

The Intergenerational Transmission of Skills dataset

Citation for published version (APA):

Jacobs, B., Vermeulen, S., & van der Velden, R. (2021). *The Intergenerational Transmission of Skills dataset*. ROA. ROA Technical Reports No. 007 <https://doi.org/10.26481/umarot.2021007>

Document status and date:

Published: 11/10/2021

DOI:

[10.26481/umarot.2021007](https://doi.org/10.26481/umarot.2021007)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

The Intergenerational Transmission of Skills dataset

Babs Jacobs
Stan Vermeulen
Rolf van der Velden

ROA Technical Report

ROA-TR-2021/7

Researchcentrum voor Onderwijs en Arbeidsmarkt | ROA
Research Centre for Education and the Labour Market | ROA

The Intergenerational Transmission of Skills dataset

Babs Jacobs
Stan Vermeulen
Rolf van der Velden

ROA-TR-2021/7
October 2021

Research Centre for Education and the Labour Market
Maastricht University
P.O. Box 616, 6200 MD Maastricht, The Netherlands
T +31 43 3883647 F +31 43 3884914

secretary-roa-sbe@maastrichtuniversity.nl
www.roa.nl

ISSN: 2666-884X

Table of Contents

1. Preface	1
2. Education cohorts	2
2.1 <i>Sampling</i>	2
2.2 <i>Content of the education cohort studies</i>	3
2.3 <i>School performance test</i>	4
2.4 <i>Intelligence test</i>	4
3. Register information	6
3.1 <i>Legal parents and children</i>	6
3.2 <i>CITO-test</i>	6
3.3 <i>Other education information</i>	7
3.4 <i>Financial information</i>	8
3.5 <i>Background information</i>	9
4. ITS dataset: combining survey and register data	9
4.1 <i>Descriptive statistics</i>	11
4.2 <i>Representativeness</i>	15
5. References	18

1. Preface

This technical report describes the construction of the Intergenerational Transmission of Skills (ITS) database. This database is created within the ITS program founded and coordinated by Rolf van der Velden. The aim of this program is to get a deeper understanding of the role of intergenerational transmission of skills in various life domains (i.e., education, health, crime etc.). The foundation for the ITS database is provided by linking Dutch cohort survey data gathered in the 1970's and 1980's containing the skill measures of the parent generation, to register data available at Statistics Netherlands, in particular the NCO database (for more information, please click [here](#)), which includes children's educational career and current characteristics of their parents and the CITO database, which includes the skill measures of the children's generation

This technical report provides a detailed overview of the construction of this database. First, information is provided on the education cohorts that started in the 1970's and 1980's. This description includes the sampling of the cohorts, content of the survey, information on the school performance test and intelligence test. Then, we describe the register information. Detailed information on the CITO-test is provided, as well as information on the education registers. We also include information on the financial situation and background. Finally, the construction of the database is explained. Here we include descriptive statistics on the database and demonstrate the representativeness of the data. In the future, this database will continue to expand with new data. This will be described in follow-up versions of this technical report.

We gratefully acknowledge a data grant received from the Dutch Ministry of Education, Culture and Science and the Netherlands Initiative for Education Research (NRO: grant 405-17-900) to construct this database.

2. Education cohorts

Since the mid-1970s, education cohort studies were undertaken by Statistics Netherlands on a regular basis. For the creation of the ITS dataset, we made use of the three education cohorts that started in the 1970's and 1980's. We linked the data from these education cohorts to register data on original respondents as well as their offspring. In the following, we will discuss the sampling procedure and content of the education cohorts.

2.1 Sampling

In 1977, Statistics Netherlands started the cohort study, called "Sociaal Milieu Voortgezet Onderwijs" (SMVO) (English translation: 'social origin and secondary education'). This cohort study is a national representative panel of children that entered secondary education for the first time, around the age of 12, in school year 1977/'78. In this longitudinal survey, students were followed throughout their educational career. A stratified two-stage cluster sample was drawn. Six strata were distinguished to avoid underrepresentation of certain tracks. Within each track, the aim was to draw at least 700 students per social origin group and gender (CBS, 1982). A sample of school classes was then drawn per stratum. The total sample consists of 37,280 students from 1,275 schools. This is about 15 percent of the total student population at the time. Analyses on the overall non-response showed that the distribution of students across tracks was representative. The non-response across regions was not entirely random: Partial non-response was more likely among underachievers, dropouts as well as among students with a relatively low teacher advice or with low-educated parents (CBS, 1982, 1991).

The second cohort study "Schoolkeuze Lager Voorgezet Onderwijs" (SLVO) ('school choice primary and secondary education') started in school year 1982/'83 in the last year of primary education. A random sample was drawn from the population of 8,745 schools. From the 977 schools selected, 669 schools were willing to participate. In total, 16,813 students were included in the study. This amounts to almost 8 percent of the primary school student population at that time. Again, the regional non-response was not random: Schools in the four biggest cities in the Netherlands were less willing to participate (CBS 1991). As in the 1977 cohort, the partial non-response occurred for variables that negatively affect school careers (i.e., disadvantaged background, low performance scores or low teacher advice).

The third cohort study "Voortgezet Onderwijs Cohort Leerlingen" (VOCL) ('secondary education cohort study students') is a nationally representative panel of children who were born in 1977 and entered secondary education for the first time in the school year 1989/1990. The sampling unit were schools in regular secondary education. In small schools all classes were included, and in bigger schools a random selection of classes was sampled (Kuyper & van der Werf, 2007). The survey comprises 19,524 students from 381 schools, amounting to 10.5% of the entire student population in the first year of secondary school at the time. The sample deviates only slightly from the total population of schools. There are small deviations present in the distribution of urbanization (large cities are slightly underrepresented) and school types (Driessen & van der Werf, 1992).

2.2 Content of the education cohort studies

In all education cohort studies, the position of the students in education was determined annually (i.e., track and grade). In the first year, additional information was gathered on parents' level of schooling and socio-economic status, children's performance scores in math, language, and nonverbal intelligence. In the 1989 education cohort, this additional information was not only collected in the first year, but also in the third and fifth year of the survey. As in the first year, students in the third year of secondary education had to do performance tests and an intelligence test. Moreover, in the fourth cohort year the same information was collected among children who were delayed (i.e., repeated a year). An extensive questionnaire was administered among students in the two highest tracks of secondary education in the last two or three years of secondary education. This questionnaire contained questions on the choice of subjects, future prospects, study skills and homework behavior. Another difference compared to the other education cohorts is that the students completed questionnaires on motivation and school experiences in the first and third year of the survey. Moreover, Statistics Netherlands has information on the students who left education without a diploma from the Dutch Labour Force Survey ('EBB'). For more detailed information on the cohorts, see CBS (1982), CBS (1991), Blommers (1983); Hustinx, Kuyper, van der Werf, and Zijlsling (2005), Kuyper and van der Werf (2007).

2.3 School performance test

In the first year of data gathering, students did a school performance test. To test the performance in math and language, test developer CITO (in English ‘Central Institute for Test Development’) developed a short performance test. The test was conducted in classroom and administered by the teacher. It was constructed with existing test items from the nation-wide CITO-test. More information on this nation-wide CITO-test is presented in paragraph 3.2.

In the 1977 education cohort, the performance test lasted two hours. For mathematics, 25 items were included, for language 45 items. The number of correctly answered questions gives the total test score. Even though all participating schools took part in the testing, in some cases it was impossible to link the test scores to individual students due to missing or incorrect information for identification on the test forms (CBS, 1982). The share of non-linkable data ranges from 15 percent in the higher track to 36 percent in the lower track. This illustrates that the sample might not be representative for the lowest performance group (specifically for students in special education).

In the 1982 and 1989 education cohorts, the test consisted of three elements, i.e., language, math, and information processing. Each element consisted of 20 multiple choice questions. The students in the 1982 cohort had 35 minutes to complete the items for math and language, and 45 minutes for the information task. In the 1989 education cohort, the total time for the test was 1 hour and 35 minutes. The share of missing data is negligible in the 1982 cohort, while in the 1989 cohort 7.8 percent of observations had missing test data in either the math or language domain. The test scores were standardized to mean zero and unit variance within each cohort using the full original cohort.

2.4 Intelligence test

The so-called TIB-test (‘Test di Intelligenza Breve’) was administered for the 1977 education cohort. This test consists of 33 items including five figures per item. For each item, the student had to choose the figure that was different from the other four. Similar to the school performance test, linking the intelligence test scores, was in some cases problematic due to missing or incorrect information for identification on the test forms.

Two elements of the German PSB test (‘Prüfsystem für Schul- und Bildungsberatung’) were administered among the 1982 and 1989 cohorts. It concerns two non-verbal subtests of the PSB, namely PSB-3 (reasoning) and PSB-8 (abstraction). Both subtests consist of 40 items and

the students need to finish as many items as possible in a certain time slot. In order to maximize comparability between cohorts, we used the scores on the reasoning subtest for the 1982 and 1989 as our measure of intelligence. The intelligence test scores are only available for the parents at age 12, not for their offspring.

2.4 (Grand)parent questionnaire

To gather background information on the family, parents of the respondents (hereafter referred to as grandparents as these are the grandparents of the respondents' offspring) filled in a questionnaire. From this, grandparental characteristics are available in terms of highest level of schooling and social status. The highest educational level of both the grandmother and grandfather was assessed using a set of questions with a closed response format. In the 1982 cohort, some upper secondary education qualifications were not clearly included in the response options, which resulted in underreporting in this category.

The social status of the grandparents was measured in broad categories, differentiating blue collar workers, employer without staff, employers with staff, lower white-collar workers, middle white-collar workers, professionals and other (unemployed and out of labor force). This social status was constructed using multiple questions from the questionnaire. First, respondents were asked whether they worked. If not, their main status was asked. If he or she did work, the status in employment (i.e., self-employed, employed or working in family business), occupation, economic sector in employment and supervisory status was asked. The grandparent who filled in the questionnaire was asked to fill in these questions also for his or her partner.

3. Register information

We link the data from the education cohorts to register data from various domains. This is possible because the data of the survey cohorts include unique identifiers for every individual that can be used to link the survey data to other datasets available at Statistics Netherlands. For the latest cohort we could rely on an existing unique personal identifier, which made the linking process successful in 98 percent of the cases. For the 1977 and 1982 cohorts the identifying information was based on name, address, place of residence, gender, and birth date of the respondents. This information allowed us to link 81 percent (1977 cohort) and 88 percent (1982 cohort), respectively, of the original students to existing identifiers in the registers.

3.1 *Legal parents and children*

The unique feature of the ITS dataset is that these identifiers make it possible to link the original cohort data to the data of their (legal) children. This is possible using a database at Statistics Netherlands containing all individuals in the municipal personal records database. This database links children to their legal parents.^{1 2} For most children, they can be linked to either their mother or their father from the education cohort surveys. For a small number of children (N=365), both parents were present in one of the education cohort surveys.

3.2 *CITO-test*

The identified children are then linked to the available CITO-test information. Statistics Netherlands has this information from school year 2005/06 till 2018/19.³ The test is a nation-wide standardized skills test. It measures what a student has learned in comparison to other students in eight years of primary education. The content of the test is closely related to the curriculum for the basic skills of language and arithmetic, and is constructed on the expectation that these are highly predictive for a successful completion of tracks in secondary education.

-
1. Note that the legal parents do not necessarily have to be the biological parents and that legal parents do not necessarily live in the same household as their children.
 2. This micro dataset is called 'KINDOUDERTAB'
 3. This micro dataset is called 'CITOTAB'. In the year 2019/20 the test was not taken due to COVID-19.

From school year 2014/2015 onwards, it is compulsory for students to make a high-stakes test in the last year of primary education (CvTE, 2015). Since then, the government makes the CITO-test available to all schools (more than 80% of the schools administer the CITO-test), but schools can also choose another test supplier approved by the Minister of Education, Cultural Affairs and Science for this final test.

In the Netherlands, track allocation to secondary education is based on the score of the national test as well as the primary school teacher's advice. All test scores are converted into a 'test advice', indicating the track most suitable for the student. From school year 2014/2015 onwards, the teacher gives a recommendation for the track placement in secondary education before the high-stakes test is administered. In case the track placement based on the CITO-test score is higher than the teacher's recommendation, the school needs to reconsider the advice. The school is responsible for doing this in consultation with the parents or legal guardians.

Students are tested on their language skills, math skills and world orientation. Before the school year 2014/2015, there was a separate subdomain called study orientation (Van Boxtel, Engelen, & de Wijs, 2011). From 2014/2015 onwards this has been included in the tests for language skills and arithmetic. World orientation is a non-compulsory component. The total scores per subdomain vary over the years, with a total score for language skills ranging from 100 in 2009/2010 to 135 in 2014/2015, and a total score for math skills ranging from 60 in 2009/2010 to 85 from 2014/2015 onwards.

Before school year 2014/2015, the test was administered in February, but from school year 2014/2015 onwards this was done between 15 April and 15 May (the exact dates vary yearly). In total, the test takes three consecutive days. In the ITS dataset, the total scores on language and math were standardized to mean zero and unit variance within the school year, using only the schools for which information was available for all the years that were included in the data.

3.3 Other education information

Additional information on the educational pathways of the children was added from the Netherlands Cohort Study on Education (NCO) (for more information see Haelermans et al. (2020)) and register information on enrollments in education.⁴ These datasets include annual

4. This micro dataset is called 'ONDERWIJSINSCHRTAB'.

information on the student's track placement, grade and diploma. The NCO dataset also includes detailed information on the attended school and background information on the individual and their parents. The NCO data is available from schoolyear 2007/08 onwards, the register information on enrollments in education from school year 2000/01 onwards.

For all children, we converted the teacher advice as well as the initial track in secondary education into years of schooling using the so-called educational ladder by Bosker and Velden (1989). The following categories are distinguished: 6 years for university preparatory education (in Dutch 'voorbereidend wetenschappelijk onderwijs' (vwo)), 5 years for senior general secondary education (in Dutch 'hoger algemeen vormend onderwijs' (havo)), 4 years for the highest track within preparatory vocational secondary education (in Dutch 'vmbo gemengde en theoretische leerweg' (vmbo gl/tl)), 3 years for second highest track within preparatory vocational secondary education (in Dutch 'vmbo kaderberoepsgerichte leerweg' (vmbo kb)), 2 years for lowest track within preparatory vocational secondary education (In Dutch 'vmbo basisberoepsgerichte leerweg' (vmbo bb) and 1 year for practical education (in Dutch 'praktijkonderwijs'). For the children who are not yet in a single track, because they are still in a mixed track in secondary education, we use the average years of schooling of the tracks they still have direct access to, given the mixed track they are in. For example, children within the mixed track havo/vwo get 5,5 years of schooling.

3.4 Financial information

Information on the parents' financial situation three to one years before the child took the CITO-test was added. Specifically, we have information on parents' personal income, household income, and household wealth from the administrative data of Statistics Netherlands.⁵ Parental personal income is based on the percentile of the parent in the Dutch income distribution (sources of income include income from labor, owned companies, unemployment benefits and social security). Household income is operationalized by the percentile of the household in the Dutch distribution of yearly spendable income. Both income measures are available from 2003 onwards. Household

5. The micro datasets including household income information are called 'Integraal Huishoudens Inkomen (IHI)' for 2003 to 2010 and 'Inkomen van huishoudens (INHATAB)' from 2011 onwards. The micro datasets on the personal level are 'Integraal Persoonlijk Inkomen (IPI)' for 2003 to 2010 and 'Inkomen van personen (INPATAB)' from 2011 onwards.

wealth is based on the percentile of the household in the Dutch distribution of the household's total wealth, determined by assets minus debts, and is available from 2006 onwards.

3.5 Background information

Basic background information on the children and their parents was added including information such as sex, birth date, country of origin, and migration background. It includes all people who were in the municipal personal records database from 1 October 1994 onwards.⁶

4. ITS dataset: combining survey and register data

Here we describe our main dataset, which combines survey and register data, and discuss its representativeness. The ITS dataset combines all aforementioned data from the education cohorts and the registers. Table 1 provides an overview of the merging process, and Figure 1 provides a graphical overview of the raw datasets used in creating the ITS dataset.

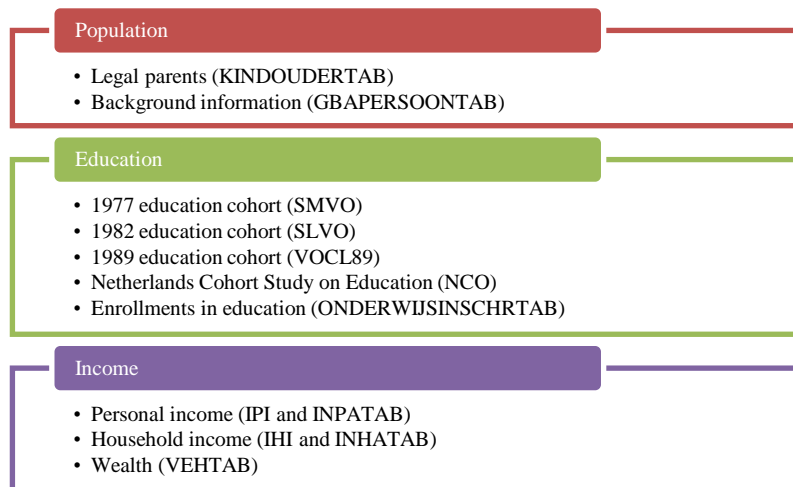
Table 1: Overview of the sample selection process

	Pooled		1977 Cohort		1982 Cohort		1989 Cohort	
	Parents	Children	Parents	Children	Parents	Children	Parents	Children
Original file	73,617		37,280		16,813		19,524	
Observations with linkable ID	64,046		30,171		14,764		19,111	
Observations with test data	58,090		25,462		14,613		18,015	
Observations with children	44,605	96,035	19,633	42,616	11,315	24,330	13,657	29,089
Observations with children with test scores	25,483	41,409	13,248	22,241	7,649	12,872	4,586	6,361
Observations with non-missing household income information	25,211	41,134	13,129	22,080	7,574	12,762	4,508	6,292
Observations with non-missing personal income information	24,053	38,635	12,515	20,846	7,227	11,963	4,311	5,884
Observations with non-missing household wealth information	23,309	36,973	12,084	19,655	7,292	11,984	3,933	5,334
Observations with all non-missing financial information	21,927	34,455	11,398	18,428	6,880	11,140	3,649	4,887

Data sources: Administrative data; ITS survey dataset.

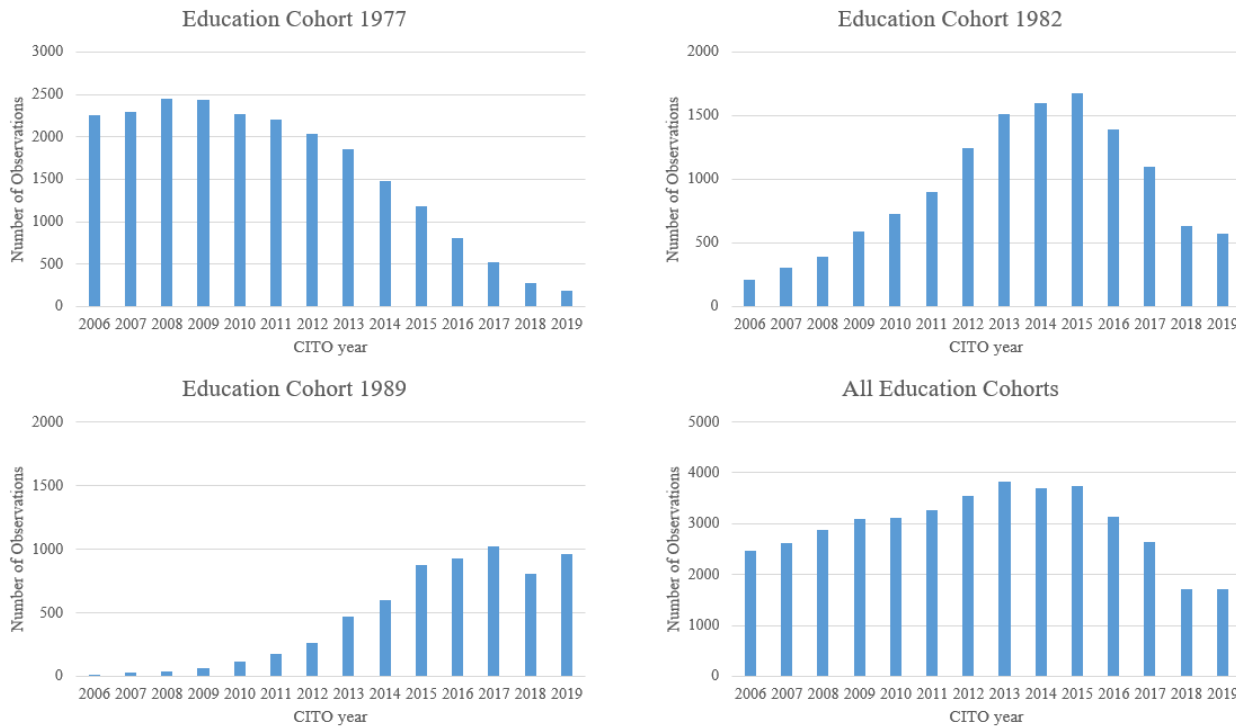
⁶ This micro data set is called 'GBAPERSONTAB'

Figure 1: Data overview



Note that only those parents are included in our parent sample whose children took the CITO-test between 2006 and 2019. This implies that for the 1977 cohort, we observe parents who are relatively old when they had children, while for the 1989 cohort we observe relatively young parents. Further, the distribution of children by education cohort in terms of the year they took the CITO-test is not uniform. As can be seen from Figure 2, which shows the distribution of children by test year, most children of cohort 1977 took their test in 2006-2012, while most children in cohort 1989 took their test in 2016-2019. The distribution of the total sample is more evenly spread over the test years.

Figure 2: Distribution of children by test year and by cohort



Data sources: Administrative data; ITS survey dataset.

4.1 Descriptive statistics

The selectivity of our sample with respect to age also has implications for parents' education and skills. Because more highly educated people tend to enter parenthood at a later age, the parents from the 1977 cohort whose children we observe in our data are positively selected in terms of their education and cognitive skills. The parents from the third cohort entered parenthood relatively young and therefore tend to have slightly lower educational attainment and cognitive skills, while the parents from the second cohort (around aged 12 in 1982) fall somewhere in between. This can be seen in tables 2, 3 and 4, where we show the descriptive statistics for the ITS sample at the child and parent level, and at the *child*parent* observation level, respectively.

Table 2: Descriptive Statistics – Child sample

Variables		Pooled	Cohort		
		(1)	1977 (2)	1982 (3)	1989 (4)
Child Characteristics					
Math skills	Mean	0.047	0.119	0.009	-0.126
	SD	0.97	0.95	0.97	1.00
Language skills	Mean	0.068	0.149	0.019	-0.133
	SD	0.96	0.92	0.96	1.00
Gender	Female	0.50	0.50	0.51	0.51
Migration background	Yes	0.08	0.07	0.07	0.12
CITO year	2006	0.06	0.10	0.02	0.002
	2007	0.06	0.10	0.02	0.005
	2008	0.07	0.11	0.03	0.01
	2009	0.07	0.11	0.05	0.01
	2010	0.08	0.10	0.06	0.02
	2011	0.08	0.10	0.07	0.03
	2012	0.09	0.09	0.10	0.04
	2013	0.09	0.08	0.12	0.07
	2014	0.09	0.07	0.12	0.09
	2015	0.09	0.05	0.13	0.14
	2016	0.08	0.04	0.11	0.15
	2017	0.06	0.02	0.09	0.16
2018	0.04	0.01	0.05	0.13	
2019	0.04	0.01	0.04	0.15	
Observations	Total number	41,409	22,241	12,872	6,361

Notes: Table reports means, SD, and shares for the variables indicated in column (1) for the pooled sample as well as the three education cohorts separately. The type of statistic reported is indicated in column (2). If neither Mean, SD, or Total number is specified, the reported statistic refers to the share with in the sample indicated in the top row. Children’s cognitive skills are standardized with mean zero and SD one in the full sample of children taking the test in their cohort based on administrative data. Children’s gender, CITO-year and migration background are taken from the administrative data. *Data sources:* Administrative data; pooled ITS survey dataset.

Both parents and their children from the 1977 cohort have on average higher math and language skills than the parents and children from the 1989 cohort. Parents from the 1977 cohort are more highly educated, and score higher on their IQ test. The parents and children from the 1982 cohort fall somewhere in the middle. In terms of their financial situation, the parents from the 1977 cohort are better off than those of the other cohorts. In part, this may be due to the returns to their higher cognitive skills as well as their older age. Further, it is noticeable that in the 1989 cohort, the share of parents and children with a migration background is higher than in the earlier cohorts. This is likely due to increased migration to the Netherlands over the years.

Table 3: Descriptive Statistics – Parent sample (one observation per parent)

Variables		Cohort			
		Pooled	1977	1982	1989
		(1)	(2)	(3)	(4)
Parent Characteristics					
Math skills	Mean	0.047	0.156	-0.015	-0.164
	SD	0.982	0.971	0.985	0.963
Language skills	Mean	0.060	0.146	0.004	-0.096
	SD	0.969	0.947	0.986	0.973
IQ scores	Mean	0.044	0.136	0.010	-0.051
	SD	0.978	0.957	0.990	0.978
Personal income percentile	Mean	62.44	64.70	61.86	56.86
	SD	28.61	28.95	28.28	27.32
Household income percentile	Mean	71.19	72.82	71.12	66.57
	SD	22.01	21.88	21.81	22.04
Household wealth percentile	Mean	56.82	61.95	55.41	43.90
	SD	25.78	24.97	25.22	24.38
Gender	Female	0.54	0.50	0.56	0.62
Education	Low	0.25	0.23	0.27	0.30
	Medium	0.44	0.48	0.40	0.40
	High	0.23	0.25	0.24	0.17
Migration background	Yes	0.09	0.07	0.08	0.14
Number of siblings	0 siblings	0.06	0.06	0.06	0.06
	1 sibling	0.39	0.35	0.43	0.41
	2 siblings	0.28	0.30	0.25	0.23
	>2 siblings	0.22	0.24	0.19	0.17
Age at time of birth child	Mean	31.53	33.44	30.80	27.22
Age at time of CITO child	Mean	43.84	45.78	43.09	39.48
Grandparent Characteristics					
Education	Primary education	0.19	0.15	0.26	0.20
	Lower secondary education	0.31	0.31	0.34	0.27
	Higher secondary education	0.28	0.33	0.18	0.34
	Tertiary education	0.16	0.18	0.14	0.14
Social status	Blue collar worker	0.29	0.29	0.30	0.29
	Employer – without staff	0.07	0.08	0.06	0.05
	Employer – with staff	0.05	0.06	0.05	0.04
	Lower white-collar worker	0.11	0.12	0.11	0.09
	Middle white-collar worker	0.18	0.20	0.16	0.17
	Professionals	0.12	0.12	0.10	0.12
	Other	0.14	0.12	0.14	0.20
Age at time of birth grandfather	Mean	30.45	31.37	29.70	29.04
Age at time of birth grandmother	Mean	27.85	28.69	27.26	26.41
Observations	Total number	25,483	13,248	7,649	4,586

Notes: Table reports means, SD, and shares for the variables indicated in column (1) for the pooled sample as well as the three education cohorts separately. The type of statistic reported is indicated in column (2). If neither Mean, SD, nor Total number is specified, the reported statistic refers to the share with in the sample indicated in the top row. Parent cognitive skills standardized with mean zero and SD one in the full sample of participants from each education cohort. Household income is based on the percentile of the household in the Dutch distribution in terms of yearly spendable income. Parent personal income is based on the percentile of the parent in the Dutch income distribution (sources include: labor income, owned companies, unemployment benefits and social security). Household wealth is based on the percentile of the household in the Dutch distribution in terms of the household's total wealth, determined by assets minus debts. Income and wealth data are taken from the administrative data in the child's test-taking year. Parent education is measured as the highest educational degree obtained by the parent observed in the survey data. In parent education, "low" denotes maximum lower secondary education (ISCED 1 or 2); "medium" denotes higher secondary or upper secondary vocational education (ISCED 3 or 4); "high" denotes tertiary education, consisting of higher vocational education and university (ISCED 5 and above). Grandparent education is measured based on the highest level of education of both grandparents. Social status is based on the occupation type of the main breadwinner in the parent household at the time of the parents' skill assessment. Apart from income and wealth, which are taken from administrative data, all (grand-)parent characteristics stem from the survey datasets. (Grand-)parent characteristics are reported at the parent level. *Data sources:* Administrative data; pooled ITS survey dataset.

Table 4: Descriptive Statistics – Parent sample (one observation per child*parent combination)

Variables		Cohort			
		Pooled	1977	1982	1989
		(1)	(2)	(3)	(4)
Parent Characteristics					
Math skills	Mean	0.100	0.220	0.027	-0.173
	SD	0.98	0.97	0.99	0.97
Language skills	Mean	0.105	0.202	0.036	-0.092
	SD	0.96	0.93	0.99	0.98
IQ scores	Mean	0.070	0.176	0.029	-0.057
	SD	0.97	0.94	0.98	0.98
Personal income percentile	Mean	63.29	66.36	61.67	55.72
	SD	28.84	28.77	28.65	27.79
Household income percentile	Mean	72.50	74.38	72.18	66.54
	SD	21.84	21.54	21.64	22.18
Household wealth percentile	Mean	58.08	63.29	56.05	43.42
	SD	25.86	24.82	25.33	24.51
Gender	Female	0.53	0.48	0.57	0.63
Education	Low	0.24	0.21	0.25	0.30
	Medium	0.44	0.48	0.41	0.40
	High	0.25	0.28	0.25	0.17
Migration background	Yes	0.08	0.07	0.08	0.15
Number of siblings	0 siblings	0.06	0.06	0.05	0.05
	1 sibling	0.37	0.34	0.41	0.40
	2 siblings	0.28	0.30	0.26	0.23
	>2 siblings	0.23	0.25	0.21	0.19
Grandparent Characteristics					
Education	Primary education	0.19	0.14	0.26	0.20
	Lower secondary education	0.31	0.30	0.34	0.27
	Higher secondary education	0.29	0.33	0.19	0.34
	Tertiary education	0.17	0.19	0.14	0.14
Social status	Blue collar worker	0.28	0.28	0.29	0.28
	Employer – without staff	0.08	0.09	0.07	0.05
	Employer – with staff	0.05	0.06	0.05	0.04
	Lower white-collar worker	0.11	0.12	0.11	0.09
	Middle white-collar worker	0.19	0.21	0.16	0.17
	Professionals	0.12	0.13	0.11	0.12
	Other	0.13	0.12	0.13	0.21
Age at time of birth grandfather	Mean	30.57	31.47	29.76	29.06
Age at time of birth grandmother	Mean	27.99	28.81	27.36	26.42
Observations	Total number	41,474	22,241	12,872	6,361

Notes: Table reports means, SD, and shares for the variables indicated in column (1) for the pooled sample as well as the three education cohorts separately. The type of statistic reported is indicated in column (2). If neither Mean, SD, or Total number is specified, the reported statistic refers to the share with in the sample indicated in the top row. Parent cognitive skills standardized with mean zero and SD one in the full sample of participants from each education cohort. Household income is based on the percentile of the household in the Dutch distribution in terms of yearly spendable income. Parent personal income is based on the percentile of the parent in the Dutch income distribution (sources include: labor income, owned companies, unemployment benefits and social security). Household wealth is based on the percentile of the household in the Dutch distribution in terms of the household's total wealth, determined by assets minus debts. Income and wealth data are taken from the administrative data in the child's test-taking year. Parent education is measured as the highest educational degree obtained by the parent observed in the survey data. In parent education, "low" denotes maximum lower secondary education (ISCED 1 or 2); "medium" denotes higher secondary or upper secondary vocational education (ISCED 3 or 4); "high" denotes tertiary education, consisting of higher vocational education and university (ISCED 5 and above). Grandparent education is measured based on the highest level of education of both grandparents. Social status is based on the occupation type of the main breadwinner in the parent household at the time of the parents' skill assessment. Apart from income and wealth, which are taken from administrative data, all (grand-)parent characteristics stem from the survey datasets. (Grand-)parent characteristics are reported at the child level. *Data sources:* Administrative data; pooled ITS survey dataset.

4.2 Representativeness

As the three education cohort studies were conducted to be representative of the Dutch student population at the time of the surveys, it is worth showing how much the subsample of parents observed in the ITS-dataset differs from the rest of the surveyed individuals. Therefore, Table 5 compares the characteristics of the parent subsample of the education cohorts included in the ITS dataset to the characteristics of the other respondents of the education cohorts.

Table 5: Representativeness of the parent sample for the education cohort studies

Variables		Pooled		Cohort 1977		Cohort 1982		Cohort 1989	
		Parents	Others	Parents	Others	Parents	Others	Parents	Others
Parent characteristics									
Language skills	Mean	0.060	-0.037	0.146	-0.109	0.004	-0.003	-0.096	0.032
	SD	0.969	1.017	0.947	1.024	0.986	1.011	0.973	1.007
Math skills	Mean	0.047	-0.030	0.156	-0.116	-0.015	0.013	-0.164	0.054
	SD	0.982	1.010	0.971	1.006	0.985	1.012	0.963	1.006
IQ scores	Mean	0.044	-0.024	0.136	-0.081	0.010	-0.009	-0.051	0.016
	SD	0.978	1.011	0.957	1.016	0.990	1.008	0.978	1.006
Gender	Female	0.54	0.49	0.50	0.53	0.56	0.45	0.62	0.44
Education	Low	0.25	0.32	0.23	0.40	0.27	0.25	0.30	0.25
	Medium	0.44	0.39	0.48	0.43	0.40	0.36	0.40	0.35
	High	0.23	0.21	0.25	0.16	0.24	0.25	0.17	0.27
Migration background	Yes	0.09	0.09	0.07	0.07	0.08	0.07	0.14	0.13
Number of siblings	0 siblings	0.06	0.07	0.06	0.08	0.06	0.07	0.06	0.06
	1 sibling	0.39	0.38	0.35	0.32	0.43	0.45	0.41	0.43
	2 siblings	0.28	0.25	0.30	0.27	0.25	0.23	0.23	0.23
	3 or more	0.22	0.18	0.24	0.22	0.19	0.18	0.17	0.11
Grandparent characteristics									
Grandparental education	Primary education	0.19	0.20	0.15	0.21	0.26	0.26	0.20	0.15
	Lower secondary	0.31	0.28	0.31	0.29	0.34	0.32	0.27	0.25
	Higher secondary	0.28	0.27	0.33	0.27	0.18	0.18	0.34	0.34
	Tertiary education	0.16	0.16	0.18	0.13	0.14	0.16	0.14	0.21
Social status	Blue collar	0.29	0.28	0.29	0.29	0.30	0.28	0.29	0.26
	Emp. w/o staff	0.07	0.06	0.08	0.06	0.06	0.05	0.05	0.04
	Emp. with staff	0.05	0.04	0.06	0.05	0.05	0.04	0.04	0.04
	Lower W. Collar	0.11	0.10	0.12	0.11	0.11	0.11	0.09	0.09
	Middle W. Collar	0.18	0.17	0.20	0.16	0.16	0.17	0.17	0.19
	Professionals	0.12	0.11	0.12	0.09	0.10	0.11	0.12	0.15
	Other	0.14	0.20	0.12	0.24	0.14	0.14	0.20	0.18
Age at t.o.b. grandfath.	Mean	30.45	30.62	31.37	31.50	29.70	30.17	29.04	29.80
Age at t.o.b. grandmoth.	Mean	27.85	27.91	28.69	28.68	27.26	27.53	26.41	27.20
Observations		25,483	48,134	13,248	24,032	7,649	9,164	4,586	14,938

Notes: Table reports means, standard deviations, and shares for the variables indicated in the first column for the pooled sample as well as the three education cohorts separately. The type of statistic reported is indicated in the second column. If neither Mean, SD, or Total number is specified, the reported statistic refers to the share with in the sample indicated in the bottom row. For each cohort, the left column summarizes the characteristics of the parents in our main estimation sample, the right column describes the same characteristics for the other participants in the education surveys. Parental cognitive skills are standardized with mean zero and standard deviation one in the full sample of participants from each education cohort. Grandparental education is measured by four categories of the highest level of education of both grandparents. Social status is measured by seven categories of occupational status of the main breadwinner in the parent household. Grandparent education and social status refer to time when parents took the skill test. All (grand)parental characteristics are taken from the survey datasets. (Grand)parental characteristics are reported at the parent level. *Data sources:* Pooled ITS survey dataset.

As expected, a positively (negatively) selected subsample of the 1977 (1989) respondents in terms of math and language skills, IQ, and highest obtained level of education, is included in the ITS-dataset, with the 1982 cohort being highly representative of the overall student population in the cohort data. The same pattern is visible for the education level of the parents of the surveyed individuals (the grandparental generation in the ITS-dataset) as well as their social status (based on the social status of the grandparent). In the 1982 and (especially) the 1989 cohorts, female respondents are more likely to be included in the ITS-dataset, as women tend to have children at an earlier age than men. For the 1977 cohort, the reverse is observed. It is further worth noting that because of the opposite direction of selectivity in the 1977 and 1989 cohorts, the pooled ITS dataset is fairly representative of the three pooled cohort studies. However, since the 1977 education cohort was larger than the other two, a somewhat positively selected sample of parents remains.

Table 6: Representativeness of the child sample for the total CITO sample

Variables		Pooled	1977	1982	1989	Children not in sample
Child characteristics						
Math skills	Mean	0.048	0.120	0.009	-0.123	0.006
	SD	0.97	0.95	0.97	0.99	0.998
Language skills	Mean	0.066	0.149	0.019	-0.132	0.006
	SD	0.95	0.92	0.96	1.00	0.998
Gender	Female	0.50	0.50	0.51	0.51	0.50
Migration background	Yes	0.08	0.07	0.07	0.12	0.22
CITO-year	2006	0.06	0.10	0.02	0.002	0.08
	2007	0.06	0.10	0.02	0.005	0.07
	2008	0.07	0.11	0.03	0.01	0.07
	2009	0.07	0.11	0.05	0.01	0.07
	2010	0.08	0.10	0.06	0.02	0.07
	2011	0.08	0.10	0.07	0.03	0.07
	2012	0.09	0.09	0.10	0.04	0.08
	2013	0.09	0.08	0.12	0.07	0.08
	2014	0.09	0.07	0.12	0.09	0.08
	2015	0.09	0.05	0.13	0.14	0.08
	2016	0.08	0.04	0.11	0.15	0.07
	2017	0.06	0.02	0.09	0.16	0.06
2018	0.04	0.01	0.05	0.13	0.06	
2019	0.04	0.01	0.04	0.15	0.05	
Observations	Total number	41,409	22,241	12,872	6,361	1,848,605

Notes: Table reports means, SD, and shares for the variables indicated in column (1) for the pooled sample as well as the three education cohorts separately. The type of statistic reported is indicated in column (2). If neither Mean, SD, or Total number is specified, the reported statistic refers to the share with in the sample indicated in the top row. Children's cognitive skills are standardized with mean zero and SD one in the full sample of children taking the test in their cohort based on administrative data. Children's gender, CITO-year and migration background are taken from the administrative data. *Data sources:* Administrative data; pooled ITS survey dataset.

Table 6 shows the representativeness of the children in the ITS survey dataset for the student population taking the CITO-test over the period 2005/2006 to 2018/2019. Overall, the children observed in the ITS dataset score slightly above average on the math and language subscales of the CITO-test. This is driven by higher scores of children from the (larger) 1977 parental cohort, with children from the (smaller) 1989 parental cohort scoring well below average. The most striking difference between the children of the ITS dataset and the general CITO-test taking student population is the relatively low share of children with a (first or second generation) migration background present in the ITS survey dataset. This difference is most likely due to the fact that we only include children in the ITS dataset whose parents were resident in the Netherlands around age 12 in 1977, 1982, and 1989. All students with parents that migrated to the Netherlands at a later age, are not in our dataset. Nevertheless, the increasing amount of students with a migration background over time can already be seen from the differences in the share of children with a migration background in the ITS dataset between the 1977 and 1982 cohorts (0.07) and the 1989 cohort (0.12).

5. References

- Blommers, A. J. (1983). *Het onderzoek schoolloopbaan en herkomst bij het voortgezet onderwijs (SLVO-1982)*. Voorburg: CBS.
- Bosker, R., & Velden, R. D. (1989). *Schooleffecten en rendementen*. In J. van Damme & J. Dronkers (Eds.), *Jongeren in school en beroep* (pp. 25-40). Amsterdam: Swets & Zeitlinger.
- CBS. (1982). *Schoolloopbaan en herkomst van leerlingen bij het voortgezet onderwijs 1977*. The Hague: CBS.
- CBS. (1991). *Schoolloopbaan en herkomst van leerlingen bij het voortgezet onderwijs: vergelijking tussen de cohorten 1977 en 1982*. Voorburg/Heerlen: CBS.
- CvTE. (2015). *Verantwoording Centrale Eindtoets po*. Utrecht: CvTE.
- Driessen, G., & van der Werf, G. (1992). *Het functioneren van het voortgezet onderwijs. Beschrijving steekproef en psychometrische kwaliteit instrumenten*. Groningen: RION, Instituut voor Onderwijsonderzoek, Rijksuniversiteit Groningen.
- Haelermans, C., Huijgen, T., Jacobs, M., Levels, M., van der Velden, R., van Vugt, L., & van Wetten, S. (2020). Using data to advance educational research, policy and practice: Design, content and research potential of the Netherlands cohort study on education. *European Sociological Review*, 36(4), 643-662.
- Hustinx, P. W. J., Kuyper, H., van der Werf, M. P. C., & Zijssling, D. (2005). *Beschrijving leerlingbestanden vocl '89*. Groningen: GION, Gronings Instituut voor onderzoek van onderwijs, opvoeding en ontwikkeling.
- Kuyper, H., & van der Werf, M. P. C. (2007). *De resultaten van VOCL '89, VOCL '93 en VOCL '99: Vergelijkende analyses van prestaties en rendement (9066909528)*. Groningen: Universiteit Groningen.
- Van Boxtel, H., Engelen, R., & de Wijs, A. (2011). *Wetenschappelijke verantwoording van de Eindtoets 2010*. Arnhem: CITO.