17,031	13,857	81%	72%	14,043
MCS5	2012	16,393	13,287	81%	69%	13,469
MCS6	2015	15,415	11,726	76%	61%	11,872
MCS7	2018	14,496	10,626	73%	55%	10,757

NOTE: the numbers reported in this tables are taken from the User Guide to the MCS, surveys 1 to 5 (CLS, 2020a) and the technical reports for MCS6 (Ipsos MORI, 2017) and MCS7 (Ipsos MORI, 2019).

2.6. United States

The DICE project used multiple cohorts from the Early Childhood Longitudinal Studies (ECLS). The ECLS is a multi-cohort study aimed at providing data on child development, school readiness, and early school experiences. The ECLS data is sponsored by the National Center for Education Statistics (NCES) within the Institute of Education Sciences (IES) of the U.S. Department of Education. Below a brief review of the main characteristics of the relevant ECLS cohorts is presented, further details can be found at https://nces.ed.gov/ecls/.

ECLS-B

The Early Childhood Longitudinal Study – Birth Cohort (ECLS-B) selected a nationally representative probability sample of children born in the United States in 2001. Births were sampled within a set of primary sampling units (PSUs) and in some cases secondary sampling units (SSUs) in order to control data collection costs. Children were mostly sampled via registered births from the National Center for Health Statistics (NCHS) vital statistics system. The core ECLS-B sample comprises births sampled within 96 PSUs and represents all infants born in the United States in the year 2001. In addition, to provide adequate precision in the American Indian/Alaska Native analysis domain, a supplementary sample of 18 PSUs was selected from a frame comprising areas where the population has a higher proportion of American Indian/Alaska Native births.

The ECLS-B target population consists of all children born in the United States in the year 2001 with the following exceptions: Children born to mothers less than 15 years of age, children who died before the 9-month assessment, and children who were adopted prior to the 9-month assessment.

The ECLS-B was designed to support statistical analyses for the following analytic domains that were derived from information on the sampled birth certificates:

- Child's race/ethnicity:
  - American Indian/Alaska Native;
  - Chinese;
  - Other Asian or Pacific Islander (i.e., excluding Chinese);
  - Hispanic;
  - Black, non-Hispanic; or
  - White, non-Hispanic;
- Birth weight:
  - Very low (less than 1,500 grams);
  - Moderately low (1,500 to 2,500 grams); or
  - Normal (greater than 2,500 grams); and
- Plurality:
  - Twin; or
  - Nontwin (single birth and other multiple births)

These domains were cross-classified to define 36 distinct birth certificate sampling strata, or case strata. Six strata required oversampling: the American Indian, Chinese, and Other Asian/Pacific Islander groups; those with very low birth weight and those with moderately low birth weight; and twins.

Over 14,000 births were sampled and fielded in the first wave. This sample size was designed to produce survey estimates with specified precision both overall and for specific analytic domains. The sample of about 14,000 births yielded 10,700 cases for the 9-month wave that had a completed interview with the child's parent, and assessments were conducted with 10,200 of these children. These 10,700 cases were fielded in the 2-year wave, and of these, 9,850 parent interviews were completed and 8,950 children were assessed. The third wave followed all sample children with a completed parent interview at the 2-year wave and 9-month wave respondents sampled as American Indian/Alaska Natives who did not respond at the 2-year wave. Of these 9,900 cases, 8,950 yielded a completed interview with the child's parent at the preschool wave. Of these 8,950 cases, 8,750 cases had a completed child assessment.

ECLS-K:2011

A brief overview of the ECLS-K:2011 is presented below. For more information on the study design and cohort profile, see Tourangeau et al. (2019; 2015).

Population represented by sample at baseline and sample design. The ECLS-K:2011 follows a nationally representative sample of children from kindergarten through elementary school. To achieve such a representative sample, a three-stage process was used. Firstly, 90 geographic areas (i.e., counties[3]) across the country were sampled as PSUs from a sampling frame of 3,141

[3] Note that larger counties in the sampling frame were treated as discrete PSUs while smaller contiguous counties were combined into PSUs.

counties in the United States. Of these 90 PSUs, 10 PSUs larger in size were sampled with certainty, the remaining PSUs were instead sampled using a stratified sampling procedure: they were grouped into 40 strata defined by Metropolitan Area status, census geographic region, size class, per capita income, and race/ethnicity of 5-year-old children. Two PSUs were then selected in each stratum.

The second stage of sampling involved selecting public and private schools with kindergarten programs (or that had kindergarten-aged children enrolled). In order to sample schools, separate sampling frames for public schools and for private schools were built using the school frames constructed for the 2010 National Assessment of Education Progress (NAEP). Again, schools were selected - within the 90 PSUs - using probability sampling proportional to size. Moreover, small schools were clustered for sampling in order to give them a better chance of being selected.

The third stage of sampling involved selecting children within sampled schools. Since the goal of the sample design was to obtain an approximately self-weighting sample of children[4], two independent sampling strata were formed within each selected school – one for API children and one for all other children; then, within each stratum, children were selected using equal probability sampling.

Once the children were sampled, parents were contacted to obtain parental consent for the child to be assessed at schools by trained assessors and for the parents themselves to be interviewed by phone. Moreover, information about teachers and schools was collected using a pencil and paper survey administered to teachers and school administrators.

Baseline cohort and follow-up waves. At baseline, during the 2010-2011 school year, data was collected for a sample for 18,174 children (15,953 from public school and 2,221 from private schools) from a sample of eligible children of 20,234 (survey response rate of 95%). In terms of schools sampled, a total of 1,352 schools were originally samples (1,052 public and 300 private). The 18,174 children that were part of the study at baseline attended around 1,310 schools, of which approximately 970 were schools that were originally sampled and approximately 340 schools to which sampled children transferred during the base year (between the Fall and Spring data collection). Subsequent waves were conducted as the children progressed through Fifth grade (see Table 2.4).

[4] The exception being Asians, Native Hawaiians, and Other pacific Islanders (APIs) that needed to be oversampled.

Figure 2.4: ECLS-K:2011 data collection waves

School year	Grade[1]	Data collections[2]
2010–11	Kindergarten	Fall 2010
		Spring 2011
2011–12	First grade	Fall 2011
		Spring 2012
2012–13	Second grade	Fall 2012
		Spring 2013
2013–14	Third grade	Spring 2014
2014–15	Fourth grade	Spring 2015
2015–16	Fifth grade	Spring 2016

Figure taken from (Tourangeau, et al., 2015)

Note that there are two rounds of data collection (Fall 2011, First grade and Fall 2012, Second grade) that include only a subsample of the original ECLS-K:2011 sample roughly 1/3 of the full sample. This subsample, sometimes called Fall sample, was selected through a three-step procedure: 1) 30 PSUs were sampled from the full list of 90 PSUs included in the study,[5] 2) then, all eligible schools within the sampled PSUs with children sampled in the base year were selected, and 3) children in these schools that were respondents in the base year and who had not moved outside of the country were included.[6]

ECLS-K: 1998

Under construction

[5] Of these 30 PSUs, 10 were the large PSUs included in the ECLS-K:2011 with certainty; the other 20 were selected by equal probability sampling 20 strata from the original 40. Then, one PSUs was sampled within each stratum also with equal probability.

[6] Note that a subsample of eligible students that had moved to another school in the same PSU or another samples PSU were followed in their new school. For students that moved to a new school outside of the sampled PSU, at attempt was made to complete a parent interview, but not a child assessment.

3. DICE approach to weights construction

3.1. Overview

To deal with non-random sampling design and unit non-response (e.g. between-wave attrition), DICE analyses use inverse probability weighting. Most of the datasets used in DICE are provided with a suite of longitudinal weighting variables and other survey design variables, such as stratum and PSU indicators, along with detailed guidance on their application. Where possible, we follow this guidance using the svy set of commands in the Stata software. Occasionally, pre-constructed weights provided in the survey datafiles are either non-existent or insufficient for the intended analyses.

Two surveys – the Netherland's Gen-R and Japan's LSN – do not contain any pre-constructed weights at all, so we construct our own and apply these in order to adjust for longitudinal attrition. That is, our method is designed to re-weight the sample remaining at the target analysis wave so it is representative of the sample observed in the initial baseline wave.

On two other occasions, we construct supplementary weights that are combined with the official survey weights to address specific issues of under-representation. First, analyses involving linked administrative data from the UK's National Pupil Database (NPD) to the Millennium Cohort Study have a target population of cohort members who attend state schools in England only, a restricted sub-sample of the full MCS sample, which includes cohort member in independent (private) schools in all four countries of the UK. Supplementary weights based on the probability of inclusion in the NPD linked dataset were developed and combined with the pre-constructed MCS weights designed for application to the target population as a whole. Second, for German NEPS-SC1, selective survey participation at the original baseline wave led to sample distributions of socio-demographic characteristics that deviated notably from official statistics on the target population. Raking weights, constructed using micro census data from the German Federal Statistical Office, were developed to calibrate the original NEPS sample to the national population.

Below, we apply an outline of the procedure used to generate longitudinal weights for Gen-R, the LSN and the MCS-NPD linked data, followed by further dataset specific-details. The procedure used to generate raking weights for the NEPS-SC1 concludes the section.

3.2. DICE-constructed longitudinal weights

The methodological literature mentions various approaches to non-response adjustment in longitudinal studies (see e.g., (Little & Rubin, 2002)), one of which is Inverse Probability Weighting (IPW). This technique works by giving more weight to those observations provided by cases whose predicted probability of being productive is low (e.g., Seaman, White, Copas, & Li (2012)). These predicted probabilities are generated from the estimates of logistic regressions fitted to the probability of being productive.

Thus, DICE-constructed weights were constructed as follows:

1. Generate predicted probabilities by estimating a logistic regression fitted to the probability of being productive in the target wave, with predictors taken from a baseline wave. Multiple imputation was used to account for any missing item information in the baseline predictors. Predictors used in these regressions were

country-specific (more details on them are presented when discussing weights in the subsequent chapters referring to specific analysis samples).

2. Compute the inverse of the predicted probabilities.
3. Divide the sample into decile groups based on the inverse of the predicted probabilities, i.e., based on the predicted probability of non-response, (following Chen et al., 2012)
4. For each decile group, take the average observed response rate. The inverse of this average response rate is then used as the longitudinal stage weight (i.e. to weight baseline to target wave sample).
5. Where necessary, this weight was multiplied by prior weights, either provided or constructed, that were designed to adjust for non-representativeness of the stage baseline

Gen-R

This section provides details of the variables used in construction of longitudinal weights for the age 3, age 5/6, and age 9 waves of the Rotterdam birth cohort Generation R. The Generation R study is a population-based prospective cohort study from fetal life until young adulthood and is situated in Rotterdam, the second largest city in the Netherlands. Eligible mothers were those who were resident in the study area at their delivery date and had an expected delivery date from April 2002 until January 2006. In total, 9778 mothers were enrolled in the study. The overall response rate based on the number of children at birth was 61% (Kooijman et al., 2016). Mothers of 9749 live born children agreed to participate. For the included twin pairs, we randomly excluded one of each pair leaving a baseline sample of N=9,628. As with many longitudinal studies, sample attrition did not occur uniformly across social groups in Gen-R. The aim of the created weights is to weight back to the original baseline sample (N=9,628), i.e., to correct for selective attrition. (Calibrated weights are not constructed for Gen R because the baseline sample was not intended to be representative of the national Dutch population).

Separate sets of weights were constructed for the maximum potential sample at the following measurement occasions:

1) age 3 survey
2) age 3 lab visit
3) age 5 survey/ age 6 lab visit
4) age 9 lab visit

A household was considered to have participated in a given wave when there was a response to at least one question in the relevant measurement occasion(s). For the age 5 survey/ age 6 lab visit intermediate weight, households with a valid response at either age 5 or 6 (or both) were considered productive (coded 1 in the dependent variable for response). After creating intermediate weights for the age 5/6 - age 9 sample, we created a longitudinal weight for the age 9 sample to be able to weight back to the baseline sample (age 9 longitudinal weight=age 5/6 weight × age 9 weight). Thus, in total 5 sets of weights were created.

Table 3.1 Number of Respondents for the 4 measurement occasions (baseline N = 9628)

Measurement occasion	**Number of respondents**
Age 3 survey	4,958
Age 3 lab visit	5,052
Age 5/6 survey/lab-visit	Combined age 5/6 sample: 7,206 - Age 5 survey: 6,269 - Age 6 lab-visit: 6,607
Age 9 lab visit	5,784

Age 3 sample and age 5/6 sample weights

To deal with missing data in the covariates used to predict probability of response at the various measurement occasions, we used the Multiple Imputation Chained Equations (MICE) technique. In MI, missing data are replaced by data drawn from an imputation model. This is done M times, generating M complete datasets. The general idea behind MICE is to impute multiple variables iteratively via a sequence of univariate imputation models, one for each imputation variable, with fully conditional specifications of prediction equations: all variables except the one being imputed are included in a prediction equation. The mi impute chained command imputes variables in order from the most observed to the least observed (Stata manual).

We imputed the following variables at baseline, generating 25 complete datasets: age of mother, pre-pregnancy bmi, birthweight, gestational age at birth, education of mother, household income, ethnicity of the child, gender (See Figure 3.1). One fully observed predictor was included in the MICE model: partner status. We fit the logit model to predict productivity separately on each of the 25 imputed datasets and combined the results.

	Observations per m			
Variable	Complete	Incomplete	Imputed	Total
AGE_M_v2	9626	2	2	9628
BMI_0	7025	2603	2603	9628
WEIGHT	9541	87	87	9628
GESTBIR	9539	89	89	9628
EduM	8435	1193	1193	9628
INCOME	6684	2944	2944	9628
Ethn_child	8980	648	648	9628
GENDER	9625	3	3	9628

Figure 3.1. Imputed cases per covariate for inclusion in the logit response models age 3 and 5/6.

Productivity variables were defined that are equal to 1 for all cases in the "productive" samples and missing for all other cases for all previously listed measurement occasions. For the first three measurement occasions we included the following variables in the logistic regression models:

- education of mother at baseline

- household income at baseline
- partner status
- age of mother
- pre-pregnancy BMI
- child gender
- child ethnicity
- birthweight
- gestational age at birth

To reduce bias, we also included missingness indicator dummies for variables with > 2% missing as fully observed predictors in the logit model. This was the case for the following variables:

- education of mother
- household income
- pre-pregnancy bmi
- child ethnicity

To make the predicted response probabilities more precise in cases of correlated predictors, we included the following interaction term in the logit response model: low educated × non-white. In the models predicting response at the age 3 samples (survey and lab visit) we additionally added the following interaction terms (these were not included in the model predicting response at age 5/6 as they did not significantly predict response at age 5/6):

- low educated × no partner
- non-white × no partner

To further reduce possible bias, we included an interaction of the missing indicators to models predicting response at age 3 and age 5/6: missing education × missing income.

Age 9 sample weights

To create the age 9 intermediate weights, we predicted productivity at age 9 in relation to age 5/6 sample characteristics. We first employed MICE to impute missingness in the age 5/6 covariates (See Figure 2). The imputed covariates were then used to predict productivity at age 9. We again generated 25 complete datasets, this time additionally imputing education of mother and household income at age 5 of the child. The MICE model included three fully observed predictors this time: partner status, age of the mother, and child gender.

	Observations per m			
Variable	Complete	Incomplete	Imputed	Total
BMI_0	5387	1819	1819	7206
BMI5child	6604	602	602	7206
WEIGHT	7191	15	15	7206
GESTBIR	7152	54	54	7206
EduM5	6178	1028	1028	7206
INCOME5	5825	1381	1381	7206
Ethn_child	7038	168	168	7206

Figure 3.2. Imputed cases per covariate for inclusion in the logit response model age 9.

To predict productivity at age 9, we included the following (imputed) variables in the logistic regression models:

- education of mother at age 5
- household income at age 5
- partner status
- age of mother
- pre-pregnancy bmi
- child gender
- child ethnicity
- birthweight
- gestational age at birth
- BMI of the child at age 5

To reduce bias, we also included missingness indicator dummies for variables with > 2% missing as fully observed predictors in the logit model. This was the case for the following variables:

- education of mother at age 5
- household income at age 5
- pre-pregnancy BMI
- child ethnicity

To make the predicted response probabilities more precise in cases of correlated predictors, we included the following interaction terms in the logit response model:

- low educated × non-white
- low educated × no partner
- non-white × no partner

As these were all n.s. they were dropped from the final model. To further reduce possible bias, we included an interaction of the missing indicators: missing education × missing income. This was also non-significant and dropped from the final model.

For all measurement occasions, weights were then derived by taking the inverse of the predicted probability of response (1/pr.). We split the sample into deciles based on the predicted probability on non-response. The weights then become the inverse of the response rate in each of the decile cells (see Chen et al, 2012).

Comparing unweighted results in Table 3.2 for baseline versus the Age 3 survey, the age 3 lab visit, the age 5/6 sample, and the age 9 sample we see that attrition was higher among lower educated respondents. Comparison of the weighted percentages of these measurement occasions with the equivalent percentages from the baseline sample show that the attrition weights are effective in reducing the impact of differential attrition by highest parental education.

Table 3.2 Baseline Unweighted Percentages for Maternal Education Versus Unweighted and Weighted Percentages at Subsequent Measurement Occasions

	Baseline	**Age 3 survey**		**Age 3 lab visit**		**Age 5 survey / Age 6 lab visit**		**Age 9**	
	U.	U.	W.	U.	W.	U.	W.	U.	W.
Maternal education									
High	42.83	57.54	42.83	50.52	43.09	47.84	41.95	51.35	41.99
Middle	30.66	27.26	30.94	28.95	30.57	30.23	30.68	30.09	31.08
Low	26.51	15.19	26.22	20.53	26.35	21.93	27.37	18.55	26.93
Missings on maternal education	1,193	-	-	-	-	-	-	-	-
N	9,628	4,732	4,732	4,584	4,584	6,570	6,570	4,991	4,991

Note. U=unweighted; W=Weighted.

LSN21

The LSN21 does not contain weight variables. The Japanese team produced weight variable to adjust attrition from first wave (age 0) to other wave (age 1 to 4) using the following steps:

- Created "continuation" variable as dummy variable in order to recognize continuing to answer from first wave. The case of the respondent participated the survey at each wave is 1 and otherwise 0.
- Run logit regression for the probability of continuation from first wave in each wave. Explanatory variables[7] in the logit regression model are followed[8]:
    - Regional block (6-category)
    - City size (3-category)
    - Father' age at childbirth (9-category)
    - Mother's age at childbirth (6-category)
    - Father / Mother's nationality (dummy variable: Majority=1, Otherwise=0)
    - Main family work (7-category)
    - Respondent in the first wave (dummy variable: Mother=1, Otherwise=0)

[7] In order to be able to include all cases in the estimation, missing values of each variable are given a separate category code.

[8] We cannot use parental education background because this information is contained in wave 2.

- o Number of living together (6-category)
  - o Household income (5-category)
- Calculated to inverse probability of continuation from the result of logit regression as weight variable in each wave.

Table 3.3 shows the weighted and unweighted distribution of mother's age at first wave at each wave. Weighted distribution at ages 1 to 4 is much closer to first wave (age 0) distribution than unweighted distribution.

Table 3.3. Mother's age distribution in 2010 cohort (first wave)

	Unweighted					Weighted			
	age0	age1	age2	age3	age4	age1	age2	age3	age4
Under 19	0.87	0.59	0.51	0.44	0.39	0.9	0.9	0.91	0.92
20-24	9.28	7.91	7.55	7.01	6.64	9.33	9.33	9.33	9.35
25-29	28.88	28.5	28.38	28.09	27.93	28.85	28.86	28.86	28.82
30-34	36.79	37.86	38.15	38.62	38.9	36.76	36.75	36.75	36.75
35-39	20.81	21.66	21.87	22.25	22.46	20.79	20.79	20.77	20.79
40 over	3.37	3.47	3.53	3.59	3.68	3.36	3.36	3.37	3.37

To confirm income distribution (Table 3.4) and parental education distribution (Table 3.5 and 3.6), as same as Table 3.3, weighted distribution at each age is similar to first wave (age 0) distribution.

Table 3.4. Income[9] distribution in 2010 cohort (first wave, 10 thousand yen)

	Unweighted					Weighted			
	age0	age1	age2	age3	age4	age1	age2	age3	age4
1-359	19.18	17.87	17.53	16.96	16.56	19.2	19.21	19.23	19.23
360-559	34.42	34.64	34.51	34.57	34.38	34.39	34.39	34.38	34.39
560-849	28.05	29.13	29.61	30.02	30.4	28.04	28.02	28.02	28.03
850-	13.47	14.28	14.39	14.72	15.06	13.47	13.46	13.47	13.46
0 or missing	4.89	4.09	3.96	3.73	3.6	4.92	4.92	4.91	4.90

Table 3.5. Parental education distribution in 2010 cohort

	Unweighted					Weighted			
	age0	age1	age2	age3	age4	age1	age2	age3	age4
Missing	15.18	1.96	7.35	5.28	5.02	2.34	8	5.98	5.83
High	42.87	49.55	47.84	49.65	50.49	48.26	46.25	47.53	47.95
Middle	25.57	29.56	27.93	28.34	28.21	29.63	27.98	28.52	28.49
Low	16.38	18.93	16.87	16.73	16.28	19.78	17.76	17.97	17.74

[9] It is household income before tax at first wave (age 0) and have not imputed missing cases.

Table 3.6. Parental education distribution in 2010 cohort (exclude missing)

	Unweighted					Weighted			
	age0	age1	age2	age3	age4	age1	age2	age3	age4
High	50.54	50.54	51.64	52.41	53.16	49.41	50.28	50.55	50.91
Middle	30.15	30.15	30.15	29.92	29.7	30.34	30.42	30.34	30.25
Low	19.31	19.31	18.21	17.67	17.14	20.25	19.31	19.11	18.83

In the following country-specific subjections, were provide summary information on official weights and survey measures provided by the survey administrators, with references to the relevant documentation. Further details on the DICE-constructed weights are also provided.

MCS-NPD linked sample

Under construction

3.3. DICE-constructed raking weights for the NEPS-SC1

Weights in NEPS-SC1 take into account the sampling procedure (design weights) and are calibrated to the "total number of births per month per municipality" (Würbach et al. 2016: 10). Further, the NEPS data center estimated models on panel drop-out relying on information from previous interviews. The (inverse) probability of taking part is used to construct weights in later waves. Additionally, the NEPS data center does not only provide weights for parents, but also for children taking part in tests. As children are tested at home, in (almost) all cases there is a valid parent interview and most, but not all children take part in testing. Because of this, there are weights for parents and for "accordant participation probabilities π_{t1comp}" (Würbach et al. 2016: 8), too.

However, there is no adjustment to socio-demographic distributions known from official statistics (or some other reliable resources). As survey participation is selective, (unweighted and weighted) distributions of key socio-demographics noticeably deviate from official statistics (see Tab. 3.7 and 3.8 for the level of deviation).

Based on micro census data collected in 2016, the German Federal Statistical Office (Statistisches Bundesamt, short: Destatis) published data on families of children being 3 to below 6 years old. As children of SC 1 are born from February to June 2012 (addresses of children born in July 2012 were not used, see above) and DICE Analysis 2 uses data collected in 2015 and 2016 (waves 4 and 5), children are 3 to 4 years old. Therefore, micro census data from 2016 should fit.

The information available specifies whether children in this age group are raised by a single parent and whether children and/or their parents have an immigrant origin (Destatis 2017: 120, Tab. 5.1.1, line 53; 126, Tab. 5.3.1, line 54). Immigration background means that the child or at least one (co-residing) parent are a) non-German citizens and/or b) became naturalized and/or c) immigrated to Germany as Ethnic Germans ("(Spät-)Aussiedler") (Destatis 2017: 20). After sending a request to Destatis, the Leipzig PI Prof. Thorsten Schneider received the distribution of parents' highest school leaving degree based on micro census data form the same year and for children in the same age range (Destatis 2019). In addition, the share of children living in East Germany, including Berlin, as a fourth benchmark (Destatis 2020a) was used, as it is known that child care offers, family demographics and so forth widely differ between East and West Germany (Kreyenfeld et al. 2017; Destatis 2018).

Consequently, the distributions of three major demographics and one regional indicator are available.

Notes:

1. The definition of parental education and immigration status is not the same as that which is used in DICE. This should be no problem. For poststratification, variables were generated in a way fitting the definition of Destatis.

2. Children between 3 and below 6 years old in 2016 belong to different birth cohorts (2011 to 2013). Families with very short spacing between births have some higher probability to contribute to the totals. This should be only a "threat" if such very short spacing happens in a sizable quantity and (!) if short spacing varies measurably in one of the three characteristics.

The population totals on parental education, single parent, immigration status and region were used to rake the already existing NEPS weights (Kalton and Flores-Cervantes 2003: 86). Other terms sometimes used for *raking* are poststratification or iterative proportional fitting (DeBell and Krosnick 2009). Raking was done in Stata (version 15) with the ado-file IPFWEIGHT (Bergmann 2011): "ipfweight performs a stepwise adjustment (known as iterative proportional fitting or raking) of survey sampling weights to achieve known population margins. The iterative process is repeated until the difference between the sample margins and the known population margins is smaller than a specified tolerance value or the specified maximum number of iterations is obtained. Additionally, thresholds for maximum and minimum weighting factors can be specified as well as a simple replacement of missing values" (https://ideas.repec.org/c/boc/bocode/s457353.html).

Sometimes, the construction of weights leads to outliers. To prevent highly influential cases due to such outliers, DeBell and Krosnick (2009: 8) suggest *to top-code/ to trim weights* to 5 times of the average weights (for alternative thresholds and procedures used in different surveys see Van de Kerckhove et al. 2014). DeBell and Krosnick (2009: 8) also suggest to repeat raking after top-coding until weights fit population totals or repetition does not further improve weights. As a default, 100 iteration were used to trim the data.

In addition, very small weights to 1/5th of the average weights were trimmed, although DeBell and Krosnick (2009: 8) militate against using a lower threshold, as weights close to zero didn't bear the risk of distort population estimates. On the other side, "those with small weights make but minimal contributions" (Kolenikov 2014: 29). Therefore, both (i.e., upper and lower) thresholds were used.

How many cases ranged below 1/5th of resp. 5 times above the mean before trimming?

Wave 4, parent interview:	159 and 60 out of 2,474 cases
Wave 4, parent interview and child testing:	175 and 59 out of 2,320 cases
Wave 5, parent interview:	168 and 56 out of 2,376 cases
Wave 5, parent interview and child testing:	162 and 47 out of 2,025 cases

Tables 3.7 and 3.8 provide an overview on how well official NEPS weights, post-stratified as well as post-stratified and trimmed weights fit to key distributions of official statistics. The post-stratified weights lead to a perfect fit. After trimming, distributions are shifted in disfavor of low qualified, single and immigrant parents. They are the least likely to take part and therefore receive extra-ordinary large weights in the post-stratification procedure. Although 2.3 to 2.5% of weights were trimmed, distributions still resemble the official distributions.

*Table 3.7: Population totals and the marginal distribution in NEPS SC 1, **wave 4** (year: 2015, children's age: 37-39m), in the case of no weights at all, official NEPS weights, post-stratified (and trimmed) weights*

	population totals	unweighted		official NEPS weights		post-stratified weights	
	%	%	n	%	n	%	n
			parent took part in interview				
Highest school leaving degree							
none or low[1]	18.5	5.49	136	8.51	211	18.50	458
intermediate	29.1	17.47	433	24.79	614	29.10	720
entrance qualification for applied universities	10.3	7.99	198	9.15	227	10.30	255
general university entrance qualification	42.1	68.69	1,707	57.27	1,419	42.10	1,042
Single parent							
no	85.0	91.85	2,276	90.17	2,234	85.00	2,103
yes	15.0	8.15	202	9.83	244	15.00	371
Immigrant family							
no	62.8	73.41	1,819	76.40	585	62.80	1,534
yes	37.2	26.59	659	23.60	1,893	37.20	920
Region							
West Germany	80.2	79.34	1,966	80,18	1,987	80.20	1,984
East Germany	19.8	20.62	511	19,63	486	19.80	490
			parent took part in interview plus child was tested				
Highest school leaving degree							
none or low[2]	18.5	5.21	121	8.12	189	18.50	429
intermediate	29.1	17.21	400	25.21	586	29.10	645
entrance qualification for applied universities	10.3	8.09	188	9.16	213	10.30	239
general university entrance qualification	42.1	69.32	1,611	57.21	1,329	42.10	977
Single parent							

	population totals	unweighted		official NEPS weights		post-stratified weights	
	%	%	n	%	n	%	n
no	85.0	92.00	2,138	90.39	2,101	85.0	1,972
yes	15.0	8.00	186	9.61	223	15.0	348
Immigrant family							
no	62.8	74.48	1,731	76.68	1,786	62.80	1,457
yes	37.2	25.52	593	23.14	538	37.20	863
Region							
West Germany	80.2	79.26	1,842	80.14	1,862	80.20	1861
East Germany	19.8	20.74	482	19.86	462	19.80	459

Sources: Population totals: highest school degree (Destatis 2019), as distribution did not sum up to 100%, 0.1% were added to all categories except “intermediate”, here 0.2% were added; single parent (Destatis 2017: 120, Tab. 5.1.1, line 53), immigrant family (Destatis 2017: 126, Tab. 5.3.1, line 54), region (Destatis 2020a); doi:10.5157/NEPS:SC1:6.0.0, wave 4, own calculations.

[1] Including 6 parents with other degree, [2] Including 5 parents with other degree.
Note: Probabilities do not always sum up to 100%, because of rounding and/or missing values. To make sure that totals of highest school leaving degree add up to 100%, we added 0.1, 0.2, 0.1 and 0.1 % to the categories according to the above listed order.

Table 3.8: Population totals and the marginal distribution in NEPS SC 1, ***wave 5*** *(year: 2016, children’s age: ~48m), in the case of no weights at all, official NEPS weights, post-stratified (and trimmed) weights*

	population totals	unweighted		official NEPS weights		post-stratified weights	
	%	%	n	%	n	%	n
	parent took part in interview						
Highest school leaving degree							
none or low[1]	18.5	4.75	113	5.99	143	18.50	440
intermediate	29.1	17.26	411	24.90	593	29.10	691
entrance qualification for applied universities	10.3	7.64	182	9.21	219	10.30	245
general university entrance qualification	42.1	70.18	1,671	59.62	1,420	42.10	1,000
Single parent							
no	85.0	91.46	2,177	91.33	2,175	85.0	2.020
yes	15.0	8.53	203	8.49	202	15.0	356
Immigrant family							
no	62.8	74.17	1,766	76.92	1,832	62.80	1,492

	population totals	unweighted		official NEPS weights		post-stratified weights	
	%	%	n	%	n	%	n
yes	37.2	25.83	615	23.08	549	37.20	884
Region							
West Germany	80.2	80.01	1,905	82.21	1,957	80.20	1,906
East Germany	19.8	19.91	474	17.72	422	19.80	470
	parent took part in interview plus child was tested						
Highest school leaving degree							
none or low[2]	18.5	3.79	77	5.37	109	18.50	375
intermediate	29.1	17.69	359	25.86	525	29.10	589
entrance qualification for applied universities	10.3	8.08	164	9.77	198	10.30	208
general university entrance qualification	42.1	70.23	1,425	58.70	1,191	42.10	853
Single parent							
no	85.0	92.16	1,870	91.67	1,860	85.00	1,721
yes	15.0	7.84	159	8,33	169	15.00	304
Immigrant family							
no	62.8	75.75	1,537	77.26	1,568	62.80	1,272
yes	37.2	24,25	492	22.74	491	37.20	753
Region							
West Germany	80.2	79.37	1,679	81.51	1,654	80,20	1,624
East Germany	19.8	20.58	440	18.45	374	19,20	401

Sources: Population totals: highest school degree (Destatis 2019), as distribution did not sum up to 100%, 0.1% were added to all categories except “intermediate”, here 0.2% were added; single parent (Destatis 2017: 120, Tab. 5.1.1, line 53), immigrant family (Destatis 2017: 126, Tab. 5.3.1, line 54), region (Destatis 2020a); doi:10.5157/NEPS:SC1:6.0.0, wave 5, own calculations.

[1] Including 7 parents with other degree, [2] Including 5 parents with other degree.
Note: Probabilities do not always sum up to 100%, because of rounding and/or missing values. To make sure that totals of highest school leaving degree add up to 100%, we added 0.1, 0.2, 0.1 and 0.1 % to the categories according to the above listed order.

4. Coding of parental education

4.1. Parental education as the key stratification variable in DICE

There are a number of practical arguments for selecting parental education as the stratification variable in studies of childhood inequality, in preference to measures of income, occupational status or an SES index that combines all three. Education is a relatively stable measure of long-term access to social and economic resources. Particularly at the stage of the lifecycle when children are young, the current income of parents tends to fluctuate and is a noisy measure of lifetime earnings and hence of long-term access to social and economic resources. Measurement of occupational status is complicated when a large fraction of parents, and particularly mothers, are out of the labour force or combining work with caring responsibilities. Use of single indicator, as opposed to an index, enables examination of whether the correlation between components differs across countries in ways that are predictive of child development. For example, we can explore the possibility that income disparities between parental education groups are smaller in some countries (for example due to greater progressivity of the tax and transfer system), leading to smaller education-related gaps in children's material environments and ultimately in their developmental outcomes.

From a human capital perspective, education is viewed as in investment undertaken early in life that yields returns in the labour market, where the nature of those returns will depend heavily on national institutional and policy structures such as the fiscal and wage bargaining systems. In theory, therefore, DICE defines education groups of parents with common 'raw' levels of human capital and explores how that capital is rewarded differently in different contexts, in terms of the resources it generates both for parents and their children. This perspective is also helpful from a public policy point of view, in that it conceptualises education as a structural form of parental advantage and factors that are more responsive to short-term interventions, such as income and employment behaviours, as mechanisms by which that advantage is transmitted.

DICE is in alignment with several prominent strands of literature that focus on inequalities by parental education. Work on the 'diverging destinies' hypothesis from family demography explores how the disparity in resources available to children in different parental education groups is magnified in different times and places by its co-variation with maternal age, employment, single parenthood and parental time with children (McLanahan, 2004; Bianchi, 2000). Originally this work was primarily focused on the moderating role of cohort change in a single country – the US – but more recent work has built understanding of cross-national variation in education-related differences in family processes (Amato et al., 2016; Bernardi & Boertien, 2017). Maternal education is also widely (if often implicitly) used as the key measure of SES in developmental psychology studies, where it is consistently one of the strongest single predictors of a wide spectrum of developmental outcomes. The preference for maternal, as opposed to combined parental, education is likely due at least in part to tradition and ease of measurement, but it also reflects theoretical ideas about the importance of interactions between the child and the primary caregiver (traditionally the mother) and the influence of education on those interactions (Harding et al., 2015). Note that the only difference in categorization of families according to maternal or highest parental education occurs in families in which the father (or partner) has a higher qualification than the mother. In DICE we use the highest parental qualification to categorize families and so effectively assume the socioeconomic advantages associated with the presence of a more educated father outweigh the disadvantages of caregiver interactions with a less educated mother. Of course, as gender roles become more outdated, the assumption that mothers interact more with children than fathers becomes less tenable and the arguments for giving primacy to maternal education diminish.

4.2. Measurement of parental education in DICE

The key variable by which we analyse group-based inequalities is the highest qualification attained by a parent who is co-resident with the child at the time at which the main child outcome variable is measured.

In practice, this was achieved by firstly coding a parental education variable for the main parent/carer of the cohort member and, if present, for the secondary parent/carer of the cohort member. Then, the final DICE parental education indicator was coded by taking the highest value out of the two, or - if only one parent was present - the highest qualification attained by the only parent/carer present. If two parents were present, but the qualifications of one were unknown, we used the highest qualification of the other parent.

The harmonization of different systems of national educational qualifications is challenging. The International Standard Classification on Education (ISCED) is commonly used for this purpose but in the specific case of these six countries, ISCED levels tend to equate qualifications that have quite different implications for life chances and family resources in different countries. We therefore developed our own coding system to categorize parental education after extensive discussion between the national teams.

We define high education as a first/bachelor's university degree or higher, requiring 3-4 years of full-time study at the tertiary level, in all countries. The definition of low education differs between countries with comprehensive systems (i.e. little or no tracking below age 16; the US, UK, France, and Japan) and those with early tracking and a high degree of academic/vocational specificity (Germany and the Netherlands). For the first group, low education is defined as no qualification beyond the expected standard, i.e. the target of the education system for all children in compulsory education. In the US, Japan and France this is a high school diploma (Baccalauréat general in France); in the UK this is attainment of at least a grade C qualification at the end of compulsory schooling (age 16). In the second group, low education is defined as no attainment beyond the intermediate/junior secondary track. The medium education group is all those who do not fall in either the high or low categories. In the US, for example, this category would include those with some education beyond high school but without a bachelor's degree. Table 4.1. below gives details of the way specific national qualifications are allocated

A key issue in cross-national research is whether family background is defined in relative (within-country) or absolute criterion-referenced terms. Most research uses a relative approach in which advantage and disadvantage are denoted by membership of top and bottom quantile groups, for example in terms of income or a composite SES index. This approach is consistent with a theoretical perspective in which SES is a measure of one's ranking within a hierarchy and it bypasses knotty issues of how and where absolute thresholds should be drawn for different populations. A relative approach is harder to implement when education is the stratifying variable because qualifications are typically on an ordinal rather than continuous scale and the distribution can rarely be sliced into equal-sized quantiles across countries. In addition, there are conceptual reasons why a purely relative approach based on within-country rank is problematic when studying the moderating role of country context. It is not possible to tell whether cross-country differences in outcome gaps reflect a difference in the mechanisms that transmit parental resources to children (a genuine moderating or buffering effect) or whether they simply reflect differences in the composition of the top and bottom groups. To give an example, larger child outcome gaps between the top and bottom income quintile groups in the US than in other countries may occur because income matters more for children's life chances in the US (e.g. because of a greater role for the market), or because the two groups are respectively richer and poorer than their equivalents in other countries, or both (Bradbury et al., 2019). An absolute approach to the categorization of family background helps to clarify the mechanisms involved by eliminating the second of these explanations.

In reality, our approach to education categorization balances several principles and so is perhaps something of a hybrid between the relative and absolute perspectives. Our high education category distinguishes families in which a parent has the equivalent of a bachelor's degree, or 3 to 4 years of full-time tertiary education, in every country. The global nature of the market for postgraduate university study, where tertiary qualifications from many systems are now routinely accepted as proving eligibility for the same programme, provides some justification for the equivalence of this category. The division of non-degree educated families between the low and medium categories is harder to rationalise in this way. Here we felt it important that thresholds be meaningful within each national context so that as far as possible our definitions have face validity in relation to domestic research, policy and practice, as well as from a comparative perspective. For example, high school graduation in the US and Japan is a major life milestone and there is a clear social division between those who go on to acquire further qualifications and those who do not. Similarly, in the UK the decision of whether to continue in education past the compulsory age of 16 is widely perceived as a key transition point. In line with the relative perspective, we did consider the distribution of families across education groups when defining our thresholds so as to avoid highly unequal sizes both within and across countries. However, we did not massage thresholds purely to equalize proportions cross-nationally, and so allowed for absolute differences in national education distributions. An important advantage of this approach is that it means we can retain the same national coding systems across cohorts and datasets within a country, providing continuity in the reference groups as children age and average levels of education rise over time. Ultimately, any crude classification of families according to socioeconomic background will have its strengths and weaknesses. We believe our system has advantages in terms of transparency, theoretical interpretation and relevance to a range of national and international audiences, while potential weaknesses in harmonisation can be discussed ex-post in relation to specific patterns of results.

Unless explicitly stated in a specific paper, these definitions and operationalisation of DICE core variables apply to every paper in which they appear. These variables formed the core of the harmonisation effort carried out for the DICE project.

Table 4.1. National coding schemes for DICE parental education indicators

Country	**High education**	**Medium education**	**Low education**
France	Higher diploma (more than 2 yrs post-baccalauréat), equivalent to at least an undergraduate degree	Intermediate post-bac diploma (usually vocational, two-year courses)	General high school certificate ("baccalauréat") Technical or vocational high school diploma Lower secondary school certificate (CAP/BEP/equivalents) - GCSE level Lower secondary school certificate ("BEPC") - Middle school level Primary school certificate ("CEP") No diploma
Germany	University – traditional University – applied sciences	Secondary level II (university entrance qualification) + VET (standardized vocational training) Secondary level II (university entrance qualification) Secondary level I (intermediate) + VET	Secondary level I (intermediate) Secondary level I (low) + VET Secondary level I (low) No certificate
Japan	Graduate school University	Junior college or technical college Specialized vocational school (after graduated from senior high school) Other	Senior high school High school Special training school (after graduated from junior high school) Junior high/elementary school
Netherlands	University Education (WO) and post-hbo education Higher professional education (HBO) e.g. heao, higher agricultural & horticultural, hts, hbo-v, pabo	Senior secondary vocational education (MBO) e.g. meao, senior secondary agricultural & horticultural school and mts General secondary education (HAVO, VWO, HBS, MMS, lyceum, athenaeum, gymnasium)	Junior general secondary education (MULO, MAVO, LAVO) Pre-vocational secondary education (VMBO) e.g. lts, lhno, home economics

Country	High education	Medium education	Low education
			school, leao, junior secondary agricultural & horticultural school Pre-vocational education (VBO) Special secondary education (VSO-LOM, VSO-MLK) School for the physically or visually handicapped or with hearing impairment Special primary education (MOK,MLK,ZMLK) Special primary education (BLO,LOM) Primary education Other
UK	Higher degree First degree Professional qualifications at degree level	Diplomas in higher education Nursing or other medical qualifications A/AS/S level NVQ/SVQ/GSVQ Level 3 Trade apprenticeship	O-level/GCSE grades A-C NVQ/SVQ/GSVQ Level 2 O-level/GCSE grades D-G NVQ/SVQ/GSVQ Level 1 Other academic or vocational qualifications, including overseas None of these qualifications
US	Professional degree after bachelor's degree (MD, DDS, JD, LLB,etc.) Doctorate degree (PHD, EDD) Master's degree (MA, MS) Graduate or professional school but no degree Bachelor's degree	Associate's degree Some college but no degree Voc/tech diploma after high school Voc/tech program after high school but no voc/tech diploma	High school diploma or equivalent 12th grade but no diploma 1st-11th grades No formal schooling

5. Other core harmonised variables

<< Under construction >>

Child age at time of testing

DICE TARGET DEFINITION: age of cohort member at the time of interview or at the time of assessment (recorded in months).

Note that sometimes this information was directly available in the data used, other times it had to be computed using the date of birth and the date of interview/assessment.

Child gender

DICE TARGET DEFINITION: binary variable - the cohort member is male (=0) or female (=1).

Mother's age at child's birth

DICE TARGET DEFINITION: 5-category variable - <20, 20-24, 25-29, 30-34, 35 or more.

Note that sometimes this information was directly available in the data used, other times it had to be computed using the date of birth of the cohort member and the date of birth of the mother.

An important note is necessary to clarify who exactly was identified as the 'mother' as three alternative operationalisation of the concept were available:

BIOLOGICAL MOTHER: biological parent who is female

MOTHER: biological, adoptive, foster, step-parent who is female

SOCIAL MOTHER: main parent/carer of the cohort member (might be mother, father, grandparent, sibling, other relative)

The operationalisation 1 was preferred for all research done on infancy for which it was particularly important to be able to identify teenage pregnancies.

For papers focusing on later childhood and adolescence period, operationalisation 2 was used – unless differently noted.

Family structure

DICE TARGET DEFINITION: the aim of the employed common operationalisation was to code family structure as shown below:

Child lives with two biological/adopted/foster parents

Child lives a single biological/adopted/foster parent only (no cohabitant partner, could be in a non-cohabiting relationship)

Child lives with one biological/adopted/foster parent and one other parent (e.g., step-families)

Other (child not residing with a biological/adopted/foster parent)

Children in shared custody arrangements are treated as living with two parents.

Parents living away from the child for work-related reasons are considered non-resident parents (so most such families would end up in category 2).

This operationalisation allowed, whenever deemed relevant and appropriate, to code a binary indicator identifying single biological/adopted/foster parent only (=1) versus all other types of families (=0). This is, of course, always mentioned in the specific papers in which family structure is conceptualised as such.

Immigration/foreign-born parent

DICE TARGET DEFINITION: binary variable - at least one parent/carer who co-resides with the child at the time of outcome measurement was born abroad (=1) versus no co-residing parent/carer was born abroad (=0).

Notes for specific countries:

In **France**, this indicator was operationalised as ‘parent/carer who co-resides with the child at the time of outcome measurement does not have French nationality.’

For the **US**, those born in the US Territories (Puerto Rico, Guam, American Samoa, US Virgin Islands, Mariana Islands, Solomon Islands) were treated as "born in the US" (coded 0). This indicator is missing for a large sub-set (approx. 23.7%). This seems to be due to a large number of missing cases for resident father's data. While approximately 6,900 respondents reported that they live with a partner/spouse at the time of survey, only approximately 5,250 resident father's provided their country of birth (approx. 24% missing).

Bibliography

Amato, P. R., Booth, A., McHale, S. M., & Van Hook, J. (2016). *Families in an era of increasing inequality.* . Springer International Pu.

Bernardi, F., & Boertien, D. (2017). Non-intact families and diverging educational destinies: A decomposition analysis for Germany, Italy, the United Kingdom and the United States. *Social Science Research, 63,* 181-191.

Bianchi, S. (2000). Maternal employment and time with children: Dramatic change or surprising continuity? *Demography, 37*(4), 401–414.

Charles, M. A., Thierry, X., Lano, J.-L., C. B., & Marie-Noelle Dufourg, R. P. (2020). Cohort Profile: The French national cohort of children (ELFE): birth to 5 years. *International Journal of Epidemiology, 49*(2), 368-369j.

Chen, Q., Gelman, A., Tracy, M., Norris, F. H., & Galea, S. (2012). *Weighting adjustments for panel nonresponse.* . Columbia University. , New York, NY. Retrieved May 7, 2022, from http://stat.columbia.edu/~gelman/research/unpublished/weighting%20adjustments%20for%20panel%20surveys.pdf

CLS. (2020a). *Millennium Cohort Study. User Guide (Surveys 1-5). 9th Edition.* London: Centre for Longitudinal Studies. UCL.

CLS. (2020b). *Millennium Cohort Study. Sixth Survey 2015-2016. User Guide. 2nd Edition.* London: Centre for Longitudinal Studies. UCL.

CLS. (2020c). *Millennium Cohort Study. Age 17 Sweep (MCS7). User Guide.* London: Centre for Longitudinal Studies. UCL.

de Onis, M., Onyango, A. W., Borghi, E., Siyam, A., Nishida, C., & Siekmann, J. (2007). Development of a WHO growth reference for school-aged children and adolescents. *Bulletin of the World Health Organization, 85,* 660-667. doi:https://doi.org/10.1590/S0042-

Harding, J. F., Morris, P. A., & Hughes, D. (2015). The relationship between maternal education and children's academic outcomes: A theoretical framework. *Journal of Marriage and Family, 77*(1), 60-76.

Ipsos MORI. (2017). *Millennium Chort Study Sixth Sweep (MCS6). Technical Report. 2nd Edition.* Ipsos MORI.

Ipsos MORI. (2019). *Millennium Cohort Study. Seventh Sweep (MCS7).*

Jaddoe, V. W., Mackenbach, J. P., Moll, H. A., Steegers, E. A., Tiemeier, H., Verhulst, F. C., . . . Hofman, A. (2006). The Generation R Study: Design and cohort profile. *European Journal of Epidemiology*.

Jaddoe, V. W., van Duijn, C. M., Franco, O. H., van der Heijden, A. J., van Ijzendoorn, M. H., de Jongste, J. C., . . . al, e. (2012). The Generation R Study: design and cohort update 2012. *European Journal of Epidemiology*.

Jaddoe, V. W., van Duijn, C. M., van der Heijden, A. J., Mackenbach, J. P., Moll, H. A., Steegers, E. A., . . . Hofman, A. (2008). The Generation R Study: design and cohort update until the age of 4 years. *European Journal of Epidemiology*.

Jaddoe, V. W., van Duijn, C. M., van der Heijden, A. J., Mackenbach, J. P., Moll, H. A., Steegers, E. A., . . . Hofman, A. (2010). The Generation R Study: design and cohort update 2010. *European Journal of Epidemiology*.

Kooijman, M. N., Kruithof, C., van Duijn, C. M., Duijts, L., Franco, O., van IJzendoorn, M. H., . . . Rivadeneira. (2017). The Generation R Study: design and cohort update. *European Journal of Epidemiology*, 1243-1264.

LIfBi. (2020). *Study Overview. NEPS Starting Cohort 2 - Kindergarten. From Kindergarten to Elementary School. Waves 1 to 9.* Bamberg: Leibniz Institute for Educational Trajectories. Retrieved from https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC2/9-0-0/SC2_Overview_W1-9.pdf

Little, R., & Rubin, D. (2002). *Statistical Analysis with Missing Data.* New Jersey, NJ: Wiley.

McLanahan, S. (2004). Diverging destinies: How children are faring under the second demographic transition. *Demography, 41*(4), 607-627.

Seaman, S., White, I., Copas, A., & Li, L. (2012). Combining multiple imputation and inverse-probability weighting. *Biometrics, 68*(1), 129-137.

Skopek, J., Pink, S., & Bela, D. (2012). *Data Manual Starting COhort 2: From KLindergarten to Elemntary School.* Bamberg: NEPS Data Center.

Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., Hagedorn, M. C., & Daly, P. (2015). *Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011). User's Manual for the ECLS-K:2011 Kindergarten Data File and Electronic Codebook, Public Version.* U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Tourangeau, K., Nord, C., Lê, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L., & Njarian, M. (2019). *Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011) User's Manual for the ECLS-K:2011 Kindergarten–Fifth Grade Data File and Electronic Codebook, Restricted Version (NCES 2019-101).* U.S. Department of Education. Washington, DC: National Center for Education Statistics.