InGRID

Supporting expertise in inclusive growth

www.inclusivegrowth.eu

Deliverable D8.6

HARMONISING LONGITUDINAL DATA ON EDUCATIONAL CAREERS

Mapping and matching harmonisation possibilities

Flore Debruyne, Monica Wouters & Nele Havermans

2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 730998

Abstract

This report focuses on opportunities to harmonise longitudinal datasets regarding education and educational outcomes that exist across Europe. More precisely, we mapped the longitudinal datasets on education via a literature review and collected data through (i) an expert questionnaire and (ii) technical reports and websites. Next, we explored the possibilities of matching these longitudinal datasets and provided methodologies to harmonise the datasets. The appropriate methodology to harmonise depends on factors (such as sample size) that need to be weighed by the researcher. This report provided the foundations and presented the tools to harmonise longitudinal datasets on educational careers across Europe.

This report constitutes Deliverable 8.6, for Work Package 8 of the InGRID-2 project.

July, 2021

General contact: inclusive.growth@kuleuven.be
p.a. InGRID
HIVA - Research Institute for Work and Society
Parkstraat 47 box 5300, 3000 LEUVEN, Belgium

For more information flore.debruyne@kuleuven.be

Please refer to this publication as follows:
Debruyne, F., Wouters, M. & Havermans, N. (2021), *Harmonised longitudinal data on educational careers. Mapping and matching harmonisation possibilities (Deliverable D8.6)*, Leuven, InGRID-2 project 730998 – H2020

This publication is part of the InGRID-2 project, this project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 730998.

InGRID
Supporting expertise in inclusive growth

Contents

List of acronyms and abbreviations

Admin	Panel of Linked Administrative Data
ASAtS	Assessment of Students' Attitudes towards Studying
DAR	Danish administrative register
DNT	Danish National Test
ECHP	European Community Household Panel
ELFE	*Etude Longitudinale Française depuis l'Enfance*
EU	European Union
EU-SILC	European Union Statistics on Income and Living Conditions
FiD	*Familien in Deutschland*
HLCS	Hungarian Life Course Survey
ICCS	International Civic and Citizenship Education Study
ICT	Information and communication technology
IEA	International Association for the Evaluation of Educational Achievement
LFS	Labour Force Survey
LISA	*Lesen in der Sekundarstufe*
LiSO	*Loopbanen in het Secundair Onderwijs*
MCS	Millennium Cohort Study
NABC	National Assessment of Basic Competencies
NEPS	National Education Panel Study
OECD	Organisation for Economic Co-operation and Development
PIAAC	Programme for the International Assessment of Adult Competencies
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
SES	Socioeconomic status
SiBO	*Schoolloopbanen in het basisonderwijs*
SOEP	Socio-Economic Panel
TIMSS	Trends in International Mathematics and Science Study
TNA	Transnational access
TOSCA	Transformation of the Secondary School System and Academic Careers
TREE	Transitions from Education of Employment
IDA	Integrative Data Analysis
RE	Random-effects
FE	Fixed-effects

List of figures

List of tables

1. Introduction

Data on educational careers are essential for the study of equity and efficiency of education systems. Yet, European comparative research in education currently depends largely on cross-sectional sources of the International Association for the Evaluation of Educational Achievement (IAE) or the Organisation for Economic Co-operation and Development (OECD), such as the Programme for International Student Assessment (PISA), the Programme for the International Assessment of Adult Competencies (PIAAC), Progress in International Reading Literacy Study (PIRLS), or International Civic Citizenship Education Study (ICCS). In addition, the longitudinal datasets that are available for European countries, such as the European Community Household Panel (ECHP), Labour Force Survey (LFS) and European Union Statistics on Income and Living Conditions (EU-SILC; for children) include very little information on education careers. However, in several European countries national longitudinal datasets are available that could be exploited more intensively for comparative research. Consequently, Task 5 in Work Package 8 of the InGRID-2 project targets this research gap and aims to explore opportunities for harmonisation of longitudinal datasets on education and educational outcomes across Europe.[1]

Two main objectives can be distinguished within Task 5. Firstly, we aim to identify and document longitudinal datasets on education. Therefore, we create an inventory of European micro-level datasets on educational careers starting from the entry into early childhood education and care through primary, secondary, and tertiary education as well as training in later life. Within the second objective, we explore the possibilities of sharing or merging the longitudinal datasets on education. We aim to facilitate comparative research (for a selection of countries) by exploring new possibilities for transnational access and providing methodologies to harmonise available datasets in Europe.

This report starts with a description of the methodology that was used to create an inventory of the datasets. The inventory is reported in Chapter 3. Within this chapter, we present the data collection, data content, transnational access (TNA), and refer to sources where more information on the specific dataset can be found. Chapter 4 elaborates more on data harmonisation. The five steps of harmonisation are introduced and the different methodologies for harmonisation are outlined. Lastly, the research is concluded in the discussion. The limitations of this research are described and recommendations for future harmonisation initiatives are offered.

1 'Data harmonisation is the process of making data from different sources more similar. This could be data collected at different sweeps or time periods within the same study or it could be data collected by separate studies' (CLOSER, 2021a, para. 1).

2. Methodology

In this chapter we describe how we created an inventory of the existing longitudinal datasets regarding educational careers in Europe based on a standardised questionnaire. We first lay out how we selected the datasets that were suitable for our inventory. Next, we describe the two parts of our data collection. Finally, we present an overview of the datasets in our inventory.

2.1 Data selection

The exploration of longitudinal datasets and their corresponding contact persons, started with a literature review. Via Limo Search and Google Scholar we searched for papers on educational careers that referred to longitudinal datasets with a focus on comparative studies. Menard (2002) defined longitudinal research as 'research in which (i) data are collected for each item or variable for two or more distinct time periods, (ii) the subjects or cases analysed are the same or at least comparable from one period to the next; and (iii) the analysis involves some comparison of data between or among periods' (p. 2). Furthermore, we only selected longitudinal datasets on educational careers from European countries as the focus of InGRID-2 lies on integrating European research infrastructures. Datasets outside the EU-28 and the Schengen Area were not included.[2]

2.2 Data collection

We collected the information for the meta-data sheet concerning longitudinal datasets on educational careers in two ways: (i) through experts filling out a questionnaire and (ii) through filling out the standardised questionnaire ourselves based on the datasets' technical reports and websites.

Firstly, we asked the author(s) of the papers selected in our literature review (N=32) by email to fill out a questionnaire on the dataset(s) they used or constructed for their research. The questionnaire was also distributed during an expert workshop that took place in Berlin (Germany) from the 27th to 29th of November 2019. During this three-day workshop 'Comparative analyses of longitudinal educational outcomes', 21 experts on educational research and datasets presented their work with, or on longitudinal educational datasets.[3] The questionnaire examined the basics of the dataset and consisted of two main parts: (i) questions about the design of the dataset (time and location of data collection, sampling) and (ii) questions about the content of the dataset (questionnaire, technical report). In advance, the questionnaire was piloted with two respondents via email. A full overview of the questionnaire can be found in Appendix 1. Completing the questionnaire took approximately 30 to 45 minutes. The questionnaire was filled out per dataset. Additionally, respondents had the possibility to submit the form more than once if they could provide information about more than one such dataset.

Secondly, the research team collected information on those datasets for which no questionnaire was filled out. We searched for information on the dataset's website or in the technical reports and

2 Because this research began before Brexit we considered the EU-28 instead of the current EU-27 and, thus, included the United Kingdom in our selection of countries.

3 The programme (including a list of the attending experts) is available on the project website: http://www.inclusivegrowth.eu/expert-workshops/call-23-expert-workshop-diw

papers. The datasets for which we neither received a questionnaire nor found sufficient information on are not part of the meta-data in this report.

2.3 The inventory

As seen in Table 1, the data collection resulted in a meta-data sheet based on the standardised questionnaire. An inventory of 19 datasets was designed, answering the first goal of Task 8.5. For 11 datasets, information was acquired through the questionnaire distributed via mail or in the expert workshop. Information on the remaining 8 datasets was collected from the websites of the datasets or from technical reports and papers.

Table 1. An inventory of longitudinal datasets on educational careers

Source of information	Dataset	Acronym	Country
Questionnaire	The Panel of Linked Administrative Data of CERS Databank	Admin1, Admin2, Admin3	Hungary
	Assessment of Student's Attitudes towards Studying	ASAtS	Switzerland
	Danish Administrative Registers	DAR	Denmark
	Danish National Test	DNT	Denmark
	Hungarian Life Course Survey	HLCS	Hungary
	Lesen in der Sekundarstufe	LISA	Germany
	National Assessment of Basic Competencies	NABC	Hungary
	National Educational Panel Study	NEPS	Germany
	Schoolloopbanen in het basisonderwijs	SiBO	Belgium (Flemish region)
	Transformation of the secondary school system and academic careers	TOSCA	Germany
	Longitudinal dataset from the Netherlands		The Netherlands
Online documentation	European Community Household Panel	ECHP	Belgium, Denmark, Germany, Ireland, Greece, Spain, France, Italy, Luxembourg, Netherlands, Austria, Portugal, Sweden, and the United Kingdom
	Etude Longitudinal Français depuis L'Enfance	ELFE	France
	Familien in Deutschland	FiD	Germany
	Loopbanen in het Secundair Onderwijs	LiSO	Belgium (Flemish region)
	Millennium Cohort Study	MCS	The United Kingdom (England, Scotland, Wales, and Northern Ireland)
	Socio-Economic Panel	SOEP	Germany
	Trends in International Mathematics and Science Study	TIMSS	59 countries around the world[4]
	Transitions from Education to Employment	TREE	Switzerland

* Datasets that have no official name were given a standard name after the country in which they were collected. The cell of the datasets for which no acronym exists was simply left empty.

4 An overview of the participating countries can be found at https://www.iea.nl/studies/iea/timss

3. Overview of the datasets in the inventory

In this chapter, we give an overview of the datasets in the inventory. As one of the aims of this report is to provide a summary of the existing datasets and their content to enable researchers to select datasets for comparative research, the datasets were organised by country and the most relevant information is presented. For each dataset we describe the data collection method, the content, publications, the data access, and we refer to sources where more information on the dataset can be found. This information was collected from the expert questionnaire (when possible) and supplemented with online information. Additionally, sources where more information can be found were mentioned when possible.

A summary of the datasets in the inventory can be found in Table 2. In this table, we include the content of the longitudinal datasets (respondents, content of questionnaire) and how the datasets were designed. The content of the datasets is discussed in more detail in Chapter 5.

For the discussion of the design of the datasets, we specify to which type of longitudinal data the dataset belongs. We distinguish between six types of longitudinal data: (i) panel surveys, (ii) cohort study, (iii) birth cohort, (iv) multicohort, (v) administrative data, and (vi) a quasi-longitudinal method. In *panel surveys* a sample of respondents is contacted and surveyed on multiple occasions (Gayle & Lambert, 2018). The respondents are not necessarily individuals. For instance, a household panel survey gathers information about the household as a whole in order to view individuals in the context of their household (CLOSER, 2021b). A *cohort* is a special type of panel as the individuals in the panel share a common characteristic or experience within a given period of time (Gayle & Lambert, 2018; Menard, 2002). For example, individuals in a *birth cohort* are born around the same time (Gayle & Lambert, 2018; Menard, 2002). Furthermore, a longitudinal panel design that includes multiple cohorts is a *multicohort study* (Menard, 2002). Moreover, while *administrative data* is not necessarily collected by government departments and agencies for research purposes, it contains repeated observations on the same units (CLOSER, 2021b; Gayle & Lambert, 2018). Thus, it is a great source for longitudinal research. Some studies, however, do not fall completely under the definition of longitudinal research (Menard, 2002). When it is not possible or preferable to collect data from the same respondents for the entire scope of the research, a *quasi-longitudinal method* can be used. Within this method, data is collected from groups of respondents covering a shorter period of time than the full research period (Oud, 2001). This method is applied within the Trends in International Mathematics and Science Study (TIMSS) as it repeatedly follows fourth and eighth grade students (International Association for the Evaluation of Educational Achievement, 2021).[5] Table 2 shows which method was used for each dataset.

3.1 Belgium

3.1.1 Schoolloopbanen in het basisonderwijs (SiBO)

Schoolloopbanen in het basisonderwijs (SiBO; School Trajectories in Primary Education) was a large-scale cohort study in which students were followed during elementary school. The schools were located in

5 The relationship between grade and age can be found in Table a1 at Appendix 2.

the Flemish Community of Belgium, which is the northern Dutch-speaking part of the country. The SiBO-project started in the school year 2002-2003 by following a cohort of approximately 6,000 children embedded in 196 schools who attended the last kindergarten class. During the 2010-2011 school year, the last round of data for SiBO were collected (Vandenberghe et al., 2012). Hence, the school careers of a cohort of pupils were followed until the end of primary education and even until the first year of secondary education in the case of pupils who did not have to repeat a grade. The data were collected via (reading, spelling, and mathematics) tests, written questionnaires, or in some cases limited observations (Gadeyne et al., 2006; Vandenberghe et al., 2012). In 2014, when the cohort members were circa 17 years old, they participated in a follow-up data collection, consisting of a mathematics test and a student questionnaire (Vanwynsberghe et al., 2017c).

SiBO was designed to investigate the school careers of children through primary education. The researchers collected data on student, class, and school level. Firstly, student level variables included socioeconomic background, cognitive development, sociocognitive variables, social development, dynamic and affective development, wellbeing, class climate and learning development, and environmental characteristics. Secondly, class level variables included class characteristics, teacher characteristics, teacher beliefs or opinions, teacher perceptions, and didactics. Lastly, the board of directors and school team completed a questionnaire to collect school level variables. This enabled researchers to use the SiBO data to uncover some of the essential characteristics of primary education through information about pupils' careers.

Publications using the SiBO-dataset are listed at different locations. Firstly, Vandenberghe et al. (2012) listed publications from 2005 to 2012 on the dataset. Publications from 2012 to 2016 can be found at https://steunpuntssl.be/Publicaties. Since 2016, five other journal articles have been published. Anumendem et al. (2017) investigated the growth in reading comprehension and mathematics achievement in primary school. Furthermore, the long term effect of primary schools on educational positions (Vanwynsberghe et al., 2017a), on mathematics achievements (Vanwynsberghe et al., 2017b), and on non-cognitive outcomes (Vanwynsberghe et al., 2017c) have been examined. Verhaeghe et al. (2018) studied whether group composition effects explain why socioeconomic and ethnic achievement gaps were not completely reduced or even expanded throughout primary school careers. Lastly, Verschueren et al. (2019) investigated the perspectives of teachers, peers, and students on the social acceptance of high-ability students.

Access to the SiBO dataset can be obtained by filling out a form on http://steunpuntsono.be/. It should be noted that the Flemish government retains the intellectual rights of the data. Hence, if anything is published using the data, the SiBO-project or the policy research centre (i.e., *Steunpunt Onderwijsonderzoek*) should be mentioned. Furthermore, researchers are obliged to inform the government on their results before making these available to the general public.

Further information on the dataset is available in Dutch and English. The *Steunpunt studie- en schoolloopbanen* (SSL) website is predominantly in Dutch but has a small English section where a detailed version of the multiannual programme and a summary of the Annual Report 2013 can be found.[6] Moreover, details on the dataset are described in English in the articles where SiBO-data were used (e.g., Anumendem et al., 2017; Gadeyne et al., 2006; Vanwynsberghe et al., 2017a, 2017b, 2017c, 2019; Verhaeghe et al., 2018; Verschueren et al., 2019).

3.1.2 Loopbanen in het Secundair Onderwijs (LiSO)

Loopbanen in het Secundair Onderwijs (LiSO; Trajectories in Secondary Education) is a cohort study following approximately 6,500 students throughout secondary education (Stevens et al., 2015). The data collection started in 2013 when students were in seventh grade, starting their secondary education in the Flemish Community (Stevens et al., 2015). In the first year, two waves of data collection took

6 https://steunpuntssl.be/

place: one at the beginning of the school year and one at the end (LiSO, 2017b). From then on, data collection took place annually until 2019, when students were in twelfth grade and leaving secondary education (LiSO, 2017b). Although the initial plan was to follow the cohort members in later life by using administrative data (LiSO, 2017b), the Flemish government decided to stop funding the project (LiSO, 2017a). The respondents were not only students who completed tests and filled in questionnaires but also their parents, the teachers, the school team, and the school principal were asked to participate by completing questionnaires (LiSO, 2017b).

The LiSO-project collected information on students' school trajectories in secondary education. The variables in the LiSO-project are not limited to cognitive outcomes (e.g., mathematics, reading comprehension, and French) but also include non-cognitive outcomes (e.g., well-being and mindset) and school career characteristics (e.g., choice of study and remaining in school; LiSO, 2017b; SONO, 2021).

At this moment, only a selected group of researchers have access to the anonymised data. It may be possible to release the data to other researchers (after signing an agreement) but no final decision has been made on this yet (LiSO, 2021).

More information in Dutch can be found on the LiSO-project website.[7] The website includes general information, technical reports,[8] and the (predominantly Dutch) publications using the LISO-data.[9]

3.2 Denmark

3.2.1 Danish administrative register (DAR)

The Danish administrative registers are administrative data on the Danish population collected by Statistics Denmark. While the collection of data based on administrative registers began in the 1970s, the production of statistics based on those registers was 'not sufficiently comprehensive and well-established until 1981' (Statistics Denmark, 2014, p. 1). Moreover, the start of the data collection on educational registers depends on the considered register. The educational registers cover the entire Danish public school population (Jensen, 2020).

The Danish administrative register does not only contain data on all levels of education but also covers many other fields of research, such as health and labour market participation. The educational register data contain 'complete detailed educational histories, including detailed codes for the type of education followed (level, subject, and educational institution) and the dates for entering and exiting the education, along with an indication of whether the individual completed the education successfully, dropped out or is still enrolled as a student' (Joensen & Nielsen, 2009, p. 184). Furthermore, background information on pupils, siblings, parents, and grandparents can be found in the non-educational registers (Andersen et al., 2018).

As the administrative data are comprehensive, many articles based on these data were published. Research includes (but is not limited to) the link between educational achievement and labour market outcomes (Joensen & Nielsen, 2009), the effect of educational frameworks on academic achievement and labour market outcomes (Jensen, 2020), and how reading and writing support each other (Andersen et al., 2018).

The administrative data are available for researchers but there are conditions to the data access for non-Danish researchers. How access to the data can be granted is described on the Statistics Denmark website.[10]

7 https://lisoproject.be/
8 https://lisoproject.be/onderzoek
9 https://lisoproject.be/resultaten2
10 https://www.dst.dk/en/TilSalg/Forskningsservice

3.2.2 Danish National Test (DNT)

The Danish National Test (DNT) is a multicohort study that is collected every year in order to document socioeconomic inequality in students' national test scores (Vad Andersen, 2021). The cohort members are followed from grades 6 to 8. Additionally, their ninth grade exam results can be linked to their test scores (Beuchert & Nandrup, 2017). There are twelve tests for different grade levels and subjects. Ten of those are mandatory and cover the subjects Danish/reading, mathematics, English, physics/chemistry, geography, and biology (Beuchert & Nandrup, 2017). The other two (one in grade 5 and the other in grade 7) are voluntary for the school and are on the subject of Danish as a second language (Beuchert & Nandrup, 2017). An overview of the 12 tests per grade and subject can be found in Beuchert and Nandrup (2017) and Wandall (2017). Every test can be used three times per student. Furthermore, one of these tests is mandatory to students in the public school system. This enables systemic data collection that is comparable across years (Vad Andersen, 2021).

The data are primarily used for pedagogical and administrative purposes. Access to the data can be granted by the Ministry of Children and Education for research purposes on microlevel but only from the facilities of the Danish statistical bureau.

While the Ministry of Children and Education does not provide technical reports, Beuchert and Nandrup (2017) described the technical details of the DNT. Synopses of the tests were provided by the website of the Ministry of Children and Education,[11] and Aarhus University.[12]

3.3 France

3.3.1 Etude Longitudinale Française depuis l'Enfance (ELFE)

Etude Longitudinale Française depuis l'Enfance (ELFE; French Longitudinal Study since Childhood) is a birth cohort study conducted by the French National Institute for Demographic Studies (INED) and the National Institute for Health and Medical Research (INSERM). Because no expert questionnaire was available on ELFE, information had to be drawn from the ELFE website managed by INED (2021). Midwives in 344 randomly selected maternity units across metropolitan areas helped targeting cohort members who were born in 2011 during four selection periods representing the four seasons. More than 18,000 children were selected, which amounts for 1 in 50 children who were born in 2011. The study aims to follow the cohort members from their birth until they reach the age of 20 in 2031. ELFE gathers information from the cohort members themselves, their parents, doctors, and teachers. The different stages of the data collection can be found online.[13]

The ELFE website managed by INED (2021) provides detailed information on the content of the study. The main research topics are social sciences (e.g., education and family transformations), health (e.g., pregnancy and neurodevelopmental disorders), and environment (e.g., air pollution and pesticides). Furthermore, the study aims to identify the factors that influence children's socialisation, education, and academic success. It contributes to the data collection on factors such as family and relationships, academic success, values transmitted at school, children's perception of what goes on at school, leisure activities, father's involvement, child development and behaviour, and well-being.

Based on the ELFE-data, research has been published on various topics. For example, Fischer and Thierry (2021) investigated how SES determined academic achievement while Berger et al. (2021) examined the link between childcare arrangements and language development. Furthermore, there are publications in academic journals on the methodology of the ELFE study (e.g., Vandentorren et

11 https://eng.uvm.dk/primary-and-lower-secondary-education/the-folkeskole/evaluation-tests-student-and-plans

12 https://childresearch.au.dk/en/signature-project-read/struktur-og-rammer/danish-national-tests/

13 https://www.elfe-france.fr/en/the-elfe-study/how-does-it-work/key-stages/

al., 2009). The ELFE website has a list of publications in academic journals ordered by topic and year of publication.[14]

The ELFE data is available on request. More information on how to access the data and which variables are available can be found at the ELFE website.[15]

3.4 Germany

3.4.1 Familien in Deutschland (FiD)

Familien in Deutschland (FiD; Families in Germany) is a longitudinal panel study. Schröder et al. (2013) gave an overview of the sampling and questionnaire contents. The FiD data were collected annually between 2010 and 2013. This resulted in a total of three waves. The FiD study was designed to target four different types of families: (i) families with children born between 2007 and 2010, (ii) single-parent families, (iii) low-income families, and (iv) families with more than two children. The families with children born between 2007 and 2010 were included in the cohort sample. The three other targeted respondents were identified in the screening samples. Given the large scope of respondents, there were different questionnaires for the respondents to fill in. Apart from the household questionnaires, each adult person completed a personal questionnaire. Furthermore, parents filled out a parent questionnaire for their children between 0 and 10. Note that not each adult had to complete a parent questionnaire as children who were older filled out a youth questionnaire themselves (people turning 17 during the survey year) or an adult questionnaire (respondents turning 18 during the survey year).

The survey paper by Fräßdorf et al. (2016) gave an overview on the data documentation. The FiD study focused on children and partnership. It included questions on marriages and partnerships that lasted longer than six months. Furthermore, the dataset contains general information on education, past and current labour market experiences, earnings and incomes, housing characteristics, health, and life satisfaction. More precisely, data on education concerned the schooling degrees, vocational and university degrees, time spent in education, and required formal education and on-the-job training. In the scope of this report, it is important to note that the dataset does not cover primary schooling.

According to DIW Berlin (2021), the FiD dataset and its publications are part of the Research Data Centre of the Socio-Economic Panel (SOEP). The FiD research project was carried out with a specific sample, namely four types of families, that can be integrated into SOEP because the contents are similar to each other. Furthermore, the FiD dataset is also available as an independent dataset. Publications that used the FiD dataset, and more broadly the SOEP datasets, are available at http://www.diw.de/soepsurveypapers and at http://www.diw.de/soeppapers. The data itself can be obtained from the SOEP Research Data Centre.[16]

3.4.2 Lesen in der Sekundarstufe (LISA)

The *Lesen in der Sekundarstufe* (LISA) contains information on individual and contextual determinants of reading comprehension and reading motivation in secondary education. The dataset consists of 1,508 students in 60 schools who completed a student questionnaire and took a reading achievement test. Additionally, the parents also filled in a questionnaire (Muntoni & Retelsdorf, 2019). The questionnaires and the test were administered approximately every 18 months (Retelsdorf et al., 2014), resulting in six rounds of data collection. The first wave took place in 2005 at the beginning of grade 5 when students were approximately 11 years old. The next three waves occurred at the end of grade 6,

14 https://www.elfe-france.fr/en/the-research/publications/academic-journals/
15 https://www.elfe-france.fr/en/the-research/access-to-data-and-questionnaires/
16 https://www.diw.de/en/diw_01.c.601584.en/data_access.html

at the beginning of grade 8, and at the end of grade 9 (Retelsdorf et al., 2014). Furthermore, the fifth round was completed after entering upper secondary level in grade 11 (Kampa et al., 2021). Lastly, the sixth LISA survey (LISA 6 or Educational Outcomes of Students from Vocational and Academic Upper Secondary Schools) took place in 2013 in grade 13 before reaching university entrance (Kampa et al., 2020).

Several articles based on the LISA-dataset were published. Retelsdorf et al. (2012) compared the development of reading comprehension of students at academic and non-academic track schools.[17] Furthermore, Retelsdorf et al. (2014) investigated the reciprocal effects between reading self-concept and reading achievement. Volodina et al. (2015) analysed the transition from lower to upper secondary school. Moreover, Muntoni and Retelsdorf (2019) investigated how parents' reading-related gender stereotypes affected their children's learning outcomes. Lastly, the effects of intelligence and motivation on academic achievement (Köller et al., 2019) and the relationship between personality traits and academic achievement (Meyer et al., 2019) were investigated using the LISA-dataset. A limited overview of publications where LISA-data were used can be found via the following links: https://www.iqb.hu-berlin.de/fdz/studies/LISA and https://www.iqb.hu-berlin.de/fdz/studies/LISA_6/?doi=10.5159/IQB_LISA_6_v1. Additionally, these websites also contain more information on the dataset, further documentation of the technical details, and a link to the application form for the scientific use of the LISA datasets.

3.4.3 National Education Panel Study (NEPS)

The National Education Panel Study (NEPS) is a multicohort longitudinal study, administered in yearly waves. NEPS (2021a) currently has six starting cohorts that were sampled from 2009 to 2012. With those six cohorts, the NEPS aims to follow individuals throughout their life. The target population differs for each cohort:

- starting cohort 1 (SC1): all children born in Germany from February to July 2012 and their families;
- starting cohort 2 (SC2): 4-year-olds in kindergarten;
- starting cohort 3 (SC3): students in grade 5;
- starting cohort 4 (SC4): students in grade 9;
- starting cohort 5 (SC5): students in higher education;
- starting cohort 6 (SC6): students in adult education.

The focus of the NEPS is mainly on educational careers and educational outcomes. To analyse the development of educational pathways of the cohorts, the NEPS is based on eight pillars: competence development, learning environments, educational decisions, migration background, returns to education, and personality and motivation (National Educational Panel Study, 2021b). A more precise description of the research objectives per cohort can be found at the NEPS website.[18]

There are many publications in which the NEPS data were used. For example, Hondralis and Kleinert (2021) used SC1 of the NEPS data to investigate whether the early development of children influenced their mothers' decision to return to the labour market after giving birth. Using SC2, Gil-Hernández (2021) studied a cohort of students from grades 1 to 5 to examine whether high-SES students substitute low cognitive skills in tests scores by higher noncognitive skills (conscientiousness) in the transition to academic secondary school. Furthermore, DeVries et al. (2021) worked with data from SC3. They focused on grades five, seven, and eight to assess whether self-concept and self-esteem mediated risk factors for lower academic achievement in mathematics and reading. To research the development of gender differences in ICT within adolescents aged 15 across a period of

17 'After elementary school, students in Germany are assigned to different types of school that either place a focus on students' gaining qualifications that would enable them to begin a vocational apprenticeship (non-academic track schools) or prepare them for university entrance (academic track schools)' (Retelsdorf et al., 2012, p. 649).

18 www.neps-data.de/Mainpage

three years, Gnambs (2021) drew on the data of SC4. The data of SC5 were used by Behr et al. (2021) to examine Bachelor students' motives for leaving higher education without obtaining a degree. Lastly, Granderath et al. (2021) investigated the effect of participation in adult education on life satisfaction of immigrants and natives. They worked with the data from SC6 which sampling started with adults. All publications using the NEPS data are listed on the NEPS website, which allows to filter publications by cohort, year, and type of publication.

More information can be found at the NEPS website.[19] Among others, it includes extensive information on the research questions per starting cohort, publications, and data access.

3.4.4 Socio-Economic Panel (SOEP)

The SOEP is an ongoing longitudinal multicohort study in Germany that started its yearly data collection in 1984 (Kroh et al., 2018). Kroh et al. (2018) provide information on sample sizes and attrition. Furthermore, the wave reports by Britzke and Schupp (2017; 2018; 2019) and the SOEP group (2020) provide information on the sampling of the dataset from 2016 until 2018.[20] To collect the SOEP data, interviewers aim to interview all members of a given survey household. Firstly, one person in the household is asked to complete a household questionnaire. Furthermore, adults are asked to complete the individual questionnaire. Information on children is also gathered. Children under the age of 16 complete a pre-teen questionnaire or their parents fill in a parent questionnaire or a mother-child questionnaire. Children above the age of 16 are able to fill in a youth questionnaire. Lastly, individuals are asked to complete a cognitive competence test, which assesses thinking abilities, such as memory and reasoning (CLOSER, 2021b).

The SOEP dataset contains a great deal of topics on diverse aspects of societal change. The wave reports by Britzke and Schupp (2017; 2018; 2019) and the SOEP group (2020) mentioned topics on household composition, work, financial situation, employment, earnings, health, emotional and behavioural problems, and satisfaction indicators. Moreover, educational and child-specific variables are measured on cognitive and personality development, and educational achievements.

The dataset, thus, allows for publications on a wide range of topics. Examples of topics are the integration of refugee children in and out of school (Gambaro et al., 2020), spending on children's education (Schroeder et al., 2015), the effect of day care on educational achievements (Spieß & Buechner, 2009), and private schools (Lohmann et al., 2009). Publications can be accessed on the website and can be filtered by publication series, topic, person, and year of publication.[21]

The SOEP data is available for transnational access. More information on how to obtain to dataset can be found online.[22] More information on the data is accessible on their website.[23]

3.4.5 Transformation of the Secondary School System and Academic Careers (TOSCA)

The Transformation of the Secondary School System and Academic Careers (TOSCA) is a multi-cohort study with five cohorts: TOSCA-2002, TOSCA-2006, TOSCA-10 (children in tenth grade), TOSCA-LAU (*Aspekte der Lernausgangslage und Lernentwicklung*; Aspects of the Learning situation and Learning development), and TOSCA-Sachsen (for the Saxony region). Data sampling, data size, number of waves, and wave frequency differ for each cohort. For example, the data collection of TOSCA-2006 started in 2006 and ended in 2012. Approximately 5,000 students in more than 150 secondary schools were followed bi-annually until 2012. At the end of 2016, the cohort was surveyed again. Consequently, the TOSCA-2006 cohort counts for five waves. Unlike TOSCA-2006, TOSCA-10

19 https://www.neps-data.de/Mainpage

20 The wave reports can be found at https://www.diw.de/de/diw_01.c.798307.de/soep_annual_report.html?nop=&id=diw_01.c.798307.de&von=0

21 https://www.diw.de/en/diw_01.c.629929.en/soep/soep_research_infrastructure_publications.html

22 https://www.diw.de/en/diw_01.c.601584.en/data_access.html

23 https://www.diw.de/en/diw_01.c.615551.en/research_infrastructure__socio-economic_panel__soep.html

only had two measurement points: one in 2007 and another one in 2014. As part of the data collection for this cohort, approximately 2,500 children in the tenth grade were examined. An overview of the data characteristics per cohort can be found at the TOSCA webpage.[24]

The study aimed to investigate the capabilities and constraints of the upper secondary school system. The TOSCA-dataset includes information on the transition from secondary school into university and vocational training. Furthermore, psycho-social variables were taken into account. (Eberhard Karls Universität Tübingen, 2021)

TOSCA data have been used by several publications on different topics. These include educational effectiveness (Neumann et al., 2011), effects of socioeconomic background (Parker et al., 2012), the role of motivation (Vasalampi et al., 2014), and the selection or socialisation effects with regard to personality traits (Jonkmann et al., 2014).

3.5 Hungary

3.5.1 The Panel of Linked Administrative Data (Admin1, Admin2, and Admin3)

The Panel of Linked Administrative Data is a longitudinal dataset based on administrative data linkages. The database was collected by the Databank of the Centre for Economic and Regional Studies (CERS), which integrated a survey and different administrative datasets. Based on those datasets, three data linkages were created: Admin 1 contains data from 2002 to 2009, Admin 2 covers 2003 to 2011, and Admin 3 has data from 2003 to 2017. These datasets have a sample size of nearly half the Hungarian population.

The dataset contains information on topics such as healthcare, the labour market, social transfers, firms, and education (Sebők, 2019b). More precisely, the topic of education includes details on all levels of education, educational outcomes, the Hungarian National Assessment of Basic Competencies (NABC), individual background data, school characteristics, and other contextual characteristics.

A selected list of publications using the data provided by the CERS Databank can be found at their website.[25] Topics include the efficiency of education (Hermann, 2020) and the link between educational achievements and employment (Hermann et al., 2020; Molnár, 2020). It should be noted that the list does not organise the selected publications by the dataset that was used.

The linked administrative data are available at the CERS Databank.[26] The technical details on the Panel of Linked Administrative Data were written in Hungarian by Sebők (2019a). There is, however, an English working paper of the article available online.[27] Additionally, more information can be found on the CERS Databank website.[28]

3.5.2 Hungarian Life Course Survey (HLCS)

The Hungarian Life Course Survey (HLCS) was an individual panel survey carried out annually by TÁRKI Social Research Institute between 2006 and 2012. The base of the sampling frame was the NABC conducted amongst eighth grade students at the end of the school year 2005-2006. The dataset consists of six waves. The initial sample contained 10,020 students, oversampling those with special needs and those in the lower third of competence scores. The sample size of the sixth wave was 7,092 students.

24 https://uni-tuebingen.de/en/fakultaeten/wirtschafts-und-sozialwissenschaftliche-fakultaet/faecher/fachbereich-sozialwissenschaften/hector-institut-fuer-empirische-bildungsforschung/forschung/aktuelle-studien/tosca/tosca-kohorten-im-detail/
25 https://adatbank.krtk.mta.hu/en/publikaciok/
26 https://adatbank.krtk.mta.hu/en/adatbazisok/adatkeresek-menete/
27 https://kti.krtk.hu/wp-content/uploads/2019/12/BWP1902.pdf
28 https://adatbank.krtk.mta.hu/en/

The HLCS examines the secondary school career and higher education opportunities, as well as the transition from school to work, of Roma and non-Roma students in Hungary. One of the main goals of the HLCS was to analyse the educational attainment and the disadvantages at school of the given cohort. Furthermore, the survey focused on the inequality of opportunities and dropout rates. Lastly, the HLCS examined life-style in a broader sense, with regard to leisure, network, substance consumption, crime, housing, and wealth. Each wave centred around different issues: the earlier waves focused more on family effects and early childhood, while school completion and further studies became significant from the fourth wave onwards, followed by transition to the job market in the last two waves.

Several articles were published based on the HLCS, often in combination with results of the NABC. Kertesi and Kézdi (2011) connected the national standardised test scores from the NABC with the HLCS in order to decompose the test score gap between Roma and non-Roma students. Keller (2014) used them to analyse whether self-assessment correlates with pupils' parental backgrounds and how these differences impact educational choices. Horn et al. (2016) looked into the mechanisms of the Hungarian education system that contribute to inequality based on information from the two datasets on tracking, school transitions and dropout. Hajdu et al. (2014) examined the situation of Roma youth in Hungarian secondary schools using the HLCS dataset. Horn and Keller (2015) used HLCS data when looking into the gender wage gap following labour market entrance of secondary school graduates. Furthermore, Horn (2016) investigated the effectiveness of apprenticeship training on youth employment.

Transnational access to the HLCS dataset is available on request addressed to the TÁRKI Social Science Data Archive in Hungary.[29] The survey documentation in general is available in Hungarian and English, while the variables in the files are in Hungarian. More information on how to use the dataset can be found online.[30] Furthermore, the TÁRKI data archive catalogue is available at http://old.tarki.hu/cgi-bin/katalogus/tarkiser_en.pl. When opening this website, look for data ID's TDATA-H81 to TDATA-H86 which contain further information on the dataset per wave.

3.5.3 National Assessment of Basic Competencies (NABC)

The NABC is an annually standard-based assessment compulsory for every school in Hungary. It examines the mathematical and reading literacy and is designed similarly to the OECD PISA survey (Schiltz et al., 2019). Additionally, background questionnaires on the characteristics of the students and the school are completed (Balázsi, 2006). The NABC follows cohorts of students every two years from sixth to tenth grade. This results in a sample size of approximately 100,000 students per grade and per year. The dataset is linked to student level identification numbers allowing for more detailed analyses.

The NABC aims to examine whether students are 'able to use their skills and knowledge in everyday life and use it as a basis for lifelong learning' (Balázsi, 2006, p. 1). Official publications that used the NABC dataset for their research can be found at the *Oktatási Hivatal* (Education Authority) website (2012).[31] Additionally, scientific articles using the data in peer-reviewed international journals have been published. As mentioned in Subsection 3.5.2, Kertesi and Kézdi (2011) used the HLCS and NABC datasets in their research. Horn (2013) investigated the effects of early selection on the inequalities of opportunity. Lastly, Schiltz et al. (2019) estimated the impact of high achieving peers leaving their primary school classes to go to elite academic tracks on student achievement, behaviour, and aspirations for higher education.

More information on the NABC dataset can be found in Hungarian as well as in English. Hermann and Molnár (2008) and the Oktatási Hivital (2012) website provide more information in Hungarian.

29 https://adatbank.tarki.hu/en/
30 https://tarki.hu/eng/adatbank
31 https://www.oktatas.hu/kozneveles/meresek/kompetenciameres/alt_leiras

In English, the conference paper of Balázsi (2006) gives a more detailed description of the NABC database.

3.6 The Netherlands

3.6.1 Longitudinal dataset from the Netherlands

The longitudinal dataset from the Netherlands contains information on (i) expectations and perceptions of eight characteristics of the learning environment and (ii) learning style characteristics. Based on the dataset, several articles have been published. Könings et al. (2008) investigated whether a new learning environment meets students' expectations. Furthermore, Könings et al. (2011) examined the match between students' lesson perceptions and their preferences about different characteristics of modern education. Lastly, the effects of a school reform on the changeability of students' preferences for different aspects of a learning environment were studied by Könings et al. (2012).

The data were collected over a period of three school years from five secondary schools in the Netherlands. Students filled out two questionnaires in ninth grade (or year three where students were on average fifteen years old) on learning style characteristics and expectations of the curriculum innovation in tenth grade. One year later, the now tenth-graders completed two questionnaires on learning style characteristics and their perceptions of the curriculum in tenth grade. Finally, the same two questionnaires as in tenth grade were completed in eleventh grade.

It is important to note that access to this dataset is not possible as there is no consent from participants to do this. At the time of the study, asking for consent was not yet common practice.[32]

3.7 Switzerland

3.7.1 Assessment of Students' Attitudes towards Studying (ASAtS)

The full-scale pilot of the Assessment of Students' Attitudes towards Studying (ASAtS) was conducted at the beginning of the academic year 2011-2012 at the University of St. Gallen. All 1,200 students who started the first year were asked via e-mail to fill in the questionnaire prior to their studies. 820 students completed the questionnaire. Throughout the academic year, students were asked to complete the questionnaire three more times: after handing in their first assignment in the first semester, in the middle of their second semester after receiving their grades from the first semester, and after the end of their first year (Brahm et al., 2017). The results of the longitudinal study can be found in Brahm et al. (2017).

The ASAtS dataset contains information on the development of economy and management students' study motivation in their transition from high school to university. Brahm and Jenert (2015) explained in their research the development and validation of the questionnaire that measures how students experience higher education. The questionnaire consisted of seven parts:

1. demographics (e.g., age, bachelor, nationality, gender, and former schooling);
2. attitude towards the university as an institution (e.g., reason for choosing the university);
3. attitude towards studying (e.g., self-efficacy);
4. attitude towards learning (e.g., autonomy in the learning process);
5. motivation (e.g., intrinsic and extrinsic motivation);
6. perception of the assessments (e.g., assessment quality);
7. outcome variables (e.g., attainment of study goals and satisfaction with first year).

32 This clarification was given in the expert questionnaire.

Brahm and Jenert (2015) included the questionnaire in the appendix of their article. Furthermore, they lay out how they developed and validated an instrument for the assessment of attitudes towards two objects:

1. the higher education institution;
2. the process of studying as an activity.

Different factors were measured within these objects. For the object 'higher education as an institution' these factors included: overall attitude towards the university, normative behaviour, and goals of the university. The factors within the object 'studying as an activity' consisted of: atmosphere among students, joy and anxiety when studying, previous experience in school, students' goals for studying, quality of lecturers, self-efficacy, and own activity in courses.

3.7.2 Transitions from Education to Employment (TREE)

Transitions from Education to Employment (TREE) is a longitudinal multicohort panel study following compulsory school leavers in Switzerland throughout their post-compulsory educational and labour market pathways. Because no questionnaire was completed on this dataset, information on the data collection was obtained from Gomensoro and Meyer (2017), Hupka-Brunner et al. (2021), TREE (2016) and the website of the University of Bern (2021). The TREE dataset has two cohorts with annual survey intervals up until cohort members are approximately 23 years old. Afterwards, follow-up surveys track the cohort members throughout their life. The first TREE cohort (TREE1) used the PISA survey of 2000 of ninth graders who left compulsory school that same year as their baseline survey. It has a sample size of 6,343 individuals. The cohort members completed annual surveys up until 2007 and filled out the (currently) last survey in 2019. The next survey is planned for 2024. The cohort members of TREE1 have been followed for a period of more than twenty years and have reached an average age above 35 years. The second TREE cohort (TREE2) started in 2016 with ninth graders who left compulsory school that year. TREE2 has a samplesize of 9,762 individuals. Unline TREE1, TREE2 used the 2016 *Überprüfung des Erreichens der Grundkompetenzen* (ÜGK) as the baseline survey.[33] The ÜGK is a national mathematics testing scheme that assesses the attainment of educational standards. TREE2 will conduct the surveys annually up until 2022.

TREE follows cohort members when they transition into adulthood and employment. It provides 'comprehensive data for the analysis of post-compulsory education, employment, and other pathways (e.g., family and household situation, income/financial situation, critical life events, social integration and participation, psycho-social personal characteristics, health and wellbeing)' (Gomensoro & Meyer, 2017, p. 209). This enables publications on a variety of topics such as the role of vocational education in the creation of gender segregation in employment (Heiniger & Imdorf, 2018), school to work transition (Imdorf et al., 2014; Müller & Wolter, 2014), the efficiency of secondary education (Scharenberg et al., 2017), the interplay of educational success and wellbeing (Samuel, 2014), and the interaction between social background and higher education (Murdoch et al., 2016, 2017). A complete list of publications can be found at the TREE website.[34] It is possible to sort the publications according to author, publication year, language, and type of publication.

The TREE data is publicly available for scientific use. More information on the transnational access to the data can be found online.[35] Furthermore, the TREE website provides more detailed information on the dataset.[36]

33 More information on ÜGK can be found at http://uegk-schweiz.ch/
34 https://www.tree.unibe.ch/results/scientific_publications/index_eng.html
35 https://www.tree.unibe.ch/study_profile/index_eng.html
36 https://www.tree.unibe.ch/index_eng.html

3.8 The United Kingdom

3.8.1 Millennium Cohort Study (MCS)

The Millennium Cohort Study (MCS) is a birth cohort study carried out by the Centre for Longitudinal Studies (CLS; 2021). The study is conducted in the four countries of the United Kingdom, namely England, Scotland, Wales, and Northern Ireland. The aim of the MCS is to follow children born between September 2000 and January 2002 throughout their early childhood and into adulthood. Data collection started in 2001-2002 when cohort members were nine months old. The dataset has currently seven waves with the last data collection happening in 2018 when cohort members were seventeen years old. The next wave is planned for 2022 when cohort members are 22 years old.

The MCS maps the physical, socio-emotional, cognitive and behavioural development of the cohort members (Centre for Longitudinal Studies, 2021). The dataset deals with topics related to gender roles, personality traits, and feelings about school and the future (Burston et al., 2017). Furthermore, background information on the economic circumstances and families of respondents is also collected (Centre for Longitudinal Studies, 2021).

More detailed information on the content and collection of the MCS is available in technical reports of the different waves (Burston et al., 2017; Centre for Longitudinal Studies, 2019; Gallop et al., 2013; GfK NOP Social Research, n.d.; Gray et al., 2009, 2010; Plewis et al., 2007). Moreover, the MCS website includes details on data access and a list of publications which allows filtering according to dataset and year.[37]

3.9 Cross-national longitudinal datasets

3.9.1 European Community Household Panel (ECHP)

The ECHP is a panel survey conducted between 1994 and 2001. A sample of households and persons (from the age of 16) was interviewed yearly for eight years, which resulted in a total of eight waves (European Communities, 2003). The data were collected in fourteen European member states, namely: Belgium, Denmark, Germany, Ireland, Greece, Spain, France, Italy, Luxembourg, the Netherlands, Austria, Portugal, Sweden, and the United Kingdom (Eurostat, n.d.). More information on the sample sizes and attrition per year and per country can be found in the EC Household Panel 'Newsletter' by the European Communities (2003).

The dataset contains information on a variety of different topics related to living conditions. Eurostat (2003) presented an overview of the available data. There were questions on the household level, such as financial situation and children, and on the individual level, such as employment, health, social relations, and migration. More importantly, the interviews examined the individuals' training and education. The interviews collected information on the highest educational degree that was obtained, whether vocational education and training was attended, and whether the education was payed for by their employer. Bassanini (2006) used data from the ECHP to investigate the impact of adult education and training on average wage and employment security of different labour market groups in EU countries. More information on the variables included in the dataset and how to access the microdata, can be found online.[38]

37 https://cls.ucl.ac.uk/cls-studies/millennium-cohort-study/
38 https://ec.europa.eu/eurostat/web/microdata/european-community-household-panel

3.9.2 Trends in International Mathematics and Science Study (TIMSS)

TIMSS is a quasi longitudinal study. TIMSS is conducted with fourth and eighth graders every four years since 1995 with the last wave in 2019 and the next one in 2023 (International Association for the Evaluation of Educational Achievement, 2021). The fourth grade cohort is assessed four years later at the eighth grade which results in a quasi-longitudinal design (International Association for the Evaluation of Educational Achievement, 2021). The design includes a stratified multistage sample technique with sampling the schools in a first stage and sampling classrooms from the targeted grade in the second stage (Olson et al., 2008). The sample sizes of the waves can be found in the technical reports, which can be accessed online.[39] Several countries around the world participate in the study. An overview of the participating countries including the waves in which they participated can be found at the TIMSS website.[40] It includes the following (EU-)countries which lie within the scope of this research: Austria, Belgium, Bulgaria, Croatia, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Malta, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom.

The TIMSS assesses not only mathematics and science achievements but also contextual factors at school and at home that are associated with the students' achievements. Firstly, Olson et al. (2008) presented the two domains within the mathematical and science assessments: cognitive and content. The cognitive domains are the same for mathematics and science assessments, and for the fourth and eighth graders. It assesses knowing, applying, and reasoning among the students. Within mathematics, the content domains assess numbers, geometric shapes and measures, and data display for fourth graders. For eighth graders numbers, algebra, geometry, and data and chance are being assessed. Science assessments include life science, physical science, and earth science for the fourth graders while the eighth graders are being assessed on biology, chemistry, physics and earth science.

Secondly, the study provides details on the contextual factors at school and at home that are associated with the students' achievements. This includes questions on the students' home environment and the organisation of the education system (International Association for the Evaluation of Educational Achievement, 2021).

More information on publications,[41] and transnational access to the data can be found online.[42]

39 https://timssandpirls.bc.edu/isc/publications.html

40 https://www.iea.nl/studies/iea/timss

41 https://timssandpirls.bc.edu/isc/publications.html and https://www.iea.nl/publications.

42 https://www.iea.nl/data-tools/repository/timss

Table 2. Summary of the datasets in the inventory

Country	Dataset	Design			Content			
		Type of longitudinal dataset	Data collection	Sampling (1st wave)	Content	Respondents	TNA to micro-data	Technical report
Belgium (Flemish region)	SiBO	Cohort and follow-up	**Start** 2002 **End** 2011 **Waves** Depends on variable (yearly/half-yearly)	**Design** There were different samples of schools. Students were followed throughout primary education until the first year of secondary education **Size** N=ca. 6,000	(i) Student level variables: socioeconomic background, cognitive development, socio-cognitive variables, social development, dynamic and affective development, class climate and learning environment, and environmental characteristics (ii) Class level variables: class characteristics, teacher characteristics, teacher beliefs/opinions, teacher perceptions, and didactics (iii) School level variables: board of directors and school team	Students, parents, teachers, school team, school principal, and observers **Level of education** From last year of kindergarten until grade 6	Yes **Conditions** Available upon request (i) The SiBO-project or policy research centre should be mentioned (ii) The government must be informed on the results before the general public **Language of documentation** Dutch	Yes **Language** Dutch
	LiSO	Cohort and follow-up	**Start** 2013 **End** 2019 **Waves** 7 (yearly)	**Design** Two-stage sampling design: (i) Sample of schools within a specific region (ii) Sample of seventh grade students within those schools (school year 2013-2014) **Size** N=ca. 6,500	The LiSO-data includes variables on cognitive outcomes (e.g., mathematics and reading comprehension), non-cognitive outcomes (e.g., well-being), and school trajectory characteristics (e.g., choice of study)	Students, parents, teachers, school team, and school principal **Level of education** Secondary education: from grade 7 until 12	Probably **Conditions** At the moment, only SONO researchers have access to the anonymised data. It may be possible to release the data to other researchers (after signing an agreement) but no final decision has been made on this yet **Language of documentation** Dutch	Yes **Language** Dutch

Country	Dataset	Design			Content			
		Type of longitudinal dataset	Data collection	Sampling (1st wave)	Content	Respondents	TNA to micro-data	Technical report
Denmark	DAR	Administrative panel data	**Start** Depends on educational register **End** Ongoing **Waves** Depends on register (depends on register)	**Design** Population wide administrative data **Size** Danish population of ca. 5,6 million people	The dataset contains the information that is covered by the admin registers such as the pupil register, grade register, test register, well-being, absenteeism, institutional register, school district register, day care, and after school care Background information is available in non-educational registers for pupils, siblings, parents, and grandparents. The teachers are currently not registered but will be added in the near future	Danish population **Level of education** All levels	Yes **Conditions** Non-Danish researchers can get micro-data access through an affiliation to a Danish authorised environment. Collaboration with Danish researchers is recommended to make sense of the raw data **Language of documentation** Danish	Yes **Language** English
	DNT	Multicohort	**Start** 2010 **End** Ongoing **Waves** (Yearly)	**Design** There are 12 different tests for different grade levels and subjects. Every test can be used 3 times per student. One of these tests are mandatory to students in the public school system, the other two are voluntary for the schools **Size** N=ca. 50,000 students in each year, cohort, and test	The dataset contains information on the tests, such as the item responses, total scores, and time per item	Students in public school **Level of education** Public primary and lower secondary school	Yes **Conditions** Access to micro and aggregated level data can be granted for research purposes by the Ministry of Education **Language of documentation** Danish	Yes
France	ELFE	Birth cohort	**Start** 2011 **End** Ongoing (until 2031) **Waves** (depends on the stage)	**Design** Midwives in 344 maternity units in metropolitan areas helped targeting cohort members who were born during four selection periods – representing the four seasons – in 2011 **Size** N=18,324	The ELFE study aims to investigate how the environment affects the development, health, and socialisation of children. They include factors such as family and relationships, education, school, well-being, leisure activities, and school-learning	Parents, children, doctors, and teachers **Level of education** From birth until age 20	Yes **Conditions** Available on request **Language of documentation** French	Yes **Language** French English

Country	Dataset	Design			Content			
		Type of longitudinal dataset	Data collection	Sampling (1st wave)	Content	Respondents	TNA to micro-data	Technical report
Germany	FiD	Cohort	**Start** 2010 **End** 2013 **Waves** 3 (yearly)	**Design** The FiD study was designed to target four different types of families: (i) families with children born between 2007 and 2010, (ii) single-parent families, (iii) low-income families, and (iv) families with more than two children **Size** N=17,002 individuals	The FiD dataset contains general information on education, past and current labour market experiences, earnings and income, housing characteristics, health, and life satisfaction More specifically, FiD focused on children and partnership	Families and children **Level of education** From age 0 to beyond (youth and adults)	Yes **Conditions** Available on request **Language of documentation** German English translation	Yes **Language** English
	LISA	Cohort	**Start** 2005 **End** 2013 **Waves** 6 (ca. every 18 months)	**Design** The basic design of the sample used was a two-stage stratified cluster design. The first stage consisted of a sampling of schools, and the second stage of sampling of intact classrooms from the target grade in the sampled schools. A sample of schools was selected according to the school types which exist in the German secondary school system **Size** N=1,508	The initial emphasis of the LISA project was on a longitudinal survey of the development of reading comprehension and reading motivation as well as on an investigation of influential factors such as the students' background or the learning environment	Students, their parents, and their teachers **Level of education** Grade 5 to 13	Yes **Conditions** Available on request **Language of documentation** German	No

Country	Dataset	Design			Content			
		Type of longitudinal dataset	Data collection	Sampling (1st wave)	Content	Respondents	TNA to micro-data	Technical report
	NEPS	Multicohort	**Start** Between 2009 and 2012 **End** Ongoing **Waves** (Yearly)	**Design** The sampling design differs for each cohort. For the starting cohort 1, the target population 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 **Size** N=3,481	The focus of NEPS is mainly on educational careers and educational outcomes.	Parents, children, childcare workers (and, at later points of time, teachers) **Level of education** This depends on the cohort. Currently, the starting cohort 1 data is publicly available until the age of 5 years	Yes **Conditions** Only the NEPS Scientific Use Files in the On-site version include very sensitive microdata with the lowest level of anonymisation. The analysis of this data is only possible at the Leibniz Institute for Educational Trajectories in Bamberg **Language of documentation** German English translation	Yes **Language** English German
	SOEP	Multicohort	**Start** 1984 **End** Ongoing **Waves** (yearly)	**Design** The respondents are households that were selected by random-walk. All samples are multi-stage random samples **Size** N=11,366 individuals	Range of societal topics, including education, training, and child development	Individuals in households including children from their first year on **Level of education** All kinds	Yes **Conditions** Available on request **Language of documentation** German English translation	Yes **Language** German English
	TOSCA	Multicohort	**Start** Depends on cohort **End** Depends on cohort **Waves** Depends on cohort	**Design** Differs for each cohort **Size** Depends on cohort	TOSCA is mainly about educational careers	Students, parents, and teachers **Level of education** Upper secondary and transition into university and vocational training	Yes **Conditions** Unclear **Language of documentation** German	Yes **Language** English

Country	Dataset	Design			Content			
		Type of longitudinal dataset	Data collection	Sampling (1st wave)	Content	Respondents	TNA to micro-data	Technical report
Hungary	Admin	Administrative panel data	**Start** 2003 **End** 2017 **Waves** Unclear (monthly)	**Design** 50% random sample of a nearly full population covered administrative register of the National Health Insurance Fund of Hungary **Size** Half of the Hungarian population	The dataset contains information on educational careers, educational outcomes (e.g., labour market status, wage, occupational information linked with study data), tests (the National Assessment of Basic Competencies), educational career data (e.g., tertiary, secondary and elementary education studies), individual background data (e.g., gender, age, and regional information), school characteristics (e.g., type, regional information, and school size), and other environmental characteristics (e.g., regional data)	Hungarian population **Level of education** Nursery, elementary, high-school and vocational education, higher education	Yes **Conditions** Available on request **Language of documentation** Hungarian	Yes **Language** Hungarian English
	HCLS	Cohort	**Start** 2006 **End** 2012 **Waves** 6 (yearly)	**Design** The base of the sampling frame was the NABC conducted amongst the 8th grade students at the end of the school year 2005/2006. Those who completed the competence survey or the shortened competence test (SNI sample pupils that for special education needs) could take part in the sample. The actual sampling frame was formed by students that filled out the attached family background questionnaire and sent back the parent's declaration form on the youth's participation in the research **Size** N=10,022	It examines the secondary school career and higher educational opportunities of Roma and non-Roma students in Hungary. One of the main goals of the HLCS is to analyse the educational attainment and the disadvantages in school of the given cohort. The survey also focuses on the inequality of opportunities and the dropout rate. Moreover, it examines life styles in a broad sense (free time, network, crime, housing, and wealth)	Primarily students and sometimes parents **Level of education** Ca. 13 years old (8th grade) until ca. 20 years old	Yes **Conditions** Available on request Proper reference should be made **Language of documentation** Hungarian English translation for large part	Yes **Language** Hungarian

Country	Dataset	Design			Content			
		Type of longitudinal dataset	Data collection	Sampling (1st wave)	Content	Respondents	TNA to micro-data	Technical report
	NABC	Cohort	**Start** 2008 **End** Ongoing **Waves** (yearly)	**Design** Full cohort of 6th, 8th, and 10th grade students (compulsory). Students are followed through the grades. **Size** N=ca. 100,000 per year per grade	The dataset contains information on mathematical and reading literacy, detailed family background, and information on schools and teachers	Students **Level of education** Lower and upper secondary	Yes **Conditions** Access to data is only through the secure server of the CERS-IE and if a researcher of the CERS-IE is a co-author/supervisor or if the international scholar has legal authorisation from *Oktatási Hivatal* (the Education Authority). Administrative approval of the CERS databank is compulsory **Language of documentation** Hungarian	Yes **Language** Hungarian English
The Netherlands	Longitudinal dataset from the Netherlands	Cohort	**Start** 2003 **End** 2005 **Waves** 3 (yearly)	**Design** The researcher contacted 6 secondary school in the region of the university to ask if they were interested to participate. Five schools agreed Students filled out 2 questionnaires in school year 3, then 2 questionnaires in year 4, and finally the same questionnaires in year 5. All pupils present at school at the day of data collection participated **Size** N=842	The dataset contains information on (i) the expectations and perceptions of eight characteristic of the learning environment and (ii) learning style characteristics	Students **Level of education** Secondary education: 9th grade until 11th grade	Legally prohibited **Language of documentation** Dutch	No

Country	Dataset	Design			Content			
		Type of longitudinal dataset	Data collection	Sampling (1st wave)	Content	Respondents	TNA to micro-data	Technical report
Switzerland	ASAtS	Panel	**Start** August 2011 **End** August 2012 **Waves** 4 (quarterly)	**Design** All 1,200 students who started the first year at the University of St. Gallen were asked via e-mail to fill in the questionnaire **Size** N=820	The dataset contains information on the attitudes of students towards the study environment. Two attitude objects were investigated: (i) attitude towards the university as an institution and (ii) attitude towards studying as an activity. The two attitude objects measured affective, cognitive, as well as behavioural items	Students **Level of education** Bachelor students (ca. 22 years old)	Yes **Conditions** Available on request. **Language of documentation** German English translation	Yes **Language** English
	TREE	Multicohort and follow-up	**Start** 2000 **End** Ongoing **Waves** (yearly)	**Design** The design is based on the Swiss PISA sample. **Size** N=6,343	Post-compulsory educational pathways of cohort members by following them into adulthood and employment	Individuals **Level of education** Grade 9 and beyond	Yes **Conditions** None **Language of documentation** English French German	Yes **Language** English French German
UK	MCS	Cohort	**Start** 2001-2002 **End** Ongoing **Waves** (Depends on the wave)	**Design** The study population were children born between 2000 and 2002 across England, Scotland, Wales, and Northern Ireland. The children had to be nine months old during the first wave **Size** N=18,818	The MCS maps the physical, emotional, and cognitive development of the cohort members. Furthermore, it collects information on family characteristics, and feelings about school and the future	The millennium children, their parents, **Level of education** Starting from nine months old until 17 years and through adulthood	Yes **Conditions** Available on request **Language of documentation** English	Yes **Language** English
Cross-national	ECHP	Household panel	**Start** 1994 **End** 2001 **Waves** 8 (yearly)	**Design** Household sample **Size** Depends on country	The dataset contains information on living conditions in a broad sense. The topic of training and education is a small part of the dataset and involves questions about educational training provided by the employer or the highest level of education one has completed	Households and individuals **Level of education** Wide variety	Yes **Conditions** Available on request and by submitting a research proposal **Language of documentation** English	No

Country	Dataset	Design			Content			
		Type of longitudinal dataset	**Data collection**	**Sampling (1st wave)**	**Content**	**Respondents**	**TNA to micro-data**	**Technical report**
	TIMSS	Quasi-longitudinal	**Start** 1995 **End** Ongoing **Waves** (quadrennial)	**Design** Two-stage stratified cluster design. The stages: (i) sampling schools (ii) sampling intact classrooms from the targeted grade **Size** Depends on country	Mathematical and science assessments. Furthermore, contextual factors at school and home	Students, teachers, and principals **Level of education** 4th and 8th graders	Yes **Conditions** Available on request **Language of documentation** English Language of participating country	Yes **Language** English

* When possible, the acronyms of the datasets were used. Datasets that have no official name were given a standard name after the country in which they were collected. The number of waves were only mentioned when the data collection had ended.

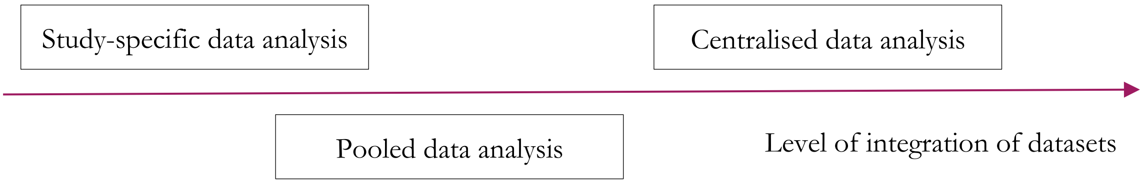
4. Harmonisation

One of the aims of this report is to explore the possibilities of sharing or merging longitudinal datasets for comparative research. Comparative research can be defined as 'describing and explaining the similarities and differences of situations or consequences among large scale social units such as regions, nations, societies and cultures' (Smelser, 1973, p. 1). The similarities and differences can be analysed both quantitatively and qualitatively. We distinguish four broad types of comparative research, each addressing a different type of research question: (i) descriptive comparison, (ii) basic explanatory research, (iii) comparison of relations, and (iv) comparative explanatory research (Esser & Vliegenthart, 2017). *Descriptive comparisons* describe the occurrences of an event or phenomena of interest and the variation in occurrences between cases (Esser & Vliegenthart, 2017). A second level of research questions are *basic explanatory questions*, where the author wants to figure out whether certain variables at the unit level impact other variables measured at the same level in different contexts (Esser & Vliegenthart, 2017; Schuck et al., 2013). A *comparison of relationships* aims to determine to which extent two or more variables co-vary (Esser & Vliegenthart, 2017). It is thus a 'robustness check to determine whether a relationship holds in various situations' (Esser & Vliegenthart, 2017, p. 15; application in Holtz-Bacha & Norris, 2001). A last type is named *comparative explanatory research*. In addition to the previous type of research questions, the different relationships across units are explained by taking the characteristics of these units into consideration (Esser & Vliegenthart, 2017; Schuck et al., 2016). Two or more levels are combined, in which the first is nested in the second and so on (e.g., classes in schools; Hanitzsch & Berganza, 2012; Schuck et al., 2016).

4.1 Integration of data for comparative research

Comparative analysis can thus be applied on different levels. This also has an influence on the structure of the data that is needed for the analysis. Building on the framework of Fortier et al. (2017) we distinguish three ways to deal with information about multiple contexts: (i) study-specific data analysis, (ii) pooled data analysis, and (iii) centralised data analysis (see also Figure 1). In *study-specific data analysis* the researchers first analyse the data per country, followed by a meta-analysis of the study-level estimates. This type of data can only be used for descriptive or basic explanatory research. If the researchers also want to compare relations or explanations across contexts, a higher level of integration of datasets is needed. For this, the researchers can *integrate the data* from the different studies into one big database and analyse this dataset as a whole (Gaye et al., 2014; Wolfson et al., 2010). While the approaches for the studies that are pooled might be different, the researchers can also apply the exact *same approach in different countries* to obtain (more) homogeneous country-specific datasets. This must unfortunately be imposed before the data are collected, and is thus not possible when combining datasets retrospectively.

Figure 1. Three ways to deal with information about multiple countries



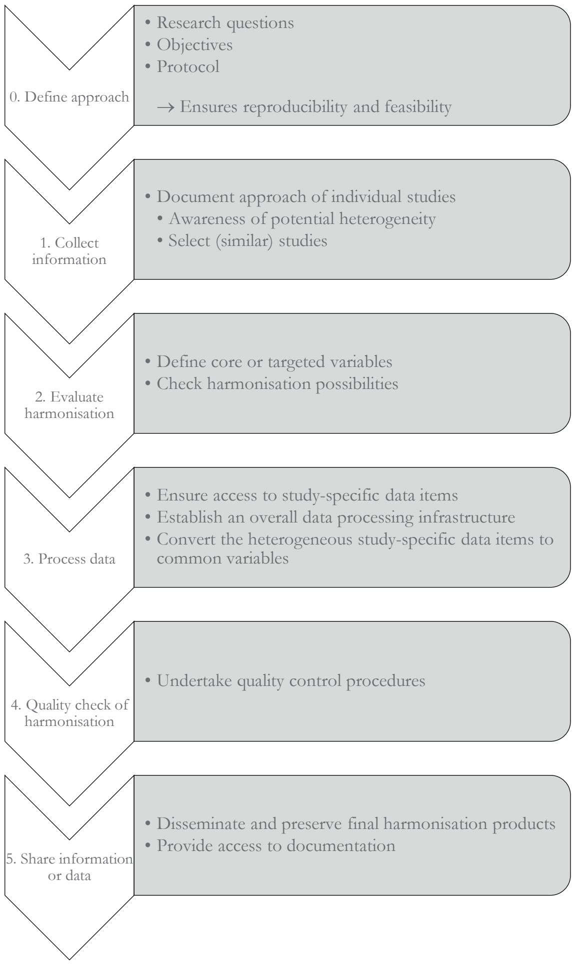
Since our goal is to determine whether, and how, combining the datasets from our inventory is possible, we will focus on the second type of data structure, namely the integration of data for comparative research. This integration can improve the quality and depth of the research design in several ways (Curran & Hussong, 2009; Fortier et al., 2017). A first advantage is that integration provides methods to test whether findings can be replicated across independent studies. If the same relationships or results are found in several similar studies, we can be more certain of the robustness of the results. Second, by pooling data from several samples, the number of observations in the analysed dataset increases. This improves the statistical power of the tests performed on this dataset. A third advantage lies in the increase in sample heterogeneity. Integrating datasets results in a more heterogeneous sample: the sample will be composed of more diverse individuals with regard to characteristics such as age, geographic location, or gender and in studies concerning fewer common events (e.g., deviant behavior), a higher rate of the behaviour or event of interest. This enables comparisons that would not be possible within the individual studies because of their small sample size. The researchers can, for example, perform a cluster analysis, which divides cases in several similar groups into clusters. Esser and Vliegenthart (2017) mentioned that there are various techniques to calculate the distance between the cases and determine the best way to cluster them into groups (see also Brüggemann et al., 2014). Additionally, the bigger sample size improves the stability of the model estimation. A fourth advantage is the broader psychometric assessments of theoretical constructs. Instruments approaching the same concept, but focusing on different aspects of this concept, can be integrated.

4.2 Data harmonisation

An important prerequisite to combining data is that the information that is compared or combined is very similar across the different studies. Fortier et al. (2017) stress the importance of data harmonisation for integration in their paper as follows: '[…] to ensure content equivalence across studies and minimise measurement/assessment error that can cause bias or impair statistical power, all such approaches require use of harmonised data. Essentially, data harmonisation achieves or improves comparability (inferential equivalence) of similar measures collected by separate studies' (p. 104).

Based on a broad consultation process with research experts, Fortier et al. (2017) put together a manual on how to harmonise existing datasets (i.e., retrospective harmonisation). They identified six steps in the harmonisation process (Figure 2). At the beginning of the research (step 0), the research questions, objectives and timeline are defined. In the timeline, sufficient time needs to be reserved for the process of getting access to the data and the collection of information on the different data sources involved in the harmonisation. The first step concerns the collection of information on the different datasets and the selection of comparable studies. Based on information on the design, timeframe and subjects in the dataset, the researchers evaluate the harmonisation potential of the datasets and select compatible studies. A second step focuses on the core variables and the evaluation of their comparability across the datasets. In the third step, the harmonised data are analysed by using appropriate statistical methods or techniques. The fourth step comprises quality control procedures, for instance to assess heterogeneity between datasets. The final step is the dissemination and preservation of the final harmonisation products.

Figure 2. Five steps in data harmonisation



Source Fortier et al., 2017

Throughout the harmonisation and integration process, it is important that the researchers give sufficient attention to the comparability of the different datasets. In order to improve the validity of the analyses on an integrated dataset, it is important that the researchers evaluate heterogeneity caused by differences in sampling, the time frame of the study, the study design and measurements of key concepts.

A first source of heterogeneity is caused by **differences in sampling**. First of all, the researchers have to establish that the populations from which the samples are drawn, are similar or at least very comparable (Curran & Hussong, 2009). Second, the researchers need to make sure they are comparing data on the same level. Some studies collect data on the household level (or other levels such as class, country, and age), while others provide information about individuals. If sufficient information is available about the family (or class, country, and age)-composition, the information on the lower level can be collapsed or aggregated on a higher level to make comparison possible. To facilitate comparison of different results even further, researchers can standardise the results. The mean result is then set to zero, while the unit standard deviation within test or year coefficients can be interpreted as standard deviations. This however has the drawback that we can only draw inferences 'compared to other pupils', rather than whether a specific pupil improved over time (Beuchert & Nandrup, 2017;

Rambøll, 2013).[43] Finally, the researchers have to determine what type of sampling mechanism was used to obtain the sample. If there are differences in the sampling procedures, this must be taken into account when interpreting or comparing the results of different studies. When the probability of selection into the sample is known, the researchers can explicitly correct for the difference by including individual-specific sampling weights directly into the statistical analysis. This method is known as the design-based procedure (Neyman, 1934). When the probability is however unknown ('nonprobability sampling'; e.g., when applying a snowball method or quota-sampling), alternative methods, such as Fisher's model-based procedure (1922) can be applied.[44]

A second potential source of heterogeneity can be caused by **differences in the time frame of the data collection**. Every observation in a sample takes place in a specific time period. It is possible that specific events that occurred during a study (or in between two points of data collection) could account for an observed effect (Curran & Hussong, 2009). Therefore, researchers should aim to integrate datasets for which the data collection took place in a similar time period. If this is not the case, the time of observations should explicitly be included in the analyses.

A third source of heterogeneity is related to **differences in the design of the study**. First of all, the repercussions of the study can have an impact on the data. If the test results in one study can have repercussions for the individual or their teacher or school, while the other test(s) have no consequences, there might be a bias although both tests measure the outcome of interest in exactly the same way. For example, Beuchert and Nandrup (2017) point to the difference in stake between the Danish exit exams and the Danish National Test. The latter has a more informal character, since it focusses on identifying the student's teaching needs rather than forming an admission condition or a reason for sanctions. This can have an effect on the performance of pupils,[45] and thus make a comparison between the two results more difficult. Second, the survey method can affect the results. The way a survey is taken (by mail, telephone, personal, online) may cause instrument bias. As stated by Curran and Hussong (2009), it is nonetheless 'unrealistic and not useful to exhaustively identify, track and code the entire set of differences in design characteristics across the set of contributing samples' (p. 90). The researcher should nevertheless be aware of the existence of heterogeneity in design between datasets and include them in the quality checks of the harmonisation process (step 4).

A fourth source of heterogeneity between datasets is associated with heterogeneity in survey design, and relates to **differences in the measurement of key concepts**. These differences must be clear to 'create a valid and reliable aggregate measure that is sensitive to potential study differences on dimensions such as design characteristics, specific items administered, subject age, and calendar year' (Curran & Hussong, 2009, p. 91). Two concepts are of special interest here: measurement invariance and measurement comparability. *Measurement invariance* applies to studies that use the same items to measure a theoretical construct. The researchers must consider the extent to which a set of items reliably and validly assesses an underlying construct in a similar way across groups or over time (Horn & McArdle, 1992; Meredith, 1993; Millsap, 1995, 1997; Thurstone, 1947). *Measurement comparability* however is relevant for studies that use a partially or wholly different scale to assess a shared underlying concept. In this case all theoretical and empirical evidence that can strengthen the confidence in whether you are assessing the same construct within each individual sample as well as within the pooled sample in a psychometrically equivalent way must be considered (Curran & Hussong, 2009). appendix 4 provides an overview of the steps that need to be taken to make sure that the same theoretical construct similarly for all individuals across the datasets are measured.

Although the researchers should be aware of these issues, scientists are still debating on the appropriate methods to address these problems (e.g., Beck & Katz, 1995; Wilson & Butler, 2007). It is

43 The researcher can only compare the position compared to the class mean over time (e.g., at t1 1 standard deviation below the mean, and at t2 only 0.5 standard deviation below the mean). But if the distribution of the test scores in the class became less unequal, a smaller deviation from the mean does not necessarily mean that the specific student's test score increased.

44 A detailed description of this method can be found in Appendix 3.

45 This effect could be positive or negative, depending on the personality of the pupil and the approach of the teacher (Beuchert & Nandrup, 2017).

however not necessary to model and identify all sources of between-study heterogeneity independently for the purposes of study integration, since there are techniques that control more globally for between-study differences. The potential sources of problems are thus not all adjusted before combining the datasets, but the possible differences between the studies are taken into account when performing the subsequent statistical analyses on the combined dataset.

4.3 Integrative data analysis for comparative longitudinal research

Once the datasets are harmonised and integrated, the researchers can perform analyses on the newly generated dataset. In order to observe the heterogeneity between different studies, it is advised to use statistical techniques that take the hierarchical structure of the data into account.

Ideally, the researchers would apply multilevel longitudinal analysis. A multilevel model is characterised by a nested structure in which the individual is nested within a country or study. By adding this additional level, this method allows to estimate a model that simultaneously evaluates the main effects of the within-sample predictors (e.g., type of sampling mechanism, geographic location, and method of data collection) on the outcome and the effect of interaction between within-sample and between-sample predictors on the outcome (Curran & Hussong, 2009). Research often refer to this method as a 'random effect', since the relationships found at the lower level may vary across higher level units. The application of random-effects (RE) might raise two issues regarding the samples: (i) one must make sure that the data sets can be considered as random draws from a homogeneous population of datasets, and (ii) the number of independent samples must be sufficiently big to allow for a reliable estimation of the random variability between and within the samples. The literature does not specify how many samples must be available (Curran & Hussong, 2009), but in the general multilevel framework 20 to 30 samples is viewed as minimum (Kreft & de Leeuw, 1998). These conditions are unfortunately difficult to meet when working with integrated data. In that case, the researcher can work with fixed-effects model or clustered standard errors.

In a fixed-effects (FE) study membership is treated as a fixed and known characteristic of each individual observation nested within that study (Curran & Hussong, 2009). Study membership is indicated by a (dummy, effect or weighted effect) code that is entered in the fitted model (similar to Fisher's model-based inferential method, 1922; see also Appendix 3). An interesting advantage of this method is the possibility to estimate the multiplicative interactions between individual characteristics and study group membership. The FE framework treats the set of independent samples as fixed and known. As a result, we can only make inferences back to the specific samples under study, while in multilevel analysis we can make inferences back to an infinite population of samples. Also, we cannot disaggregate *within-sample effects, between sample effects and cross-level interactions*, in the FE framework. It is only designed to deal with unobserved heterogeneity between different groups in the data (Miller, 2017).

A final alternative is to work with clustered standard errors. This method can be performed on a dataset with a small number of units at the higher level (Esser & Vliegenthart, 2017). The researcher can use standard errors clustered at the higher level. This however has the drawback that the standard errors tend to be too small when the cluster sizes vary a lot (MacKinnon & Webb, 2017). Clustered standard errors also do not allow distinguishing between within-sample effects, between-sample effects and cross-level interactions.

5. The inventory: harmonisation possibilities

After presenting the inventory and describing the techniques and conditions for harmonisation, we discuss the harmonisation possibilities for the datasets in the inventory. We first outline the different steps in the matching process. After this, we describe the results for the datasets in general and for the different types of educational variables.

5.1 Steps in the matching process

We scanned for matches within the pool of datasets in the inventory where harmonisation was possible. Because data harmonisation includes access to study specific data to process the data (Fortier et al., 2017) TNA to the microdata is necessary. Thus, the longitudinal dataset from the Netherlands was eliminated because TNA to the microdata is legally prohibited. This resulted in a pool of eighteen datasets that were eligible for harmonisation.

To determine the steps in the matching process, we considered two of the four aspects of heterogeneity that needed to be considered when harmonising datasets: heterogeneity due to historical time and measurement (see Section 4.2). Those biases still need to be considered by the researchers when harmonising datasets. Furthermore, the biases in datasets regarding sampling, design, and variable operationalisations were not taken into account in this chapter. It is up to the researchers themselves to determine which differences they want to allow in their harmonisation based on the sampling and design, and to appropriately evaluate and correct for these biases in their research design.

In a first step, we scanned through the datasets to categorise variables in them. We distinguished two broad categories: educational and additional variables. Within educational variables three subcategories were differentiated: formal characteristics (e.g., degrees and early school-leaving), academic outcomes (e.g., mathematics test scores), and non-academic characteristics (e.g., cognitive development and school context). The additional variables included variables such as health, social development, and socio-economic background. These categories were used to compare the datasets. It should be noted that not all variables are known and access to the dataset should be requested to gain more precise information on the availability and operationalisation of variables in a given dataset. An overview of the categories we distinguished can be found in Table 3.

Secondly, we looked at the time period in which the longitudinal study was conducted and the age of the respondents the dataset covered. We selected age and time period as time units that should be similar in order to harmonise the dataset because of heterogeneity due to historical time. Hence, we considered datasets not comparable when no overlap could be detected based on time period *or* age (see Section 5.2).

In a final step, we checked which categories of the variables overlapped (see Subsections 5.2.1 and 5.2.2). Similar variables should be compared with each other when harmonising data. The researchers should take heterogeneity due to measurement into account when further exploring the harmonisation possibilities (see Section 4.2).

Table 3. Overview of the variables in the inventory

					Educational variables																				Additional variables						
					Formal characteristics						Academic outcomes				Non-academic characteristics																
Dataset	Country	Type	Time period	Age range	Adult education	Choice of study	Degrees	Early school-leaving	Grade retention	Years in formal education	Mathematics	Language	Other	Later educational outcomes	Cognitive development	Learning styles	Relationship between parents and school or teachers	School context	School policy	Students' attitude	Students' motivation	Students' self-efficacy	Students' well-being	Teacher characteristics	Environment	Family characteristics	Health	Pregnancy	Social development	Socio-economic background	Socio-economic outcomes
SiBO	Belgium	Cohort	2002-2011	5-11 (follow-up when 17)		x			x		x	x		x	x		x	x	x		x	x	x	x	x	x			x	x	
LiSO	Belgium	Cohort	2013-2019	11-18		x	x	x	x		x	x	x		x		x	x	x		x	x	x	x	x	x			x	x	
DAR	Denmark	Administrative	1981-present	All		x	x	x		x	x	x	x														x			x	x
DNT	Denmark	Multicohort	2010-present	7-14		x					x	x	x	x																x	
ELFE	France	Birth cohort	2011-present	0-(now 10)		x	x				x	x			x	x	x	x		x		x	x	x	x	x	x	x	x	x	x
FiD	Germany	Cohort	2010-2013	All		x	x	x	x	x	x	x	x			x	x			x	x	x	x	x	x	x			x	x	x
LISA	Germany	Cohort	2005-2013	10-19		x					x	x	x		x				x	x	x			x		x			x	x	
NEPS	Germany	Multicohort	2009-present	0-79	x	x	x	x		x	x	x	x		x			x		x	x				x	x			x	x	x
SOEP	Germany	Multicohort	1984-present	All		x	x	x	x	x	x	x	x		x	x	x			x	x	x	x	x	x	x	x		x	x	x
TOSCA	Germany	Multicohort	ca. 1995-2016	15- (ca. 35)		x					x	x	x		x			x	x		x	x	x						x	x	x
Admin	Hungary	Administrative	2003-2017	All			x				x	x	x					x							x		x			x	x
HLCS	Hungary	Cohort	2006-2012	13-20			x	x	x	x	x	x	x												x	x	x		x	x	x
NABC	Hungary	Cohort	2008-present	11-16		x					x	x						x			x				x	x			x	x	
ASAtS	Switzerland	Panel	2011-2012	22		x	x						x			x				x	x	x	x	x	x				x		
TREE	Switzerland	Multicohort	2000-present	15-(now 36)		x	x	x		x	x	x			x			x		x	x	x	x	x	x	x	x		x	x	x
MCS	UK	Birth cohort	2001-present	9 months - (now 21)		x	x	x		x					x			x		x			x	x	x	x	x	x	x	x	x
ECHP	Cross-national	Household panel	1994-2001	16-further	x		x			x															x	x	x		x	x	x
TIMSS	Cross-national	Quasi-longitudinal	1997-present	9-14							x		x		x		x	x		x				x	x					x	

* The acronyms of the datasets were used. The age range was expressed in years unless mentioned otherwise. Variables that were identified within the dataset were marked with an 'x'.

5.2 General comparability between datasets

Table 4 displays the comparability between datasets based on the time period of the data collection or the age of the respondents. When no overlap based on time period or age could be detected, the cells were marked in black. For example, the LiSO and SiBO dataset cannot be compared although they have some overlap in age but no overlap regarding time period as SiBO ended in 2011 and LiSO started in 2013. When a study ended in 2011 and another started in 2011, the study was considered an overlap. The same principle was applied to the ages. For example, there was an overlap based on age between the ECHP dataset and the NABC dataset as the ECHP starts from the age of 16 of respondents and the NABC dataset includes the ages between 11 and 16. However, we did not take into account which age was reached at a particular time to match the datasets based on time period or age. We wanted to leave the decision to the researchers whether they want to include different ages at different time periods in their research. Table 4 was used as a foundation for the other categories we used to compare the datasets as the pool of datasets in the inventory for harmonisation were filtered based on time period or age due to time heterogeneity.

Table 4. Overlap between datasets based on time period or age

	Time period or age																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO		■												■			■	
LiSO												■		■			■	
DAR																		
DNT										■				■	■		■	
ELFE										■		■	■	■	■		■	
FiD																	■	
LISA														■			■	
NEPS																	■	
SOEP																		
TOSCA																		■
Admin																	■	
HLCS														■			■	
NABC														■			■	
ASAtS																	■	■
TREE																		■
MCS																		
ECHP																		■
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period or age. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

5.2.1 Educational variables

5.2.1.1 Formal characteristics

Overlap between datasets based on educational variables was detected after applying a filter on time period and age. The tables displaying those opportunities for harmonisation are displayed in Appendix 5. Subsection a5.1.1 in Appendix 5 shows the harmonisation possibilities for the datasets regarding formal characteristics. There was a considerable amount of harmonisation opportunities for formal characteristics such as choice of study, early school-leaving (including drop-out), and years of formal education. We included the category adult education (i.e., education after entering the labour market) in our linking process because not many studies included this category in their dataset. Although both NEPS and ECHP have variable(s) on adult education, they were not marked as a match because they did not overlap based on time period or age (see Table a2).

As seen in Table 3, six datasets did not contain information on the degrees: SiBO, DNT, LISA, TOSCA, NABC, and TIMSS. When investigating the harmonisation possibilities regarding the degree variable, a few options were eliminated. Table a4 displays the overlap between datasets based on variables on degrees. Although ELFE and TREE contained a degree variable, no overlap between those dataset could be detected as they did not overlap on age. However, in a few years those datasets could be a subject of comparison as ELFE participants were 10 years old at the time of the study but will be 20 years old when the study ends in 2031 (see Subsection 3.3.1).

Information regarding grade retention was detected within five datasets, namely the SiBO, LiSO, FiD, SOEP, HLCS datasets. However, no overlap between the SiBO and LiSO datasets could be detected based on time period or age. Hence, these datasets were not indicated as comparable in Table a6. Furthermore, the FiD variables were part of the SOEP dataset (see Subsection 3.4.1). These remarks should be considered when deciding which datasets to match regarding grade retention.

5.2.1.2 Academic outcomes

In addition to formal characteristics, we considered academic outcomes (such as test scores) as a subcategory. The harmonisation possibilities for those variables can be found in Section a5.1.2 of Appendix 5. Only three datasets have no variable regarding mathematics test scores: ASAtS, MCS, and ECHP. When considering test scores on language subjects (such as reading) the TIMSS dataset can be added to the datasets that have no variable regarding test scores. A broad subcategory regarding test scores on other subject was included in our matching process. For example, the DNT dataset included test scores on geography and biology (Beuchert & Nandrup, 2017). Lastly, the later educational attainment of students was considered. A single match between the DNT and SiBO datasets was found (see Table a11). Researchers should, however, consider whether the later educational outcomes of those datasets are similar enough in order to consider those datasets for harmonisation.

5.2.1.3 Non-academic characteristics

Non-academic characteristics included cognitive development (e.g., intelligence test, thinking abilities, and teacher evaluation of students' relative performance level), learning-styles, relationship between parents and the school or teachers, school context (e.g. size of classes), school policy, students' attitude (e.g., attitude towards their teachers or studying), students' motivation, students' self-efficacy, students' well-being (e.g., mental health and satisfaction), and teacher characteristics (e.g., teachers' experiences). As shown in Table 3, ELFE, FiD, SOEP, and ASAtS contain information on the students' learning style. This broad category contains items such as teachers indicating the learning process of the students or the autonomy in the learning process. This last item, encompassed in the ASAtS dataset, was also seen as an item that measured self-efficacy. Table a13 displays the overlap between datasets based on learning styles and Table a19 shows the overlap between datasets based on the students' self-efficacy.

Note that the SiBO, LiSO, LISA and TOSCA datasets contained variables regarding school policy, such as how lessons are conducted (see Table 3). However, the SiBO and LiSO datasets did not overlap based on time or age (see Table 4) and were, thus, not marked as an overlap (see Table a16).

5.2.2 Additional variables

Apart from educational variables, additional variables were considered. Those included health (e.g., eating habits, medical factors, and weight at birth), social development (e.g., socialisation, behaviour, and personality), socio-economic background (such as SES), and socio-economic outcomes (such as labour market outcomes). Variables on the environment of students include supportive home environment or learning environment. The ELFE and ECHP dataset also contained information on air pollution. Because those two datasets did not overlap based on time period or age, they were not marked as a match (see Table a22). As seen in Table a23, there was a considerable amount of datasets that contained information on family characteristics. However, some datasets contained more elaborate information on those family characteristics such as the FiD dataset which was created for this purpose. The ELFE and MCS datasets included pregnancy related variables such as the mothers' health or a maternity medical record. Because they also overlapped based on time period or age, they were indicated as a match for harmonisation opportunities in Table a25.

6. Discussion

In this final chapter, we discuss how this report laid the foundations and presented the tools to harmonise longitudinal datasets on educational careers across Europe. We recapitulate the different steps to harmonise datasets and link these to the chapters in this report. After this, the limitations of this report are presented. Lastly, we provide suggestions to improve future comparative research regarding longitudinal datasets on educational careers.

6.1 Conclusion

The objective of this report was to map the different longitudinal datasets on educational careers that exist across Europe and explore the possibilities of harmonising those datasets. Fortier et al. (2017) discussed the steps a researcher needs to take to harmonise datasets. This report provides the foundations for those steps. However, not all steps could be prepared within the scope of this research. For example, the researchers themselves need to define their approach and determine the research questions (step 0). To collect the information on the dataset (step 1), an overview of the different longitudinal datasets on educational careers in Europe was presented in Chapter 3. Understanding the data is important to reduce measurement bias (Wong et al., 2003). This implies understanding of the institutional context of the countries, such as the different educational institutions, which is necessary to make sense of the datasets and needs to be further explored by the researchers.

Fortier et al. (2017) identified how to check the harmonisation possibilities and how to evaluate the harmonisation (step 2) as a next step in the harmonisation process. This report discussed multiple harmonisation opportunities in Chapter 5. We categorised the variables in two broad categories: educational and additional variables (e.g., social development or health). The educational variables had three subcategories: formal characteristics (e.g., degree), academic outcomes (e.g., mathematics test scores), and non-academic characteristics (e.g., students' motivation). The broad categorisation of the variables resulted in a summary that visually indicated the harmonisation possibilities and can be seen in Appendix 5. We considered datasets not comparable when no overlap could be detected based on time period *or* age. We did not consider datasets as a match when overlap between datasets could be detected based on a variable but they did not overlap based on time or age. This was the case for NEPS and ECHP, which both had variable(s) on adult education but did not overlap based on time period or age. Furthermore, some datasets could gain a harmonisation possibility in a few years. While the ELFE and TREE dataset did not overlap based on age at the time of this report, they could overlap later on as ELFE is an ongoing study until the children reach the age of 20 in 2031 and TREE is also in process and, at this moment, covers children from the age of 15 to 36. Although we displayed many options for harmonisation, the harmonisation possibilities need to be further explored by the researcher who has interest in data harmonisation.

In Chapter 4, the methodologies for processing data and correcting for heterogeneity (step 3 and 4) were examined. We discussed comparative research methods such as data harmonisation and the integration process. Throughout the harmonisation and integration process, it is important that the researcher gives sufficient attention to the comparability of the different datasets. As mentioned in Curran and Hussong (2009), integrating data requires attention to heterogeneity. Heterogeneity can be caused by differences in sampling (populations for which the samples are drawn need to be similar and data should be compared on the same level), the time frame of the study (researchers should aim to integrate datasets for which the data collection took place in a similar time period), the study design (repercussions of the study can have an impact on the data), and measurements of key concepts (due to the usage of a less precise proxy of the real outcome of interest). Once the datasets are harmonised and integrated, the researcher can perform analyses on the newly generated dataset. In order to take the heterogeneity between different studies into account, it is advised to use statistical techniques that consider the hierarchical structure of the data. Furthermore, as step 3 requires access to study-specific data, we listed the ways one could ensure access to the data items in Chapter 3. Most datasets require an application before the researcher can have access to the dataset and information on the dataset. These procedures make investigating the harmonisation possibilities harder (see Sections 6.2 and 6.3).

6.2 Limitations

Two main limitations can be identified within this report: (i) unavailability of information on operationalisation of variables, and (ii) a restricted inventory. Firstly, we did not have precise information on the availability and the operationalisation of the variables. That information would be necessary to reduce measurement bias (Wong et al., 2003). The absence of this information has two consequences for this study: no specific matches and missing matches. Because we had to use broadly defined categories to match the datasets, we could not formulate specific matches between datasets. Exploring the matches in more detail will reveal that not every identified match is an ideal match answering to the specific harmonisation goals of researchers. While the ideal match ultimately depends on the research goals of the researchers who want to harmonise, there was also not enough information on the variables to distinguish the categories in greater detail.

Furthermore, it is plausible that some potential matches could not be detected during the exploration of harmonisation possibilities. Not every dataset shared a list of variables which caused the possibility that some matches may be overlooked.

The second limitation of this report is the restricted inventory as it did not deliver a full overview of the available longitudinal datasets on educational careers. Datasets that have no international publications were not detected due to the methodology used to identify the datasets. Additionally, the datasets for which we did not receive an answer on the expert questionnaire or could not find sufficient information on online could not be included in this report. Furthermore, although many other interesting longitudinal datasets on educational careers exist, only European (including the Schengen Area) datasets were included in this study.

6.3 Suggestions

This report aimed to improve European comparative research regarding longitudinal datasets on educational careers. Mapping and matching the different longitudinal datasets within this report revealed opportunities for improvement to enhance comparative research in education. Firstly, provide a detailed list of variables without the necessity to request the data first. This will help researchers decide whether data harmonisation with the dataset is feasible. Secondly, technical reports with details on topics such as data collection, attrition, or sample design need to be made available to account for heterogeneity (see Section 4.2). Thirdly, the study's website should clearly refer to the list of variables, technical reports, and other documentations. Making those documentations easily accessible will reduce a barrier for data harmonisation and reduce the time spent on the information collection (identified as step one by Fortier et al., 2017). Lastly, researchers should offer English translations of the documentation. This includes the study's website, technical reports, questionnaires, and list of variables. Those four recommendations are related to the fifth step identified by Fortier et al. (2017): share information. Taking those recommendations in mind will help improve prospective harmonisation and, thus, European comparative research on educational careers.

One way to minimise biases (see Section 4.2), such as measurement bias, is by organising cross-country longitudinal analyses. A promising initiative that aims to start in 2022 is the 'Growing Up in Digital Europe (GUIDE), EuroCohort' project funded by the European Union's Horizon 2020 programme. EuroCohort (n.d.) is Europe's first comparative birth cohort that aims to follow a sample of newborns and school age children until they reach the age of 24 years. The cohort members will be asked to complete questionnaires that aims to probe into topics such as child education and development. More information can be found on the project's website.[46] In anticipation of the start of the EuroCohort in 2022, European comparative research on educational careers can be improved by harmonisation of the existing longitudinal datasets in Europe.

46 https://eurocohort.eu/

appendix 1 Expert questionnaire

a1.1 Intro

InGRID is a research infrastructure project funded by the European Commission's Horizon 2020 programme. InGRID aims to support the social science community by promoting (among other things):
- data archives/collections that have already been specialising in integrating national data;
- existing EU-wide databases and indicator collections of relevant national institutions and policies;
- transnational access of researchers to each other's datasets;
- harmonisation/integration of datasets for (joint) transnational comparative research.

This survey aims to collect information about the existing longitudinal datasets regarding educational careers in Europe. By collecting such information, we aim to identify and document the most interesting longitudinal datasets on education and explore the possibilities of sharing or merging them, in order to enable comparative longitudinal research in education. This survey is administered in connection with the InGRID Expert workshop 'Comparative analysis of longitudinal data on educational outcomes' (27-29 November 2019, SOEP & DIW, Berlin).

The following questionnaire should not take more than 30-45 minutes to complete.

Please fill one questionnaire per dataset. You can submit this form more than once if you can provide information about more than one such dataset.

1. What is the name of the dataset that you developed or used in your research?
   ……………………………………………………………………………………………
2. What is the acronym (if applicable)?
   ……………………………………………………………………………………………

a1.2 Design of the dataset

3. What kind of longitudinal dataset is it? **Please describe it.** (e.g. panel, pseudo-panel, cohort follow-up, administrative data, census, combination of data from various surveys and/or administrative data, etc.)
   ……………………………………………………………………………………………
4. In which country (countries) or region(s) are the data collected?
   ……………………………………………………………………………………………
5. When did the data collection start?
   ……………………………………………………………………………………………
6. When did/will the data collection end?
   ……………………………………………………………………………………………

7. (So far) How many data collection rounds (waves) were there, and at what frequency (e.g. yearly/two-yearly, …)?
…………………………………………………………………………………………………

8. Please describe the sampling design (1st wave)
…………………………………………………………………………………………………

9. What is/was the sample size?
…………………………………………………………………………………………………

10. How and by whom is/was the data collection funded?
…………………………………………………………………………………………………

a1.3 Content of the dataset

11. What is this dataset about? (e.g. educational careers, educational outcomes, orientation, performance, tests, educational career data, individual background data, psychological variables, parental background, teacher characteristics, school characteristics, other environmental characteristics, other, etc.)
…………………………………………………………………………………………………

12. Who are the respondents? (e.g. students/pupils, parents, teachers, school principals, other, etc.)
…………………………………………………………………………………………………

13. What level of education, age group or grade level is/was covered?
…………………………………………………………………………………………………

14. Is a technical report available about the dataset? YES/NO
O in English? YES/NO
O in other language(s) – please specify:

15. Could you provide the URL for further information and/or attach the technical report to this filled questionnaire?
…………………………………………………………………………………………………

16. What is/was the language(s) of questionnaires and documentation?
…………………………………………………………………………………………………

17. Have any (key) research results based on these data been published? If so, please provide the references/URL:
…………………………………………………………………………………………………

18. Are the micro-data freely available? YES/NO
If they are available, could you provide the URL to the freely available dataset?
…………………………………………………………………………………………………

19. Is access to the micro-data (by international scholars) *legally* prohibited? YES/NO

If not,

20. What would be the conditions for transnational access to the micro-data?
    a. Financial (please specify: ………………………………………)
    b. Intellectual property rights (please specify: …………………….)
    c. Administrative/legal (please specify: …………………………)

21. What is the format of the database file? (CSV, etc.)

22. In the case of a 'national' longitudinal dataset, did you ever (try to) combine it with datasets from other countries/regions for comparative research and create a pooled dataset? YES/NO

If yes,

23. What were the main obstacles/difficulties?
    ……………………………………………………………………………………………

24. Do you consider that the output of the comparative research was successful?
    ……………………………………………………………………………………………

FINALLY, if we need more information about this survey, whom can we contact? (provide your contact details if applicable)
Name:
Institution:
Email:
Tel:

Thank you very much!

appendix 2 Relationship between grade and age

Table a1. Level of education: the relationship between grade and age

Grade	Age (in years)
Grade 1	6-7
Grade 2	7-8
Grade 3	8-9
Grade 4	9-10
Grade 5	10-11
Grade 6	11-12
Grade 7	12-13
Grade 8	13-14
Grade 9	14-15
Grade 10	15-16
Grade 11	16-17
Grade 12	17-18
Grade 13	18-19

* Grade was used as an indicator for students' level of education because the majority of datasets used this indicator. When another indicator was used in the dataset the level was adepted to grades, using this Table a1 as a guideline.

appendix 3 Fisher's model-based inferential method

Fisher's (1922) model-based inferential method is an inferential framework that allows the researcher to make inferences to the whole population under study, even though this is normally not possible with non-random sampling. His framework consists of four steps.[47] To start with, the researcher must formulate the statistical model that describes how the dependent variables have been generated (Sterba, 2009). Secondly, the researcher needs to impose a parametric distributional assumption on the model 'in order to convert the fixed y-values obtained for the sampled units into realisations of a random variable y' (Fisher, 1922, p. 313). If the researcher, for example, sets that the errors in our regression model are independently and identically distributed with mean 0 and variance σ2, they would assume that the error term is a random variable. In that case the dependent variable y becomes a random variable (Neter et al., 1996). By imposing this model the y-values are thus epistemically random (Johnstone, 1989), whether the researcher used random sampling or not.

The third step if often called 'Fisher's conditionality principle' (see Johnstone, 1987; Lehmann, 1993). Fisher (1922) described several circumstances under which the sampling mechanism would differ significantly from the random sampling. These circumstances have to be made explicit, and taken into account. The first one is stratification. The sampling units might be divided into non-overlapping categories (e.g., youth and elderly) before being independently selected from each stratum. If the researcher ignores this issue, the standard errors would become too large (Kish & Frankel, 1974). The researcher must thus condition their model on any strata indicators so that, after conditioning, the infinite population is 'subjectively homogeneous and without recognisable stratification' (Fisher, 1956, p. 33; Sterba, 2009). This can be done by introducing strata dummies as fixed effects. The sampling units could also be clustered. In contrast to the stratification, clustering results in standard errors that are too small (Kish & Frankel, 1974). In that case the researcher has to include cluster indicators as random effects (Raudenbush & Bryk, 2002). As a last issue Fisher pointed at the fact that the probability for sampling units to get selected might be disproportionate,[48] resulting in probabilities of selection that are related to the outcome variable even after controlling for independent variables. The exact consequence of this disproportional sampling depends on how the selection variables relate to the outcome after conditioning on independent variables (e.g., Berk, 1983; Graubard & Korn, 1996; Skinner et al., 1989; Sugden & Smith, 1984). If these three steps are taken into account, the researcher can draw inferences from nonrandom samples to infinite populations (Sterba, 2009).

47 A more detailed description of the steps can be found in Sterba (2009): https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2856970/

48 But not stratified or clustered.

appendix 4 Statistical methods to tackle measurement bias

Several methods were designed to make sure that researchers are measuring the same theoretical construct similarly for all individuals across the datasets. As a first step the researcher must identify the set of items that measure the construct of interest. but they all start with identifying the set of items that measure the construct the researcher is interested in. In the ideal situation some portion of this pooled items set contains items that are used in every study.[49] Curran and Hussong (2009) refer to these shared items as ‘anchor items’. It is not necessary that the items are worded in the precise same manner. It is sufficient if they are operating in a psychometrically similar fashion. This can be tested using statistical techniques like calculating and comparing the Cronbach’s alpha’s to check whether specific items form a reliable scale for each separate system (Esser & Vliegenthart, 2017) or exploratory factor analyses, when more dimensions are anticipated to be present.[50] However, when applying a comparative content analysis, one should gather information from the exploration of the concepts, the theoretical frameworks, expert advice and additional sources to provide a qualitative discussion on the equivalence of the items used (W. Wirth & Kolb, 2012).

As a second step we need to *select the appropriate measurement model*. This depends on the type of response scales that were used for each individual item. A Standard Linear CFA is suitable for items that are interval scaled, while nonlinear measurement models should be used when dealing with discretely scaled items since the assumption of linearity is not met (Flora & Curran, 2004). We can choose between two options, nonlinear factor analysis (NLFA; Bauer & Hussong, 2009; Curran et al., 2007, 2008; Skrondal & Rabe-Hesketh, 2004; R. J. Wirth & Edwards, 2007) or item response theory (IRT; Thissen & Wainer, 2001). We can however also evade the assumed continuous distribution by combining two or more individual items to create item parcels (Little et al., 2002; MacCallum et al., 1999).

Once the appropriate measurement model is chosen, this *model is fitted*. The specific model fitting depends on the characteristics and goals of the given integrative data analysis (IDA). Nevertheless, Curran et al. (2008) identified four general steps in the measurement portion of the analysis: (i) assessment of the dimensionality underlying the set of items; (ii) fitting measurement models within each study separately and, subsequently, across all studies simultaneously to get an understanding of the psychometric properties of the scales; (iii) assessment of measurement invariance across study, demographic group or time; (iv) calculation of the scale scores that are obtained by combining the observed pattern of responses to the items and the parameter estimates from the final measurement model. For this step one of the several available factor score estimates in the factor model (Grice, 2001) and a posterior model estimate or posterior mean estimate scoring in the IRT model (Thissen & Wainer, 2001) are used. After this step a person-specific scale score that contains information about study group and, potentially, demographic group membership is created. These scale scores serve as input for the subsequent statistical analyses. The most suitable method for this statistical analysis depends on the specific characteristics of the datasets to be pooled (Bauer & Hussong, 2009; McArdle et al., 2009).

49 In this regard previous Danish national test-scores are adjusted when new items to measure the same construct are used. By doing this, the researchers ensure comparability across cohorts (Pøhler & Sørensen, 2010).

50 More advanced techniques for testing and optimising equivalence (e.g., Latent Class Analysis (LCA), multigroup confirmatory factor analysis or congruence coefficient analysis) can be found in Davidov et al. (2014), Kühne (2018), and Wirth and Kolb (2012).

appendix 5 Harmonisation possibilities

a5.1 Educational variables

a5.1.1 Formal characteristics

Table a2. Overlap between datasets based on formal characteristics: adult education

	Adult education																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and adult education. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a3. Overlap between datasets based on formal characteristics: choice of study

	Choice of study																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and choice of study. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a4. Overlap between datasets based on formal characteristics: degree

	Degree																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and degree. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a5. Overlap between datasets based on formal characteristics: early school-leaving

	Early school-leaving																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and early school-leaving. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a6. Overlap between datasets based on formal characteristics: grade retention

	Grade retention																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and grade retention. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

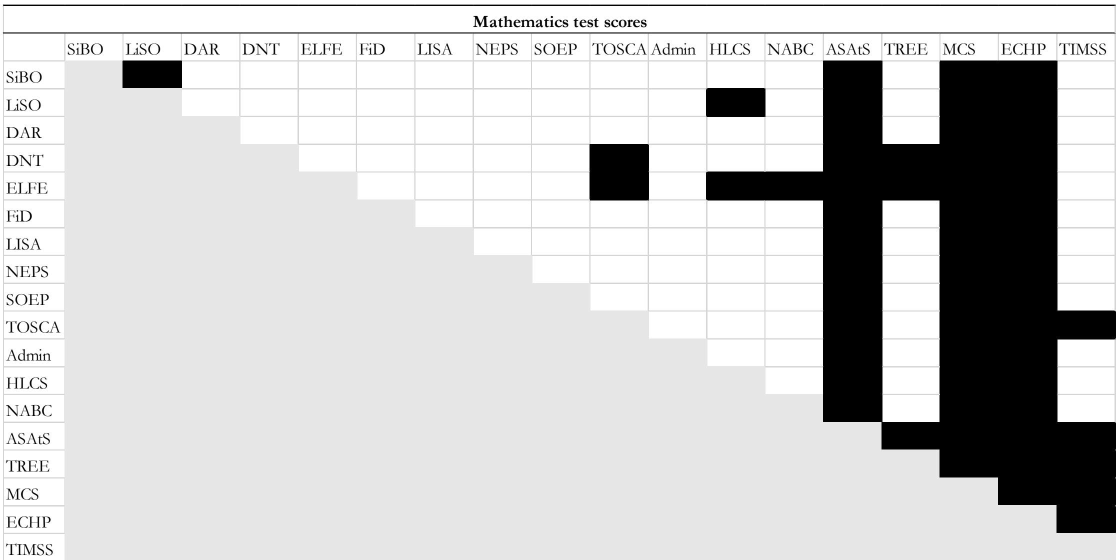
Table a7. Overlap between datasets based on formal characteristics: years in formal education

	Years in formal education																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and grade retention. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

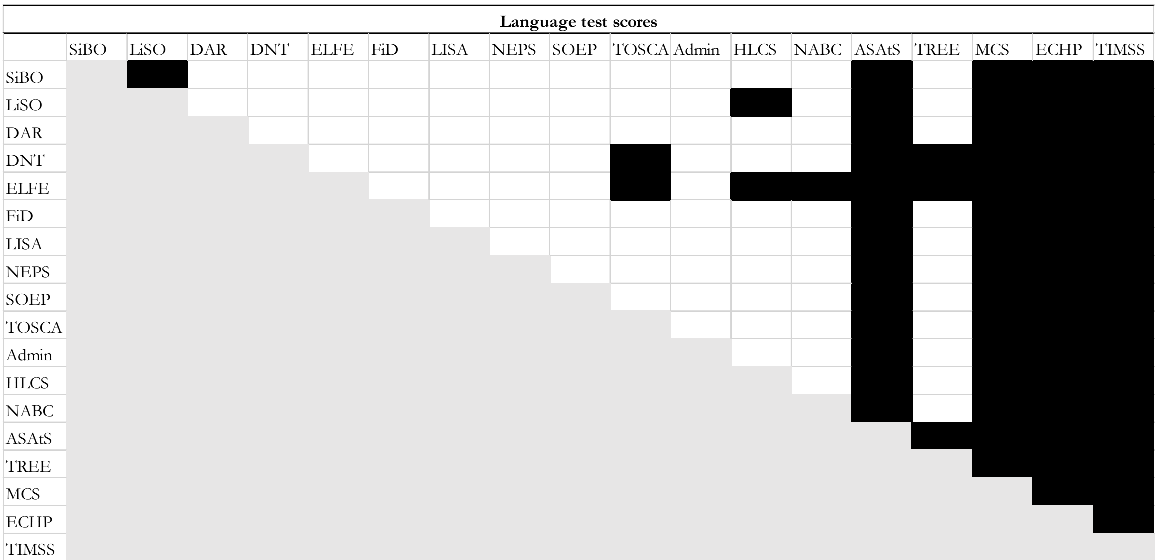
a5.1.2 Academic outcomes

Table a8. Overlap between datasets based on academic outcomes: mathematics test scores



* The cells indicate whether overlap could be detected between the dataset based on time period, age, and mathematics test scores. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a9. Overlap between datasets based on academic outcomes: language test scores



* The cells indicate whether overlap could be detected between the dataset based on time period, age, and language test scores. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a10. Overlap between datasets based on academic outcomes: other test scores

	Other test scores																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and other test scores. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a11. Overlap between datasets based on academic outcomes: later educational outcomes

	Later educational outcomes																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and later educational outcomes. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

a5.1.3 Non-academic outcomes

Table a12. Overlap between datasets based on non-academic characteristics: cognitive development

	Cognitive development																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and cognitive development. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a13. Overlap between datasets based on non-academic characteristics: learning styles

	Learning styles																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and learning styles. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a14. Overlap between datasets based on non-academic characteristics: relationship between parents and school or teachers

Relationship between parents and school or teachers																		
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and the relationship between parents and school or teachers. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a15. Overlap between datasets based on non-academic characteristics: school context

School context																		
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and school context. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a16. Overlap between datasets based on non-academic characteristics: school policy

	School policy																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and school policy. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a17. Overlap between datasets based on non-academic characteristics: students' attitude

	Students' attitude																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and students' attitude. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a18. Overlap between datasets based on non-academic characteristics: students' motivation

Students' motivation																		
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and students' motivation. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a19. Overlap between datasets based on non-academic characteristics: students' self-efficacy

Students' self-efficacy																		
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and students' self-efficacy. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a20. Overlap between datasets based on non-academic characteristics: students' well-being

	Students' well-being																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and students' well-being. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

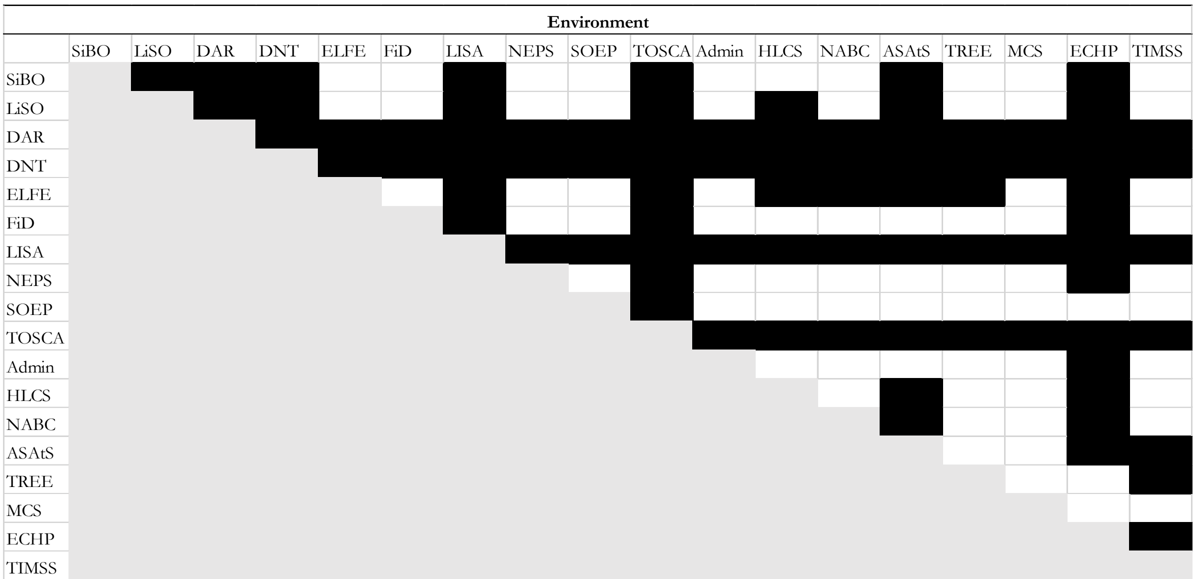
Table a21. Overlap between datasets based on non-academic characteristics: teacher characteristics

	Teacher characteristics																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and teacher characteristics. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

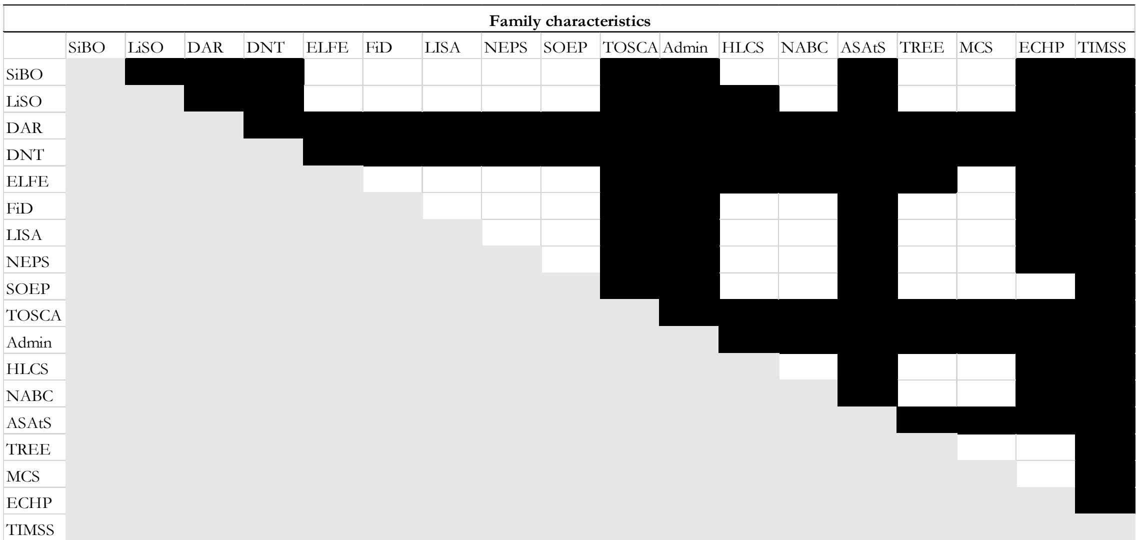
a5.2 Additional variables

Table a22. Overlap between datasets based on environment



* The cells indicate whether overlap could be detected between the dataset based on time period, age, and environment. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a23. Overlap between datasets based on family characteristics



* The cells indicate whether overlap could be detected between the dataset based on time period, age, and family characteristics. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a24. Overlap between datasets based on health

	Health																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and health. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a25. Overlap between datasets based on pregnancy

	Pregnancy																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and pregnancy. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a26. Overlap between datasets based on social development

	Social development																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and social development. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a27. Overlap between datasets based on socio-economic background

	Socio-economic background																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and socio-economic background. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

Table a28. Overlap between datasets based on socio-economic outcomes

	Socio-economic outcomes																	
	SiBO	LiSO	DAR	DNT	ELFE	FiD	LISA	NEPS	SOEP	TOSCA	Admin	HLCS	NABC	ASAtS	TREE	MCS	ECHP	TIMSS
SiBO																		
LiSO																		
DAR																		
DNT																		
ELFE																		
FiD																		
LISA																		
NEPS																		
SOEP																		
TOSCA																		
Admin																		
HLCS																		
NABC																		
ASAtS																		
TREE																		
MCS																		
ECHP																		
TIMSS																		

* The cells indicate whether overlap could be detected between the dataset based on time period, age, and socio-economic outcomes. When no overlap could be detected, the cells were marked in black. Consequently, the white cells mark that there was an overlap between the datasets based on time period and age. The gray cells on the diagonal signal the same dataset (e.g., there is no point in comparing the SiBO-dataset with the SiBO-dataset). Below the diagonal grey cells were used to indicate that the comparability had already been visualised.

References

Andersen, S. C., Christensen, M. V., Nielsen, H. S., Thomsen, M. K., Østerbye, T., & Rowe, M. L. (2018). How reading and writing support each other across a school year in primary school children. *Contemporary Educational Psychology*, *55*, 129–138. https://doi.org/10.1016/j.cedpsych.2018.09.005

Anumendem, N. D., De Fraine, B., Onghena, P., & Van Damme, J. (2017). Growth in reading comprehension and mathematics achievement in primary school: A bivariate transition multilevel growth curve model approach. *Biometrics & Biostatistics International Journal*, *5*(4), 00137. https://doi.org/10.15406/bbij.2017.05.00137

Balázsi, I. (2006, May 23). National Assessment of Basic Competencies in Hungary. *Assessment in an Era of Rapid Change: Innovations and Best Practices*. 32nd annual conference IAEA: Assessment in an Era of Rapid Change: Innovations and Best Practices, Singapore, Singapore. https://iaea.info/documents/national-assessment-of-basic-competencies-in-hungary/

Bassanini, A. (2006). Training, wages and employment security: An empirical analysis on European data. *Applied Economics Letters*, *13*(8), 523–527. https://doi.org/10.1080/13504850500400447

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods*, *14*(2), 101–125. https://doi.org/10.1037/a0015583

Beck, N., & Katz, J. N. (1995). What to do (and not to do) with time-series cross-section data. *The American Political Science Review*, *89*(3), 634–647. https://doi.org/10.2307/2082979

Behr, A., Giese, M., Kamdjou, H. D. T., & Theune, K. (2021). Motives for dropping out from higher education—An analysis of bachelor's degree students in Germany. *European Journal of Education*, *00*, 1–18. https://doi.org/10.1111/ejed.12433

Berger, L. M., Panico, L., & Solaz, A. (2021). The Impact of Center-Based Childcare Attendance on Early Child Development: Evidence From the French Elfe Cohort. *Demography*, *58*(2), 419–450. https://doi.org/10.1215/00703370-8977274

Berk, R. A. (1983). An introduction to sample selection bias in sociological data. *American Sociological Review*, *48*(3), 386–398. https://doi.org/10.2307/2095230

Beuchert, L., & Nandrup, A. (2017). The Danish National Tests at a Glance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2996431

Brahm, T., & Jenert, T. (2015). On the assessment of attitudes towards studying—Development and validation of a questionnaire. *Learning and Individual Differences*, *43*, 233–242. https://doi.org/10.1016/j.lindif.2015.08.019

Brahm, T., Jenert, T., & Wagner, D. (2017). The crucial first year: A longitudinal study of students' motivational development at a Swiss Business School. *Higher Education*, *73*(3), 459–478. https://doi.org/10.1007/s10734-016-0095-8

Britzke, J., & Schupp, J. (Eds.). (2017). *SOEP wave report 2016* (No. 7; SOEP Annual Report, p. 222). DIW Berlin, The German Socio-Economic Panel (SOEP). https://www.diw.de/documents/publikationen/73/diw_01.c.560446.de/wave_report_2016.pdf

Britzke, J., & Schupp, J. (Eds.). (2018). *SOEP wave report 2017* (No. 8; SOEP Annual Report, p. 196). DIW Berlin, The German Socio-Economic Panel (SOEP). https://www.diw.de/documents/publikationen/73/diw_01.c.592826.de/wave_report_2017.pdf

Britzke, J., & Schupp, J. (Eds.). (2019). *SOEP wave report 2018* (No. 9; SOEP Annual Report, p. 180). DIW Berlin, The German Socio-Economic Panel (SOEP). https://www.diw.de/documents/publikationen/73/diw_01.c.632608.de/wave_report_2018.pdf

Brüggemann, M., Engesser, S., Büchel, F., Humprecht, E., & Castro, L. (2014). Hallin and Mancini revisited: Four empirical types of Western media systems. *Journal of Communication*, *64*(6), 1037–1065. https://doi.org/10.1111/jcom.12127

Burston, K., Conolly, A., Cleary, A., Dangerfield, P., Gardiner, C., Knibbs, S., Michelmore, O., Rose, N., Sheriffs, P., & Tietz, S. (2017). *Millennium Cohort Study sixth sweep (MCS6): Technical report* (p. 134) [Technical report].

Centre for Longitudinal Studies, UCL Institute of Education. http://doc.ukdataservice.ac.uk/doc/8156/mrdoc/pdf/mcs6_technical_report.pdf

Centre for Longitudinal Studies. (2019). *Millennium Cohort Study seventh sweep (MCS7): Technical report* (pp. 1–141) [Technical report]. Centre for Longitudinal Studies, UCL Institute of Education. https://cls.ucl.ac.uk/wp-content/uploads/2020/01/MCS7_Technical_Report.pdf

Centre for Longitudinal Studies. (2021). *Centre for Longitudinal Studies*. Millennium Cohort Study. https://cls.ucl.ac.uk/cls-studies/millennium-cohort-study/

CLOSER. (2021a). *Data harmonisation – the basics*. CLOSER Learning Hub. https://learning.closer.ac.uk/learning-modules/data-harmonisation/data-harmonisation-the-basics/download-this-section/

CLOSER. (2021b). *Learning Hub*. CLOSER. https://learning.closer.ac.uk/

Curran, P. J., Edwards, M. C., Wirth, R. J., Hussong, A. M., & Chassin, L. (2007). The incorporation of categorical measurement models in the analysis of individual growth. In *Modeling contextual effects in longitudinal studies* (pp. 89–120). Lawrence Erlbaum Associates Publishers.

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*(2), 81–100. https://doi.org/10.1037/a0015914

Curran, P. J., Hussong, A. M., Cai, L., Huang, W., Chassin, L., Sher, K. J., & Zucker, R. A. (2008). Pooling data from multiple longitudinal studies: The role of item response theory in integrative data analysis. *Developmental Psychology*, *44*(2), 365–380. https://doi.org/10.1037/0012-1649.44.2.365

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*(1), 55–75. https://doi.org/10.1146/annurev-soc-071913-043137

DeVries, J. M., Szardenings, C., Doebler, P., & Gebhardt, M. (2021). Subject-Specific Self-Concept and Global Self-Esteem Mediate Risk Factors for Lower Competency in Mathematics and Reading. *Social Sciences*, *10*(1), 1–17. https://doi.org/10.3390/socsci10010011

DIW Berlin. (2021). *DIW Berlin* [Text]. Integrated Studies; DIW Berlin. https://www.diw.de/en/diw_01.c.603160.en/integrated_studies.html

Eberhard Karls Universität Tübingen. (2021). *TOSCA | University of Tübingen*. Transformation of the Secondary School System and Academic Careers (TOSCA). https://uni-tuebingen.de/en/faculties/faculty-of-economics-and-social-sciences/subjects/department-of-social-sciences/education-sciences-and-psychology/research/aktuelle-studien/tosca/

Esser, F., & Vliegenthart, R. (2017). Comparative research methods. In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), *The international encyclopedia of communication research methods*. John Wiley & Sons, Inc.

Eurocohort. (n.d.). *EuroCohort: An accelerated cohort survey of growing up in Europe* (p. 8) [Briefing]. European Union. https://www.eurocohort.eu/files/filemanager/files/EuroCohort_an_accelerated_cohort_survey_of_growing.pdf

European Communities. (2003). *The EC Household Panel 'Newsletter' (01/02)* (p. 11). Office for Official Publications of the European Communities.

Eurostat. (n.d.). *Eurostat*. European Community Household Panel (ECHP). Retrieved 20 May 2021, from https://ec.europa.eu/eurostat/web/microdata/european-community-household-panel

Eurostat. (2003). *ECHP UDB: Description of variables. Data dictionnary, codebook and differences between countries and waves* (ECHP UDB DOC. PAN 166/2003-12; p. 500). European Commission, Eurostat. https://ec.europa.eu/eurostat/documents/203647/203704/ECHP_variables/477f8adb-531e-4c7d-c071-de34ddc55765

Fischer, J.-P., & Thierry, X. (2021). Are differences between social classes reduced by non-symbolic numerical tasks? Evidence from the ELFE cohort. *British Journal of Educational Psychology*, *91*(1), 286–299. https://doi.org/10.1111/bjep.12363

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A*, *222*(594–604), 309–368. https://doi.org/10.1098/rsta.1922.0009

Fisher, R. A. (1956). *Statistical methods and scientific inference* (pp. viii, 175). Hafner Publishing Co.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466

Fortier, I., Raina, P., Van den Heuvel, E. R., Griffith, L. E., Craig, C., Saliba, M., Doiron, D., Stolk, R. P., Knoppers, B. M., Ferretti, V., Granda, P., & Burton, P. (2017). Maelstrom Research guidelines for rigorous retrospective data harmonisation. *International Journal of Epidemiology*, *46*(1), 103–105. https://doi.org/10.1093/ije/dyw075

Fräßdorf, M., Siegers, R., Damerow, S., Mannschreck, M., Putzke, G., Raith, A., Scherner, N., Werneburg, J., Wittbrodt, L., & Zobel, M. (2016). *SOEP FiD - 'Familien in Deutschland', data documentation release FiDv4.0* (Research Report No. 341; SOEP Survey Papers 341: Series D). Berlin: DIW/SOEP. https://www.econstor.eu/handle/10419/146886

Gadeyne, E., Ghesquière, P., & Onghena, P. (2006). Psychosocial educational effectiveness criteria and their relation to teaching in primary education. *School Effectiveness and School Improvement*, *17*(1), 63–85. https://doi.org/10.1080/09243450500391573

Gallop, K., Rose, N., Wallace, E., Williams, R., Cleary, A., Thompson, A., Burston, K., Frere-Smith, T., Dangerfield, P., & Tietz, S. (2013). *Millennium Cohort Study fifth sweep (MCS5): Technical report* (pp. 1–122) [Technical report]. Centre for Longitudinal Studies, UCL Institute of Education. http://doc.ukdataservice.ac.uk/doc/7464/mrdoc/pdf/mcs5_survey_technical_report.pdf

Gambaro, L., Kemptner, D., Pagel, L., Schmitz, L., & Spieß, C. K. (2020). Integration of refugee children and adolescents in and out of school: Evidence of success but still room for improvement. *DIW Weekly Report*, *34*, 345–354. https://doi.org/10.18723/DIW_DWR:2020-34-2

Gaye, A., Marcon, Y., Isaeva, J., LaFlamme, P., Turner, A., Jones, E. M., Minion, J., Boyd, A. W., Newby, C. J., Nuotio, M.-L., Wilson, R., Butters, O., Murtagh, B., Demir, I., Doiron, D., Giepmans, L., Wallace, S. E., Budin-Ljøsne, I., Oliver Schmidt, C., … Burton, P. R. (2014). DataSHIELD: Taking the analysis to the data, not the data to the analysis. *International Journal of Epidemiology*, *43*(6), 1929–1944. https://doi.org/10.1093/ije/dyu188

Gayle, V., & Lambert, P. (2018). *What is quantitative longitudinal data analysis?* Bloomsbury Academic, an imprint of Bloomsbury Publishing Plc.

GfK NOP Social Research. (n.d.). *Millennium Cohort Study – Sweep 2 – Technical report* (pp. 1–48) [Technical report]. Centre for Longitudinal Studies.

Gil-Hernández, C. J. (2021). The (Unequal) Interplay Between Cognitive and Noncognitive Skills in Early Educational Attainment. *American Behavioral Scientist*, 1–22. https://doi.org/10.1177/0002764221996764

Gnambs, T. (2021). The development of gender differences in information and communication technology (ICT) literacy in middle adolescence. *Computers in Human Behavior*, *114*, 106533. https://doi.org/10.1016/j.chb.2020.106533

Gomensoro, A., & Meyer, T. (2017). TREE (Transitions from Education to Employment): A Swiss multi-cohort survey. *Longitudinal and Life Course Studies*, *8*. https://doi.org/10.14301/llcs.v8i2.424

Granderath, J. S., Martin, A., & Froehlich, L. (2021). The Effect of Participation in Adult Education on Life Satisfaction of Immigrants and Natives: A Longitudinal Analysis. *Journal of Happiness Studies*. https://doi.org/10.1007/s10902-020-00340-7

Graubard, B. I., & Korn, E. L. (1996). Modelling the sampling design in the analysis of health surveys. *Statistical Methods in Medical Research*, *5*(3), 263–281. https://doi.org/10.1177/096228029600500304

Gray, J. C., Gatenby, R., & Simmonds, N. (2009). *Millennium Cohort Study sweep 3: Technical report* (pp. 1–92) [Technical report]. Centre for Longitudinal Studies.

Gray, J. C., Gatenby, R., Simmonds, N., & Huang, Y. (2010). *Millennium Cohort Study sweep 4: Technical report* (pp. 1–98) [Technical report]. Centre for Longitudinal Studies.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, 6(4), 430–450.

Hajdu, T., Kertesi, G., & Kézdi, G. (2014). Roma fiatalok a középiskolában. Beszámoló a TÁRKI Életpálya-felmérésének 2006 és 2012 közötti hullámaiból. In T. Kolosi & I. G. Tóth (Eds.), *Társadalmi riport, 2014* (pp. 265–302). TÁRKI. http://real.mtak.hu/20255/

Hanitzsch, T., & Berganza, R. (2012). Explaining journalists' trust in public institutions across 20 countries: Media freedom, corruption, and ownership matter most. *Journal of Communication*, *62*(5), 794–814. https://doi.org/10.1111/j.1460-2466.2012.01663.x

Heiniger, M., & Imdorf, C. (2018). The role of vocational education in the transmission of gender segregation from education to employment: Switzerland and Bulgaria compared. *Journal for Labour Market Research*, *52*(1), 15. https://doi.org/10.1186/s12651-018-0248-6

Hermann, Z. (2020). The impact of decreasing compulsory school-leaving age on dropping out of school. In *The Hungarian Labour Market 2019* (pp. 70–77). Institute of Economics, Centre for Economic and Regional Studies. http://real.mtak.hu/112542/

Hermann, Z., Horn, D., Köllő, J., Sebők, A., Semjén, A., & Varga, J. (2020). The impact of reading and mathematics test results on future earnings and employment. In *The Hungarian Labour Market 2019* (pp.

45–52). Institute of Economics, Centre for Economic and Regional Studies. https://www.mtakti.hu/wp-content/uploads/2020/07/LMYB2019_InFocus_2_1.pdf

Hermann, Z., & Molnár, T. L. (2008). Országos Kompetenciamérési adatbázis. *Unpublished Manuscript*. http://adatbank.mtakti.hu/files/dokum/7.pdf

Holtz-Bacha, C., & Norris, P. (2001). 'To entertain, inform, and educate': Still the role of public television. *Political Communication*, *18*(2), 123–140. https://doi.org/10.1080/105846001750322943

Hondralis, I., & Kleinert, C. (2021). Do children influence their mothers' decisions? Early child development and maternal employment entries after birth. *Advances in Life Course Research*, *47*, 100378. https://doi.org/10.1016/j.alcr.2020.100378

Horn, D. (2013). Diverging performances: The detrimental effects of early educational selection on equality of opportunity in Hungary. *Research in Social Stratification and Mobility*, *32*, 25–43. https://doi.org/10.1016/j.rssm.2013.01.002

Horn, D. (2016). The effectiveness of apprenticeship training: A within-track comparison of workplace-based and school-based vocational training in Hungary. *Social Science Research*, *55*, 139–154. https://doi.org/10.1016/j.ssresearch.2015.09.002

Horn, D., & Keller, T. (2015). Hungary: The impact of gender culture. In H.-P. Blossfeld, J. Skopek, M. Triventi, & S. Buchholz (Eds.), *Gender, education and employment: An international comparison of school-to-work transitions* (pp. 280–296). Edward Elgar.

Horn, D., Keller, T., & Róbert, P. (2016). Early tracking and competition: A recipe for major inequalities in Hungary. In H.-P. Blossfeld, S. Buchholz, J. Skopek, & M. Triventi (Eds.), *Models of secondary education and social inequality: An international comparison* (pp. 129–147). Edward Elgar.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3–4), 117–144. https://doi.org/10.1080/03610739208253916

Hupka-Brunner, S., Jann, B., Koomen, M., Krebs-Oesch, D., Meyer, T., Müller, B., von Rotz, C., Sacchi, S., & Wilhelmi, B. (2021). *TREE2 study design* [Technical report]. TREE. 10.7892/boris.152018

Imdorf, C., Sacchi, S., Wohlgemuth, K., Cortesi, S., & Schoch, A. (2014). How cantonal education systems in Switzerland promote gender-typical school-to-work transitions. *Swiss Journal of Sociology*, *40*(2), 175–196. https://doi.org/10.7892/boris.67589

INED. (2021). *French Longitudinal Study of Children*. Ined - Institut National d'études Démographiques. https://www.elfe-france.fr/en/

International Association for the Evaluation of Educational Achievement. (2021). *TIMSS | IEA*. IEA. https://www.iea.nl/studies/iea/timss

Jensen, S. S. (2020). Effects of school-wide positive behavior support in Denmark: Results from the Danish National Register data. *School Effectiveness and School Improvement*, *0*(0), 1–19. https://doi.org/10.1080/09243453.2020.1840398

Joensen, J. S., & Nielsen, H. S. (2009). Is There a Causal Effect of High School Math on Labor Market Outcomes? *The Journal of Human Resources*, *44*(1), 171–198.

Johnstone, D. J. (1987). Tests of significance following R. A. Fisher. *The British Journal for the Philosophy of Science*, *38*(4), 481–499.

Johnstone, D. J. (1989). On the necessity for random sampling. *The British Journal for the Philosophy of Science*, *40*(4), 443–457.

Jonkmann, K., Thoemmes, F., Lüdtke, O., & Trautwein, U. (2014). Personality traits and living arrangements in young adulthood: Selection and socialisation. *Developmental Psychology*, *50*(3), 683–698. https://doi.org/10.1037/a0034239

Kampa, N., Fleckenstein, J., Schmidt, F. T. C., Meyer, J., Enke, F., Köller, O., & Leibniz-Institut Für Die Pädagogik Der Naturwissenschaften Und Mathematik (IPN) An Der Christian-Albrechts-Universität Zu Kiel. (2020). *Educational Outcomes of Students from Vocational and Academic Upper Secondary Schools (LISA 6)Lernergebnisse an beruflichen und allgemeinbildenden Gymnasien in Schleswig-Holstein (LISA 6)* (Version 1) [Data set]. IQB - Institute for Educational Quality Improvement. https://doi.org/10.5159/IQB_LISA_6_V1

Kampa, N., Fleckenstein, J., Schmidt, F. T. C., Meyer, J., Enke, F., Köller, O., & Leibniz-Institut Für Die Pädagogik Der Naturwissenschaften Und Mathematik (IPN) An Der Christian-Albrechts-Universität Zu Kiel. (2021). *Reading in Secondary Education (LISA)Lesen in der Sekundarstufe (LISA)* (Version 1) [Data set]. IQB - Institute for Educational Quality Improvement. https://doi.org/10.5159/IQB_LISA_V1

Keller, T. (2014). *Talented but unaware? An analysis of the role of self-assessment in educational transition* (Budapest Working Papers on the Labour Market BWP-2014/9). https://doi.org/10.2139/ssrn.2554261

Kertesi, G., & Kézdi, G. (2011). The Roma/Non-Roma Test Score Gap in Hungary. *American Economic Review*, *101*(3), 519–525. https://doi.org/10.1257/aer.101.3.519

Kish, L., & Frankel, M. R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society*, *36*(1), 1–37.

Köller, O., Meyer, J., Saβ, S., & Baumert, J. (2019). New analyses of an old topic: Effects of intelligence and motivation on academic achievement. *Journal for Educational Research Online*, *11*(1), 166–189.

Könings, K. D., Brand-Gruwel, S., & Elen, J. (2012). Effects of a school reform on longitudinal stability of students' preferences with regard to education. *British Journal of Educational Psychology*, *82*(3), 512–532. https://doi.org/10.1111/j.2044-8279.2011.02044.x

Könings, K. D., Brand-Gruwel, S., & Merriënboer, J. J. G. van. (2011). The match between students' lesson perceptions and preferences: Relations with student characteristics and the importance of motivation. *Educational Research*, *53*(4), 439–457. https://doi.org/10.1080/00131881.2011.625155

Könings, K. D., Brand-Gruwel, S., van Merriënboer, J. J. G., & Broers, N. J. (2008). Does a new learning environment come up to students' expectations? A longitudinal study. *Journal of Educational Psychology*, *100*(3), 535–548. https://doi.org/10.1037/0022-0663.100.3.535

Kreft, I. G. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. SAGE Publications, Inc.

Kroh, M., Kühne, S., Siegers, R., & Belcheva, V. (2018). *SOEP-Core – Documentation of sample sizes and panel attrition (1984 until 2016)* (SOEP Survey Papers 480: Series C, p. 90). DIW Berlin, The German Socio-Economic Panel (SOEP). https://www.diw.de/documents/publikationen/73/diw_01.c.579464.de/diw_ssp0480.pdf

Kühne, R. (2018). Measurement invariance. In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), *The international encyclopedia of communication research methods*. John Wiley & Sons, Inc.

Lehmann, E. L. (1993). The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two? *Journal of the American Statistical Association*, *88*(424), 1242–1249. https://doi.org/10.2307/2291263

LiSO. (2017a). *Het LiSO-project kort samengevat*. LiSO-Project. https://lisoproject.be/

LiSO. (2017b). *Onderzoek*. LiSO-Project. https://lisoproject.be/onderzoek

LiSO. (2021). *FAQ*. LiSO-Project. https://lisoproject.be/faq

Little, T., Cunningham, W., Shahar, G., & Widaman, K. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling A Multidisciplinary Journal*, *9*(2), 151–173. https://doi.org/10.1207/S15328007SEM0902_1

Lohmann, H., Spieß, C. K., & Feldhaus, C. (2009). Private schools in Germany: Attendance up, but not among the children of less educated parents. *DIW Weekly Report*, *5*(29), 203–208. http://hdl.handle.net/10419/151068

MacCallum, R., Widaman, K., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84

MacKinnon, J. G., & Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, *32*(2), 233–254. https://doi.org/10.1002/jae.2508

McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods*, *14*(2), 126–149. https://doi.org/10.1037/a0015857

Menard, S. (2002). *Longitudinal research* (2nd ed.). Sage.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/BF02294825

Meyer, J., Fleckenstein, J., Retelsdorf, J., & Köller, O. (2019). The relationship of personality traits and different measures of domain-specific achievement in upper secondary education. *Learning and Individual Differences*, *69*, 45–59. https://doi.org/10.1016/j.lindif.2018.11.005

Miller, A. (2017, August 10). When to use fixed effects vs. Clustered standard errors? *Allex P. Miller*. https://alex.miller.im/posts/when-to-use-fixed-effects-vs-clustered-standard-errors-panel-data/

Millsap, R. E. (1995). Measurement invariance, predictive invariance, and the duality paradox. *Multivariate Behavioral Research*, *30*(4), 577–605. https://doi.org/10.1207/s15327906mbr3004_6

Millsap, R. E. (1997). Invariance in measurement and prediction: Their relationship in the single-factor case. *Psychological Methods*, *2*(3), 248–260. https://doi.org/10.1037/1082-989X.2.3.248

Molnár, G. (2020). Youth in public employment, with particular emphasis on early secondary school leavers. In *The Hungarian Labour Market 2019* (pp. 115–120). Institute of Economics, Centre for Economic and Regional Studies. https://www.mtakti.hu/wp-content/uploads/2020/07/LMYB2019_InFocus_5_5.pdf

Müller, B., & Wolter, S. C. (2014). The role of hard-to-obtain information on ability for the school-to-work transition. *Empirical Economics*, *46*(4), 1447–1471. https://doi.org/10.1007/s00181-013-0709-2

Muntoni, F., & Retelsdorf, J. (2019). At their children's expense: How parents' gender stereotypes affect their children's reading outcomes. *Learning and Instruction*, *60*, 95–103. https://doi.org/10.1016/j.learninstruc.2018.12.002

Murdoch, J., Guégnard, C., Griga, D., Koomen, M., & Imdorf, C. (2016). How do second-generation immigrant students access higher education? The importance of vocational routes to higher education in Switzerland, France, and Germany. *Swiss Journal of Sociology*, *42*(2), 245–265. https://doi.org/10.1515/sjs-2016-0011

Murdoch, J., Guégnard, C., Koomen, M., Imdorf, C., Kamanzi, C., & Meyer, T. (2017). Pathways fostering mobility to higher education for vulnerable immigrants in France, Switzerland and Canada. *European Journal of Higher Education*, *7*(1), 29–42. https://doi.org/10.1080/21568235.2017.1254918

National Educational Panel Study. (2021a). *National Educational Panel Study*. Data Center > Data and Documentation. https://www.neps-data.de/Data-Center/Data-and-Documentation

National Educational Panel Study. (2021b). *National Educational Panel Study*. Project Overview > Structure. https://www.neps-data.de/Project-Overview/Structure

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear statistical models* (4th ed.). McGraw-Hill.

Neumann, M., Trautwein, U., & Nagy, G. (2011). Do central examinations lead to greater grading comparability? A study of frame-of-reference effects on the University entrance qualification in Germany. *Studies in Educational Evaluation*, *37*(4), 206–217. Scopus. https://doi.org/10.1016/j.stueduc.2012.02.002

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, *97*(4), 558–625. https://doi.org/10.2307/2342192

Oktatási Hivatal. (2012). Általános Leírás. https://www.oktatas.hu/kozneveles/meresek/kompetenciameres/alt_leiras

Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.). (2008). *TIMSS 2007 technical report*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. https://timssandpirls.bc.edu/TIMSS2007/PDF/TIMSS2007_TechnicalReport.pdf

Oud, J. (2001). Quasi-Longitudinal Designs in SEM State Space Modeling. *Statistica Neerlandica*, *55*, 200–220. https://doi.org/10.1111/1467-9574.00165

Parker, P. D., Schoon, I., Tsai, Y.-M., Nagy, G., Trautwein, U., & Eccles, J. S. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: A multicontext study. *Developmental Psychology*, *48*(6), 1629–1642. https://doi.org/10.1037/a0029167

Plewis, I., Calderwood, L., Hawkes, D., Hughes, G., & Joshi, H. (2007). *The Millennium Cohort Study: Techical report on sampling* (pp. 1–78) [Technical report]. Centre for Longitudinal Studies. http://doc.ukdataservice.ac.uk/doc/4683/mrdoc/pdf/mcs_technical_report_on_sampling_4th_edition.pdf

Pøhler, L., & Sørensen, S. A. (2010). *Nationale test og anden evaluering af elevens læsning* (1st ed.). Dafolo.

Rambøll. (2013). *Evaluering af de nationale test* (pp. 1–126). Rambøll. https://dk.ramboll.com/~/media/Files/RM/Rapporter/Nationale%20test%202013.pdf

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). SAGE Publications, Inc.

Retelsdorf, J., Becker, M., Köller, O., & Möller, J. (2012). Reading development in a tracked school system: A longitudinal study over 3 years using propensity score matching. *British Journal of Educational Psychology*, *82*(4), 647–671. https://doi.org/10.1111/j.2044-8279.2011.02051.x

Retelsdorf, J., Köller, O., & Möller, J. (2014). Reading achievement and reading self-concept – Testing the reciprocal effects model. *Learning and Instruction*, *29*, 21–30. https://doi.org/10.1016/j.learninstruc.2013.07.004

Samuel, R. (2014). The gendered interplay between success and well-being during transitions. *Educational Research*, *56*(2), 202–219. https://doi.org/10.1080/00131881.2014.898915

Scharenberg, K., Wohlgemuth, K., & Hupka-Brunner, S. (2017). Does the structural organisation of lower-secondary education in Switzerland influence students' opportunities of transition to upper-secondary education? A multilevel analysis. *Swiss Journal of Sociology*, *43*(1), 63–88. https://doi.org/10.1515/sjs-2017-0004

Schiltz, F., Mazrekaj, D., Horn, D., & De Witte, K. (2019). Does it matter when your smartest peers leave your class? Evidence from Hungary. *Labour Economics*, *59*, 79–91. https://doi.org/10.1016/j.labeco.2019.04.001

Schröder, M., Siegers, R., & Spieß, C. K. (2013). 'Familien in Deutschland' - FiD: Enhancing Research on Families in Germany. In *SOEPpapers on Multidisciplinary Panel Data Research* (No. 556; SOEPpapers on Multidisciplinary Panel Data Research). DIW Berlin, The German Socio-Economic Panel (SOEP). https://www.econstor.eu/handle/10419/74714

Schroeder, C., Spieß, C. K., & Storck, J. (2015). Private spending on children's education: Low-income families pay relatively more. *DIW Economic Bulletin*, *8*, 113–123.

Schuck, A. R. T., Vliegenthart, R., Boomgaarden, H. G., Elenbaas, M., Azrout, R., Spanje, J. van, & Vreese, C. H. de. (2013). Explaining campaign news coverage: How medium, time, and context explain variation in the media framing of the 2009 European Parliamentary elections. *Journal of Political Marketing*, *12*(1), 8–28. https://doi.org/10.1080/15377857.2013.752192

Schuck, A. R. T., Vliegenthart, R., & de Vreese, C. H. (2016). Who's afraid of conflict? The mobilising effect of conflict framing in campaign news. *British Journal of Political Science*, *46*(1), 177–194. https://doi.org/10.1017/S0007123413000525

Sebők, A. (2019a). A KRTK Adatbank Kapcsolt Államigazgatási Paneladatbázisa. *Közgazdasági Szemle*, *66*(11), 1230–1236.

Sebők, A. (2019b). *The Panel of Linked Administrative Data of CERS Databank* (No. 2; Budapest Working Papers on the Labour Market, pp. 1–11). Institute of Economics, Centre for Economic and Regional Studies. https://www.mtakti.hu/wp-content/uploads/2019/12/BWP1902.pdf

Skinner, C. J., Holt, D., & Smith, T. M. F. (1989). *Analysis of complex surveys*. Wiley.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalised latent variable modeling: Multilevel, longitudinal, and structural equation models*. Chapman and Hall/CRC. https://doi.org/10.1201/9780203489437

Smelser, N. (1973). The methodology of comparative analysis. In D. Warwick & S. Osherson (Eds.), *Comparative research methods* (pp. 45–52). Prentice-Hall.

SONO. (2021). *Kwaliteitsvol onderwijs bieden in functie van de lerende*. Steunpunt SONO. http://steunpuntsono.be/delerende/

Spieß, C. K., & Buechner, C. (2009). Children who attend formal day care do better in school—Even many years later in secondary school. *DIW Weekly Report*, *5*, 31–34.

Statistics Denmark. (2014). *The Danish system for access to micro data*. Statistics Denmark. https://www.dst.dk/Site/Dst/SingleFiles/GetArchiveFile.aspx?fi=5452354440&fo=0&ext=israel2016

Sterba, S. K. (2009). Alternative model-based and design-based frameworks for inference from samples to populations: From polarisation to integration. *Multivariate Behavioral Research*, *44*(6), 711–740. https://doi.org/10.1080/00273170903333574

Stevens, E., Dockx, J., Custers, C., Fidlers, I., De Fraine, B., & Van Damme, J. (2015). *LiSO-project: Steekproef september 2013* (SSL/2015.10/2.1; p. 77). Steunpunt Studie- en Schoolloopbanen. https://lisoproject.be/SSL_2015.10_2.1-TechnischRapport_Steekproef_2013

Sugden, R. A., & Smith, T. M. F. (1984). Ignorable and informative designs in survey sampling inference. *Biometrika*, *71*(3), 495–506. https://doi.org/10.2307/2336558

The SOEP group. (2020). *SOEP annual report 2019* (No. 10; SOEP Annual Report, p. 112). DIW Berlin, The German Socio-Economic Panel (SOEP). https://www.diw.de/documents/publikationen/73/diw_01.c.795906.de/soep_annual_report_2019.pdf

Thissen, D., & Wainer, H. (2001). *Test scoring*. Erlbaum.

Thurstone, L. L. (1947). *Multiple factor analysis*. University of Chicago Press.

TREE. (2016). Documentation on the first TREE cohort (TREE1), 2000-2016. In *TREE,—(2016). Documentation on the first TREE cohort (TREE1), 2000–2016 Bern: TREE* [Info:eu-repo/semantics/report]. TREE. https://boris.unibe.ch/99243/

University of Bern. (2021). *Transitions from Education to Employment*. Transitions from Education to Employment. https://www.tree.unibe.ch/index_eng.html

Vad Andersen, M. (2021, January 14). *Danish National Tests*. TrygFonden's Centre for Child Research. https://childresearch.au.dk/en/signature-project-read/struktur-og-rammer/danish-national-tests/

Vandenberghe, N., Maes, F., & Van Damme, J. (2012). *Een introductie in de SiBO-databank* (pp. 1–28). Studie en Schoolloopbanen. http://www.informatieportaalssl.be/SiBO-Databank/_AlgemeneIntroductie/DOC_IntroductieSiBODatabank.pdf

Vandentorren, S., Bois, C., Pirus, C., Sarter, H., Salines, G., & Leridon, H. (2009). Rationales, design and recruitment for the Elfe longitudinal study. *BMC Pediatrics*, 9, 58. https://doi.org/10.1186/1471-2431-9-58

Vanwynsberghe, G., Vanlaar, G., Damme, J. V., & Fraine, B. D. (2019). Long-term effects of first-grade teachers on students' achievement: A replication study. *School Effectiveness and School Improvement*, *30*(2), 177–193. https://doi.org/10.1080/09243453.2018.1556164

Vanwynsberghe, G., Vanlaar, G., Van Damme, J., & De Fraine, B. (2017a). Long-term effects of primary schools on educational positions of students 2 and 4 years after the start of secondary education. *School Effectiveness and School Improvement*, *28*(2), 167–190. https://doi.org/10.1080/09243453.2016.1245667

Vanwynsberghe, G., Vanlaar, G., Van Damme, J., & De Fraine, B. (2017b). Long-term effects of primary schools on mathematics achievement of students at age 17. *British Educational Research Journal*, *43*(6), 1131–1148. https://doi.org/10.1002/berj.3311

Vanwynsberghe, G., Vanlaar, G., Van Damme, J., & De Fraine, B. (2017c). Long-term effects of primary schools on non-cognitive outcomes of students at age 17. *Studies in Educational Evaluation*, *55*, 83–93. https://doi.org/10.1016/j.stueduc.2017.07.005

Vasalampi, K., Parker, P., Tolvanen, A., Lüdtke, O., Salmela-Aro, K., & Trautwein, U. (2014). Integration of personality constructs: The role of traits and motivation in the willingness to exert effort in academic and social life domains. *Journal of Research in Personality*, *48*, 98–106. https://doi.org/10.1016/j.jrp.2013.11.004

Verhaeghe, J. P., Vanlaar, G., Knipprath, H., De Fraine, B., & Van Damme, J. (2018). Can group composition effects explain socioeconomic and ethnic achievement gaps in primary education? *Studies in Educational Evaluation*, *57*, 6–15. https://doi.org/10.1016/j.stueduc.2017.07.006

Verschueren, K., Lavrijsen, J., Weyns, T., Ramos, A., & De Fraine, B. (2019). Social Acceptance of High-Ability Youth: Multiple Perspectives and Contextual Influences. *New Directions for Child and Adolescent Development*, *168*. https://doi.org/10.1002/cad.20316

Volodina, A., Nagy, G., & Retelsdorf, J. (2015). Berufliche Interessen und der Übergang in die gymnasiale Profiloberstufe: Ihre Struktur und Vorhersagekraft für das individuelle Wahlverhalten. *Zeitschrift Für Pädagogische Psychologie*, *29*(2), 89–100. https://doi.org/10.1024/1010-0652/a000154

Wandall, J. (2017). National Tests in Denmark: CAT as a pedagogic tool. *FORUM*, *59*(2), 145. https://doi.org/10.15730/forum.2017.59.2.145

Wilson, S. E., & Butler, D. M. (2007). A lot more to do: The sensitivity of time-series cross-section analyses to simple alternative specifications. *Political Analysis*, *15*(2), 101–123. https://doi.org/10.1093/pan/mpl012

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*(1), 58–79. https://doi.org/10.1037/1082-989X.12.1.58

Wirth, W., & Kolb, S. (2012). Securing equivalence: Problems and solutions. In F. Esser & T. Hanitzsch (Eds.), *The handbook of comparative communication research*. Routledge.

Wolfson, M., Wallace, S. E., Masca, N., Rowe, G., Sheehan, N. A., Ferretti, V., LaFlamme, P., Tobin, M. D., Macleod, J., Little, J., Fortier, I., Knoppers, B. M., & Burton, P. R. (2010). DataSHIELD: Resolving a conflict in contemporary bioscience - performing a pooled analysis of individual-level data without sharing the data. *International Journal of Epidemiology*, *39*(5), 1372–1382. https://doi.org/10.1093/ije/dyq111

Wong, M. Y., Day, N. E., Luan, J. A., Chan, K. P., & Wareham, N. J. (2003). The detection of gene-environment interaction for continuous traits: Should we deal with measurement error by bigger studies or better measurement? *International Journal of Epidemiology*, *32*(1), 51–57. https://doi.org/10.1093/ije/dyg002

InGRID-2
Integrating Research Infrastructure for European expertise on Inclusive Growth from data to policy

Referring to the increasingly challenging EU2020-ambitions of Inclusive Growth, the objectives of the InGRID-2 project are to advance the integration and innovation of distributed social sciences research infrastructures (RI) on 'poverty, living conditions and social policies' as well as on 'working conditions, vulnerability and labour policies'. InGRID-2 will extend transnational on-site and virtual access, organise mutual learning and discussions of innovations, and improve data services and facilities of comparative research. The focus areas are (a) integrated and harmonised data, (b) links between policy and practice, and (c) indicator-building tools.
Lead users are social scientist involved in comparative research to provide new evidence for European policy innovations. Key science actors and their stakeholders are coupled in the consortium to provide expert services to users of comparative research infrastructures by investing in collaborative efforts to better integrate microdata, identify new ways of collecting data, establish and improve harmonised classification tools, extend available policy databases, optimise statistical quality, and set-up micro-simulation environments and indicator-building tools as important means of valorisation. Helping scientists to enhance their expertise from data to policy is the advanced mission of InGRID-2. A new research portal will be the gateway to this European science infrastructure.

This project is supported by the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 730998.

More detailed information is available on the website: www.inclusivegrowth.eu

Co-ordinator
Monique Ramioul

KU LEUVEN HIVA
RESEARCH INSTITUTE FOR WORK AND SOCIETY

Partners

TÁRKI Social Research Institute Inc. (HU)
Amsterdam Institute for Advanced Labour Studies – AIAS, University of Amsterdam (NL)
Swedish Institute for Social Research - SOFI, Stockholm University (SE)
Economic and Social Statistics Department, Trier University (DE)
Centre for Demographic Studies – CED, University Autonoma of Barcelona (ES)
Luxembourg Institute of Socio-Economic Research – LISER (LU)
Herman Deleeck Centre for Social Policy – CSB, University of Antwerp (BE)
Institute for Social and Economic Research - ISER, University of Essex (UK)
German Institute for Economic Research – DIW (DE)
Centre for Employment and Work Studies – CEET, National Conservatory of Arts and Crafts (FR)
Centre for European Policy Studies – CEPS (BE)
Department of Economics and Management, University of Pisa (IT)
Department of Social Statistics and Demography – SOTON, University of Southampton (UK)
Luxembourg Income Study – LIS, asbl (LU)
School of Social Sciences, University of Manchester (UK)
Central European Labour Studies Institute – CELSI (SK)
Panteion University of Social and Political Sciences (GR)
Central Institute for Labour Protection – CIOP, National Research Institute (PL)

InGRID-2

Integrating Research Infrastructure for European expertise on Inclusive Growth from data to policy Contract N° 730998

For further information about the InGRID-2 project, please contact
inclusive.growth@kuleuven.be
www.inclusivegrowth.eu
p/a HIVA – Research Institute for Work and Society
Parkstraat 47 box 5300
3000 Leuven
Belgium