



Towards a European Panel on Education: A Feasibility Study on the Harmonisation of Longitudinal Educational Datasets

WP 3 - Deliverable 3.2

Feasibility Study

Lead authors: Emmanuele Pavolini, Gianluca Argentin

Authors: Ekaterina Kochergina, Giulietta Zanga, Ilaria Lievore, Vincenzo Mauro, Rosita Pretaroli

Main Reviewers: Roberto Ricci, Patrizia Falzetti, Emanuela Sala, Veronica Mobilio

2026



Funded by the European Union's Horizon Europe Framework Programme (HORIZONCL2-2023-TRANSFORMATIONS-01-06) under grant agreement No 101132455. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.



Deliverable Information

Project Information	
Project number	101132455
Project acronym	LINEup
Project title	LINEup - Longitudinal data for INequalities in Education
Start date	01/02/2024
Duration	36 months
Type of Action	HORIZON-RIA
Call Identifier	HORIZON-CL2-2023-TRANSFORMATIONS-01
Project website	https://www.lineup-project.eu/

Deliverable D3.2 Feasibility Study	
Work Package	WP3
Lead Partner	University of Milan (UNIMI)
Due date	31/03/2026
Submission date	31/03/2026
Dissemination level	PU-Public
Lead authors	Emmanuele Pavolini, Gianluca Argentin
Authors	Ekaterina Kochergina, Giulietta Zanga, Ilaria Lievore, Vincenzo Mauro, Rosita Pretaroli
Lead Reviewers	Patrizia Falzetti, Veronica Mobilio, Roberto Ricci, Emanuela Sala
Other Reviewers	Martin Neugebauer & Johanna Unewisse (PHKA), Luis Ortiz & Kars van Oosterhout (UPF), Léonard Moulin & Priya Maurya (INED), Ioannis Cholezas & Nikolaos Kanellopoulos (KEPE), Paulo Moreira & Carlos Alfonso (UTAD)
Version	1

History of changes

Version	Date	Change History	Organisation
0.1	12-03-2026	Initial version	UNIMI-Bicocca-FpS
0.2	18-03-2026	First Review	Comments From Lead Reviewers
0.3	21/03/2026	Second Review	UNIMI-Bicocca-FpS (revision)
0.4	26/03/2026	Third Review	Comments from Partners (revision)
1	30/03/2026	Final version	UNIMI-Bicocca-FpS

Disclaimer: This project is funded by the European Union's Horizon programme under grant agreement No 101132455. However, the views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Feasibility Study - Author Statement

Lead authors: Emmanuele Pavolini¹, Gianluca Argentin²

Authors: Ekaterina Kochergina², Giulietta Zanga¹, Ilaria Lievore³, Vincenzo Mauro⁵, Rosita Pretaroli⁵,

Lead Reviewers: Roberto Ricci⁴, Patrizia Falzetti⁴, Emanuela Sala², Veronica Mobilio³

Other Reviewers: Martin Neugebauer & Johanna Unewisse⁶, Luis Ortiz & Kars van Oosterhout⁷, Léonard Moulin & Priya Maurya⁸, Ioannis Cholezas & Nikolaos Kanellopoulos⁹, Paulo Moreira & Carlos Alfonso¹⁰

Participating Institutions:

1. Università degli Studi di Milano La Statale (UNIMI - WP3 leader), Italy
2. Università degli Studi di Milano-Bicocca (UniMiBicocca – WP3 leader), Italy
3. Fondazione per la Scuola della Compagnia di San Paolo (FpS – LINEup project coordinator), Italy
4. National Institute for the Evaluation of the Education and Training System (INVALSI), Italy
5. Università Degli Studi di Macerata (UNIMC), Italy
6. Karlsruhe University of Education (PHKA), Germany
7. Universidad Pompeu Fabra (UPF), Spain
8. Institut National D'Etudes Demographiques (INED), France
9. Centre of Planning and Economic Research (KEPE), Greece
10. Universidade de Trás-os-Montes e Alto Douro (UTAD), Portugal

CRedit Authorship Contribution Statement¹:

Emmanuele Pavolini: Conceptualisation, Methodology, Validation, Formal Analysis, Writing - Original Draft, Writing - Review & Editing, Visualisation, Supervision, Project administration

Gianluca Argentin: Conceptualisation, Methodology, Validation, Formal Analysis, Writing - Original Draft, Writing - Review & Editing, Visualisation, Project administration, Supervision

Ekaterina Kochergina: Methodology, Validation, Investigation, Data Curation, Writing - Original Draft, Visualisation

Giulietta Zanga: Methodology, Validation, Investigation, Data Curation, Writing - Original Draft, Visualisation

Ilaria Lievore: Methodology, Validation, Investigation, Data Curation, Writing - Original Draft, Visualisation

Vincenzo Mauro: Methodology, Validation.

Rosita Pretaroli: Methodology, Validation.

Infographic Layout: Andrea Succhielli, European Grants International Academy (EGInA)

Corresponding author: Veronica Mobilio, Piazza Bernini, 5 10138 Torino - Italy, +39 011 1563 0500, veronica.mobilio@fondazione scuola.it

Acknowledgement of Contributors: the authors gratefully acknowledge the researchers and institutional teams who supported this feasibility study by facilitating access to the relevant datasets and contributing to the validation of the harmonisation procedures. Contributors include colleagues from the ten institutions listed above.

How to cite:

Pavolini E., Argentin G., Kochergina E., Zanga G., Lievore I., Mauro V., Pretaroli R. (2026). *Towards a European Panel on Education: A Feasibility Study on the Harmonisation of Longitudinal Educational Datasets*. Deliverable D3.2, Horizon Europe Project LINEup.

¹ CRediT Authorship: <https://www.elsevier.com/researcher/author/policies-and-guidelines/credit-author-statement>

Abstract

The present feasibility study, conducted as part of the LINEup project (Longitudinal Data for Inequalities in Education), addresses a critical gap in European educational research: the lack of a comparable longitudinal perspective on student achievement. While international large-scale assessments like PISA provide valuable cross-sectional insights, they do not track individual progress over time, thereby limiting our understanding of how educational inequalities emerge and evolve. This study represents **one of the first systematic attempts to harmonise national longitudinal datasets across eight European countries** – including Germany, Greece, Spain (Catalonia), France, Italy, the Netherlands, Portugal, and the United Kingdom - some of which had previously remained inaccessible for comparative research.

The study adopts an **ex-post (retrospective) harmonisation approach**, applying a structured and replicable framework to improve comparability across existing datasets. It follows a “**minimum harmonisation**” logic, aiming to standardise key variables while preserving as much of the original information as possible. The analysis focuses on **eight major datasets**, encompassing both survey-based longitudinal studies (e.g., Germany’s NEPS and the UK’s Millennium Cohort Study) and administrative registers (e.g., Greece’s MySchool and Portugal’s DGEEC). Access to highly restricted administrative microdata was successfully negotiated, highlighting both the feasibility and the complexity of cross-national data integration.

The harmonisation process led to the identification of **44 target variables** across four core domains: individual student characteristics, household background, school and teacher factors, and academic performance. Given the heterogeneity of national assessment systems - ranging from standardised test scores to IRT-based measures and teacher-assigned grades - student performance was operationalised through **within-dataset standardisation (z-scores) in literacy and numeracy**, the most consistently measured domains.

A key finding of the study is that, while ex-post harmonisation enables meaningful descriptive and within-country longitudinal analyses, the **construction of a single fully comparable cross-national longitudinal dataset is currently not feasible under existing data conditions**. This limitation is driven by structural differences in data access conditions, measurement frameworks, and the absence of shared standards in longitudinal educational data production. In addition, gaps in household-level information and the uneven availability of school and teacher data limit the possibility of systematically analysing the micro-level mechanisms underlying educational inequalities.

Considering these constraints, the study identifies a **viable alternative strategy** based on coordinated, parallel analyses across harmonised national datasets, rather than full data integration. At the same time, it argues that addressing educational inequalities in a robust and comparable way requires a **shift from ex-post harmonisation to stronger coordination at the point of data production**. Proposed directions include the development of shared metadata standards, the longitudinal extension of existing international surveys (e.g., PIRLS, TIMSS), and the potential creation of a dedicated European longitudinal panel, comparable to the SHARE infrastructure in the field of health and ageing. By providing harmonised metadata and coding frameworks, this study offers a concrete foundation for advancing comparative longitudinal research within current constraints, while outlining a strategic agenda for the development of future European data infrastructures.

Table of Contents

Executive Summary	7
Core Objective.....	7
Datasets Analysed.....	7
Main Harmonisation Process.....	8
Main Obstacles.....	8
Main Conclusions and Strategic Directions.....	9
1. Introduction	10
2. Methodological Framework	12
3. Data Sources	18
3.1 The general characteristics of the datasets and the selected cohorts: an overview	18
NEPS (DE)	18
MySchool (EL).....	19
RALC (ES – Catalonia).....	19
DEPP Panels (FR).....	20
INVALSI-SNV (IT).....	20
Netherlands Cohort Study on Education (NL).....	21
Portuguese Student Register - DGEEC (PT)	21
Millennium Cohort Study (UK)	22
3.2 Cohorts selection and access to datasets	24
4. Harmonised Variables Definitions: Main Challenges and Solutions	28
4.1 Individual and Household Variables	30
Student's individual characteristics	30
Student's household characteristics	32
4.2 School and Teacher Variables	37
School variables.....	39
Teacher characteristics.....	41
4.3 Academic Performance Measures.....	43
Methodological Issues and Choices	43
Open Issues	44
5. Distribution Checks and Data Quality Patterns in the Harmonised Datasets	45
5.1 An overview by dataset	46
5.2 Distribution of Independent Variables by Country and Dataset.....	50
5.3 Description and Distribution of Dependent Variables by Country and Dataset.....	65
5.4 Key Patterns Across Datasets	80

6. Limitations.....	82
6.1 Harmonisation and Use of Individual and Household Variables.....	82
Difference in the objectives, methods, and categorisation choices of the datasets	82
Harmonised individual and household variables	83
6.2 Harmonisation and Use of School and Teachers Variables	85
6.3 Harmonisation of Academic Performance Measures	86
7. Lesson Learned: Conclusion and Discussion	89
Importance of the Feasibility study	89
Why this matters.....	89
Limits of ex-post harmonisation: a comparative perspective	90
7.1 Implications for Cross-National Comparative Research Use	90
What ex-post harmonization can support.....	90
What ex-post harmonization cannot credibly support	91
What the feasibility study helped us to do.....	91
7.2 Cost-Effectiveness Reflections: Toward a Coordinated European Data Collection Effort?	91
Option 1: Ex-post harmonisation of existing longitudinal datasets	92
Option 2: Incremental coordination of existing national studies.....	92
Option 3: Transformation of international cross-sectional studies into a longitudinal panel	95
Option 4: Creation of a European longitudinal panel on educational performances	96
7.3 Conclusions	97
8. References	98

Executive Summary

Core Objective

The primary objective of this feasibility study is to address a major gap in European educational research: **the lack of a comparable longitudinal perspective on student achievement**. While cross-sectional surveys such as PISA provide valuable snapshots of performance, they do not track individual trajectories over time, limiting the ability to understand how educational inequalities emerge and evolve.

To address this gap, the study pursued four main objectives:

- **Assessing Comparability:** Evaluating the extent to which existing national longitudinal datasets can be harmonised in terms of both learning outcomes (test scores and grades) and explanatory variables (socio-demographic and school-related factors);
- **Developing a Harmonisation Framework:** Identifying strategies and linking techniques to improve comparability of student achievement measures across heterogeneous national systems.
- **Building Technical Infrastructure:** Producing a harmonisation codebook and a reference syntax (in Stata) to support the construction of comparable variables across datasets.
- **Testing Feasibility:** Determining whether independent national datasets can be integrated into a single cross-national framework, or whether alternative analytical strategies are more appropriate.

The study focuses on literacy and numeracy, as core competencies for individual development and employability, and the most consistently measured domains across datasets.

Datasets Analysed

The analysis covers **eight countries** - Germany (DE), Greece (EL), Spain (ES - Catalonia), France (FR), Italy (IT), the Netherlands (NL), Portugal (PT), and the United Kingdom (UK) - selected to reflect the diversity of European education systems. It includes **twenty-two cohorts** drawn from **eight major longitudinal datasets**, representing the two main types of high-quality longitudinal designs: survey-based studies and administrative registers.

- **DE - NEPS:** The National Educational Panel Study, featuring Starting Cohort 2 (Kindergarten) and Starting Cohort 3 (Grade 5). These are pure longitudinal surveys with rich contextual data from parents and teachers.
- **EL - MySchool:** A unified administrative information system covering the entire student population nationwide. It tracks students longitudinally from Grade 7 through 12, focusing on attendance, marks, and graduation.
- **ES (Catalonia) - RALC:** An administrative student register for Catalonia. It provides longitudinal trajectories for students, including standardized test scores in Grade 6 and Grade 10.
- **FR - DEPP Panels:** Pure longitudinal panels (2007 and 2011 cohorts) produced by the Ministry of Education, following students from entry into secondary education.
- **IT - INVALSI-SNV:** A design that links repeated cross-sectional assessments through a unique student identifier, allowing for the construction of longitudinal trajectories. It combines test results with detailed student and school questionnaires.

- **NL - NCO:** A collection of administrative education registries covering the full student population, including standardized end-of-primary tests and secondary examination scores.
- **PT - DGEEC:** The Portuguese Student Register, integrating multiple administrative databases (MISI, E360, CIF) to follow students from Grade 5 through Grade 12.
- **UK - MCS:** The Millennium Cohort Study, a pure longitudinal birth cohort study following children born in 2000-2002 across the UK.

A major achievement was securing access to highly restricted administrative microdata in EL, ES (Catalonia), and PT, which have historically been underutilized in international comparative research.

Main Harmonisation Process

The study applies an **ex-post (retrospective) harmonisation strategy** based on a “minimum harmonisation” logic, aiming to maximise comparability while preserving as much of the original information as possible.

A total of **44 target variables were harmonised** across four core domains:

- **Student Individual Characteristics:** Standardisation was achieved for key variables such as sex, birth month/year, and country of birth. Due to varying levels of detail, the project adopted a simplified categorical variable distinguishing between native-born and foreign-born students.
- **Household Characteristics:** The team harmonised parental education (ISCED levels), occupation (ISCO codes), and employment status. However, household income could only be harmonised into national quartiles and deciles for a subset of datasets (NEPS, DEPP, MCS, NCO).
- **School and Teacher Variables:** Indicators for school type (public/private), school level, and school track (academic, technical, vocational) were successfully harmonised. The project also created compositional variables, such as the share of migrant and female students at both the school and class levels.
- **Academic Performance Measures:** This was the most challenging domain. To address extreme heterogeneity in scoring systems (ranging from raw marks to IRT-based psychometric scores), the study transformed original results into standardized z-scores within each dataset. This allows researchers to compare a student’s relative position within their national distribution over time.

This approach allows for the analysis of students’ relative performance within national contexts over time, while ensuring a minimum level of cross-country comparability.

Main Obstacles

The study identifies several structural barriers to the creation of a fully integrated European longitudinal dataset:

- **Heterogeneity in Data Access Conditions:** Procedures for accessing microdata are highly inconsistent, ranging from open scientific access to highly restricted administrative access. For the three administrative datasets (EL, ES, PT), the access process required eight to twelve months of negotiation.
- **The Household Information Gap:** While administrative registers (Greece, Portugal, Catalonia) are excellent for coverage and tracking, they often lack detailed household information such as

parental occupation, income, or home environment variables. Conversely, survey data (like the UK's MCS) have rich household data but are often weaker on school-functioning variables.

- **Measurement Heterogeneity:** National assessments are independently developed to serve national priorities. A "literacy" score in one country may reflect curriculum-based mastery, while in another, it may measure broader cognitive reasoning. The problem looks less severe for numeracy.
- **Temporal and Structural Misalignment:** Cohorts follow students at different ages and school grades, and wave frequencies do not always overlap. Institutional differences—such as tracking ages, grade repetition rules, and school structures—further complicate the comparison of student trajectories.
- **Inability to Harmonize Complex Factors:** Many critical predictors of achievement, such as student attitudes, motivation, and learning impairments, are measured too heterogeneously across datasets to be meaningfully harmonised.

Main Conclusions and Strategic Directions

The primary conclusion of the feasibility study is that the **construction of a single fully comparable cross-national longitudinal dataset through ex-post harmonisation is currently not feasible** under existing data conditions. Structural differences in data access, measurement, and timing cannot be fully resolved retrospectively without significant loss of analytical validity.

Instead, the study proposes a more viable approach based on coordinated, parallel analyses across harmonised national datasets, enabling meaningful comparison without requiring full data integration. More broadly, the findings highlight the need to **shift from ex-post harmonisation toward stronger coordination at the point of data production**. If Europe aims to generate robust and comparable evidence on educational inequalities, comparability must be built “upstream” rather than reconstructed “downstream”.

The study outlines four strategic options for the future development of European longitudinal data:

1. **Ex-post Harmonisation:** Continuing the current ad-hoc model for specific, bounded research questions.
2. **Incremental Coordination:** Creating a network of national data producers to adopt common metadata standards and international coding (e.g., ISCO, ISCED, Eurostat migration standards) during the original data collection.
3. **Longitudinal Extension of Existing Cross-Sectional Studies:** Converting existing international surveys like PIRLS or TIMSS into longitudinal panels by following the same students through the end of upper secondary education.
4. **A Dedicated European Longitudinal Panel:** Establishing a new research infrastructure for education, like the SHARE project in health and aging, ensuring common sampling, timing, and assessment instruments by design.

While the project demonstrates that meaningful harmonisation is possible for a core set of variables, it also shows that robust **comparative evidence on educational inequalities requires a systemic shift toward coordinated data production** at the European level. The feasibility study provides both the technical foundation and the strategic direction for advancing this agenda.

1. Introduction

The purpose of this feasibility study, conducted within the LINEup project (Longitudinal Data for Inequalities in Education), is to address a critical gap in European educational research: the lack of a unified, comparable, and longitudinal perspective on student achievement. This study represents one of the first systematic attempts to harmonise national longitudinal datasets on educational inequalities across multiple European countries, including contexts that have so far remained underutilised in comparative research.

While international surveys such as PISA and PIRLS provide essential cross-national evidence on student performance at specific points in time, they are cross-sectional by design and do not follow individuals over time. As a result, they offer limited insight into how achievement gaps emerge and evolve across countries. This constrains our ability to fully understand the dynamic nature of educational inequalities.

Against this background, the feasibility study aims to assess whether existing national longitudinal datasets can be harmonised and used to build a comparable cross-national analytical framework. By moving from independent national data sources to a more standardised comparative approach, the study seeks to strengthen the evidence base for EU policymaking and to enable new forms of cross-country analysis.

This feasibility study builds on previous activities carried out within Work Package 3 (WP3) of the LINEup project, which focused on mapping longitudinal datasets². WP3 included three main tasks:

- Identification of longitudinal datasets (Task 3.1)
- Analysis of longitudinal datasets (Task 3.2)
- Feasibility study (Task 3.3)

The results of Task 3.1 and Task 3.2 are presented in Deliverable D3.1, entitled *Mapping Longitudinal Datasets on Educational Inequalities in Primary and Secondary Education: Research Findings and Implications* (delivered in July 2025). The project's initial mapping involved comprehensive desk research covering 32 countries, ensuring that the project considers the diversity of data collection systems across Europe.

Building on this work, the feasibility study (Task 3.3) pursues the following objectives:

- Explore and compare the identified longitudinal datasets.
- Assess the extent to which key variables can be harmonised in terms of content and coding structure.
- Identify potential linking techniques to improve comparability of student achievement.
- Evaluate whether a fully harmonised cross-national dataset can be constructed.
- Develop a harmonisation codebook covering variables available across datasets.
- Build a reference syntax to support the generation of harmonised variables and comparative analyses.

² At the same time, WP3 is strictly linked to WP2 that provided a systematic literature review collecting information on longitudinal research designs and educational inequalities in Europe (see Deliverable 2.1 of the present Horizon Europe Project: "Tracing Educational Inequalities in Primary and Secondary Schools - Insights from a Systematic Review of Longitudinal and Repeated Cross-sectional Studies", February 2026).

Among EU-funded initiatives, the project that comes closest to the present approach is PIONEERED – *Pioneering policies and practices tackling educational inequalities in Europe* (grant agreement No. 101004392). However, the present study differs in two main respects: it focuses specifically on longitudinal datasets covering primary and secondary education, and it includes a different set of data sources (with only NEPS in common).

The feasibility study focuses on eight countries: Germany (DE), Greece (EL), Spain (ES), France (FR), Italy (IT), the Netherlands (NL), Portugal (PT), and the United Kingdom (UK). These countries were selected based on three criteria:

- First, they represent a variety of educational system structures and challenges. For example, Germany provides a model of a stratified secondary school system, while Italy and France represent more integrated school systems. Moreover, Greece and Italy display significant territorial disparities in achievement, which require specific policy attention.
- Second, these countries include both contexts with longitudinal datasets that are already relatively well known in the scientific community (e.g. NEPS for Germany, INVALSI for Italy, DEPP for France, the Netherlands Cohort Study on Education for the Netherlands, and the Millennium Cohort Study for the UK) and contexts with datasets that have so far been scarcely used in international comparative research (e.g. MySchool for Greece, RALC for Spain, and the Portuguese Student Register for Portugal-DGEEC). It must be underlined that getting access to some of these datasets is one relevant achievement of the LINEup project, given the administrative difficulties that researchers encounter when attempting to do so.
- Third, the chosen datasets are in some cases survey studies, while in others are administrative registers. They represent the main two design types of high-quality longitudinal datasets on education, coherently with what emerged from the mapping process. Finally, as just stated above, the included datasets had not been previously studied in other datasets harmonisation attempts, apart from NEPS³.

The harmonisation process includes eight datasets, one per country, encompassing both survey-based studies and administrative registers: NEPS (Starting Cohorts 2 and 3) for Germany, the Greek school register (MySchool) for Greece, the Catalan school register (RALC) for Spain, DEPP (2007 and 2011 Panels) for France, INVALSI-SNV for Italy, the Netherlands Cohort Study on Education (NCO) for the Netherlands, the Portuguese Student Administrative Register (DGEEC) for Portugal, and the Millennium Cohort Study (MCS) for the United Kingdom (see Section 3 for more detailed information on these datasets).

The analysis focuses on primary and secondary education (ages 6–18) and prioritises literacy and numeracy as core competencies. Explanatory variables include individual (student-level), family (household-level), and institutional characteristics (school- and teacher-level). The last set is clearly relevant from a policy perspective, being much more manageable through public and private intervention.

The report is structured as follows:

- Section 2 presents the methodological framework;
- Section 3 describes the datasets and selected cohorts;
- Section 4 discusses harmonisation challenges and solutions;

³ PIONEERED focused on the analysis of longitudinal datasets from the following countries: Germany, Ireland, Hungary, Luxembourg, Norway.

- Section 5 presents the distribution of variables across datasets;
- Section 6 outlines key limitations;
- Section 7 discusses lessons learned and implications.

2. Methodological Framework

The LINEup project investigates whether it is possible to create comparable longitudinal measures of educational achievement and inequalities across seven EU countries and the United Kingdom, selected based on the previous mapping of available datasets presented in Deliverable D3.1. The harmonisation strategy adopted in this study follows the **logic of “minimum harmonisation”**, prioritising comparability across datasets while preserving as much information as possible from the original variables. The feasibility study examines whether such comparability can be achieved using existing longitudinal datasets on students’ learning outcomes in primary and secondary education, collected independently in each country. All datasets have nationwide coverage, except for Spain, where no comparable national dataset is available. We use instead a dataset covering the full student population of Catalonia, one of Spain's largest regions.

In particular, the project focuses on educational achievement, especially the attainment of basic literacy and numeracy skills. Across the datasets considered, educational achievement is measured using different indicators, including standardised test scores and grades, depending on the structure of national assessment systems.

The overarching goal is to strengthen comparative evidence on the factors shaping educational outcomes over time, to understand how inequalities related to students’ ascriptive characteristics—such as sex, social origin, and migration background—evolve across different countries and educational contexts, and to inform policies aimed at improving performance in basic skills and reducing early leaving from education and training, which remains a key challenge across Europe.

Achieving comparability across independently developed longitudinal datasets represents a major challenge, given the substantial heterogeneity of national education systems, the motivations underlying data collection efforts, research designs, and the specific content covered.

The study draws on both survey data and administrative registers. Combining these sources raises specific methodological challenges, including differences in measurement strategies, population coverage, and frequency of data collection. The decision to include both types of datasets in the feasibility study is both important and strategic. On the one hand, it makes it possible to assess the extent to which survey and administrative register data can be harmonised, especially considering that, in most countries, only one of the two types of data is available. On the other hand, the two data sources complement each other in terms of strengths and weaknesses. Stand-alone longitudinal or time-series surveys often collect rich information on students’ and households’ characteristics, but they are typically limited in sample size and duration. Administrative register data, by contrast, may not have been designed to investigate in detail the factors associated with educational performance or inequality, but they cover large populations over longer periods of time.

The present Feasibility Study aims to answer three related questions:

- 1) Is it possible to create a harmonised European longitudinal dataset on educational achievement and social inequalities from originally independent national datasets?

- 2) Which variables, and which underlying phenomena, can be adequately harmonised, and to what extent would the resulting harmonised dataset be informative for the study of educational inequalities?
- 3) What are the costs or trade-offs of adopting such a harmonisation strategy rather than pursuing the alternative path of designing and implementing a single unified longitudinal study that is comparative from the outset?

To address the research questions outlined above, the study adopts a structured methodological approach based on **ex-post (retrospective) harmonisation**. This strategy refers to “a set of procedures applied to already collected data to improve the comparability and inferential equivalence” of measured concepts and variables from different surveys and datasets that are to be integrated into a single dataset (Kołczyńska, 2022). In simple terms, the verb *harmonise* refers to the process of making data coming from different surveys comparable (Lillard, cited in Tomescu-Dubrow et al., 2024). Despite the considerable number of studies aimed at harmonising data in the social sciences, there is no single established approach to retrospective harmonisation: “the terminologies, procedures, technologies and methods adopted vary markedly” (Fortier et al., 2017, p. 103). According to Kołczyńska (2022), retrospective data harmonisation is both theory-informed and data-driven. While theoretical frameworks help identify core concepts and definitions, data availability shapes the choice of variables and their operationalisation. In general, an ex-post data harmonisation project, such as the present one, involves the following stages: concept specification, data collection and preparation, harmonisation, verification, and documentation (cf. Fortier et al., 2017; Słomczyński & Tomescu-Dubrow, 2018).

In this study, we distinguish between **source variables**, that is, the original variables available in each national dataset, and **target variables**, that is, the harmonised variables constructed to enable cross-country comparative analysis.

The harmonisation process was designed to transform selected source variables into analytically coherent and comparable target variables while ensuring full transparency and reproducibility. The primary objective was to establish a structured framework allowing different phenomena to be operationalised consistently across datasets. Source variables were carefully examined and mapped, and the original categories were aggregated into reduced and interpretable classification schemes that remain theoretically meaningful and suitable for statistical analysis.

The LINEup project adopted this general harmonisation flow, adapting it to better reflect the theoretical and practical specificities of educational data. This workflow constitutes the operational framework guiding the harmonisation process implemented in the feasibility study.

The LINEup harmonisation workflow translates general harmonisation principles into a structured operational process tailored to educational data:

1. Theoretical model and concept specification.
2. Criteria for data selection: data and documentation collection
3. Source variables: first (inclusive) selection; item availability template.
4. (Re)Definition of target variables based on cross-national examination of source variables and finalisation of the source variables' list.
5. Crosswalk coding and additional checks of source variables.
6. Quality control of items, documentation, and data.

7. Final revisions of the harmonization and its documentation, including the preparation of a common variables' report.

Following previous attempts to harmonise educational databases, we adopted a **variable-by-variable approach** (i.e. as the one implemented also by the PIONEREED research project; Kroezen & Alieva, 2022).

First, we selected a set of key variables that, according to the existing literature, are associated with educational achievement and related inequalities in Europe. These variables refer to students' characteristics, including both socio-demographic and household-related dimensions, as well as to the school and classroom environment, including information on teachers.

We then identified and followed three key principles to guide the data transformation process:

- Concept validity: using variable definitions that are as similar as possible across datasets.
- Harmonisation transparency: documenting all changes made during the harmonisation process and specifying adaptations for "special cases", including both national and dataset-specific circumstances.
- Minimisation of information loss: where possible, using the same scales across datasets (i.e. variables measured using the same units across datasets), maximising the number of categories retained, and avoiding drastic reduction in the number of values.

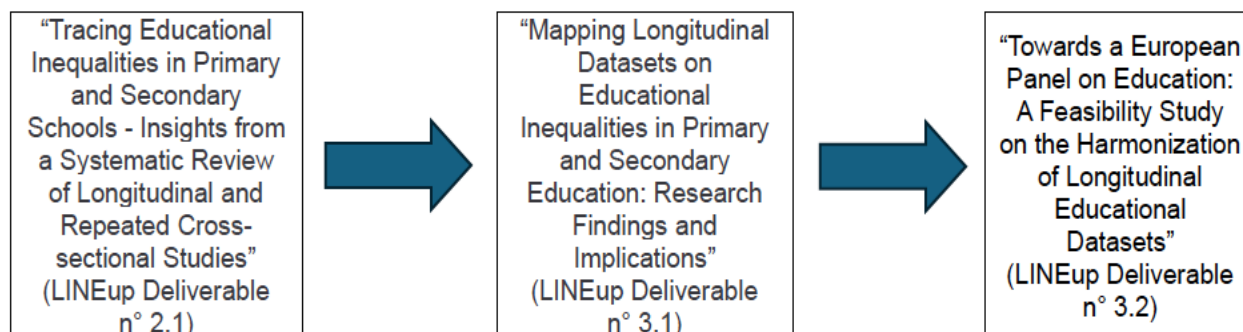
To implement these principles in the retrospective harmonisation process, the documentation accompanying each dataset provided essential support, including codebooks, questionnaires, technical reports, and other relevant sources of information. The research team consulted the available documentation to better describe the source variables. This consultation followed a hierarchical strategy, beginning with the most direct and reliable information source (namely the dataset codebook, when available) and then moving to less direct sources:

1. Analysis of codebooks
2. Analysis of questionnaires
3. Analysis of technical reports
4. Analysis of other available sources (e.g. journal articles or book chapters)
5. Consultation with the team that compiled the dataset and, through them, when possible, information requests to data providers.

After selecting the datasets and analysing the available documentation, we attempted to harmonise as many variables as possible. This phase of the process was also informed by previous tasks carried out within the LINEup project and by their related findings (Figure 1).

Specifically, deliverable D2.1 on the systematic literature review (Kamplis et al., 2025) enabled us to identify the factors, or variables, most frequently associated with educational performance and inequalities in studies based on longitudinal datasets. We used this information to define a grid of factors and variables for collecting information on all the longitudinal datasets from European countries that we were able to identify. Instead, deliverable D3.1 on the mapping of longitudinal datasets (Pavolini et al., 2025) provided information on dataset structure, frequency of data collection, accessibility of microdata, and available variables.

Figure 1: Workflow of the main LINEup deliverables



The selection of each individual variable was based both on: (a) the literature; and (b) their presence across datasets and their operational definitions.

By theoretical support, we mean that we looked for source variables that have been empirically shown to be linked to educational inequalities across countries.

More specifically, the variables were clustered into four core domains:

- Student's individual characteristics;
- Student's household characteristics;
- Teacher- and school-related characteristics;
- Student's academic performance and/or competences.

The present feasibility study follows the same cluster structure and examines how these variables are distributed and operationalized across datasets. The analysis therefore focuses on variables that are represented across the mapped longitudinal datasets, excluding those with only limited availability⁴, to maximise the number of variables that can be harmonised within the four core domains.

Each variable was harmonised according to the same procedure:

1. We assessed the variable's **availability** across the datasets.
2. We discussed the **measurement** of the target variable across different surveys.
3. We created a table with the **original coding** in each dataset.
4. We then proceeded with the **harmonisation** choices, moving from source variables to target variables, ensuring maximum comparability across these datasets.

The full list of selected variables is presented in Table 1 and is organised according to the analytical domains introduced above: students' individual characteristics, household characteristics, teacher- and school-related characteristics, and students' academic performance. To improve readability and clarity, household and school-related characteristics are further divided into more detailed subcategories below (e.g. parents' characteristics and household structure characteristics).

⁴ We considered only those variables that were available in at least one third of the 103 datasets that we mapped and analysed in the LINEup Deliverable 3.1 dedicated to the Mapping of Longitudinal Datasets.

Table 1: List of target variables

Variable name	Variable label
Students' characteristics	
tg_sex	Student's sex/gender indicator
tg_brth_month	Student's month of birth
tg_brth_year	Student's year of birth
tg_brthcntr	Student's country of birth
tg_brth_region	Region of the world where the student was born
tg_stmig	Migrant background of the student
Household: parents-related characteristics	
tg_prnt_sex	Parental sex/gender indicator
tg_mthr_brtcntr	Country of birth - mother
tg_fthr_brtcntr	Country of birth - father
tg_mthr_brtcntr_region	Region of the world where the student's mother was born
tg_fthr_brtcntr_region	Region of the world where the student's father was born
tg_mthr_occ	Occupation status - mother
tg_mthr_empl	Employment status - mother
tg_fthr_occ	Occupation status - father
tg_fthr_empl	Employment status - father
tg_mthr_edu_level	Highest level of education - mother
tg_fthr_edu_level	Highest level of education - father
Household: structural characteristics	
tg_hh_size	Household size
tg_hh_children_minor	Number of children younger than 14 living in the household
tg_hh_income_qt	Household monthly income - quartiles
tg_hh_income_dec	Household equivalised monthly income - deciles
tg_nuts1	Regional location of the household
School characteristics: school level	
tg_schl_type	School ownership, describe funding and governing models, i.e. state, private, public, or other type of governance
tg_schl_level	Level of education, i.e. kindergarten, primary, lower secondary, upper secondary
tg_schl_track	School track
tg_schl_size	Number of registered students in school
tg_schl_mg	Share of migrant students in school
tg_schl_mg_percent	Percent of migrant students in school
tg_schl_gd	Share of female students in school
tg_schl_gd_percent	Percent of female students in school
School characteristics: class level	
tg_cl_size	Number of registered students in class

tg_cl_mg	Share of migrant students in class
tg_cl_mg_percent	Percent of migrant students in class
tg_cl_gd	Share of female students in class
tg_cl_gd_percent	Percent of female students in class
Teacher characteristics	
tg_lang	Subject - literacy
tg_math	Subject - numeracy
tg_teach_year	Teacher's working experience
tg_teach_sex	Teacher's sex/gender indicator
tg_teach_empl	Teacher's employment contract
Students' academic performance	
tg_literacy_tests	Literacy Performance - tests
tg_numeracy_tests	Numeracy Performance - tests
tg_literacy_marks	Literacy Performance - marks
tg_numeracy_marks	Numeracy Performance - marks

A systematic crosswalk coding procedure was implemented to document each step of the transformation process. For each harmonised variable, the original source variable, its coding scheme, and the resulting derived target variable, together with its corresponding categories, were recorded in a master harmonisation file. Each source category was assigned to exactly one harmonised category to ensure mutual exclusivity and avoid ambiguity. At the same time, the harmonised categories were designed to be exhaustive, so that all valid original responses were represented. Special attention was paid to the consistent handling of missing values and non-substantive response categories, both due to pitfalls in data collection and panel attrition.

During the data processing phase, the research team developed a set of syntax templates in Stata 18 (StataCorp LLC, 2023), together with a related codebook. These templates include three main elements: documentation of the variables in the source data; a harmonisation template for each target variable; and relevant information on country-specific differences and/or alternative coding decisions. This enabled each team member to follow the harmonisation process consistently and to apply the templates to their own dataset and source variables. The templates also had a tool for scholars to share comments within the code, facilitating feedback and collaboration among team members.

The implementation of these tools allowed the team to automate the generation of target-variable files and to introduce the adjustments required for “special cases” at both national and dataset level (e.g. variables with highly specific content or coding structures due to country-specific features). Quality control procedures were integrated throughout the harmonisation process. At the level of individual variables, all observed source codes were checked against the crosswalk to ensure complete coverage and to prevent accidental omissions. Coding ranges were verified, and any overlaps or inconsistencies in category assignment were resolved.

At the documentation level, harmonised variables were named according to a consistent convention, and crosswalk tables were uniformly formatted to ensure clarity. At the dataset level, the resulting variables were checked to confirm that they contained only predefined target values and that their distributions were logically coherent.

In the final revision stage, variables were reviewed collectively to ensure structural coherence across socioeconomic domains. Classification schemes were assessed in terms of balance, interpretability, and conceptual alignment, while the overall framework was evaluated to ensure that similar constructs followed a consistent aggregation logic.

This final review ensured that the harmonised datasets function as a coherent analytical system rather than as a collection of independently recoded variables. As a result, researchers can apply the same analytical models across harmonised national datasets to address comparable research questions, while preserving the maximum amount of information from the original data.

3. Data Sources

3.1 The general characteristics of the datasets and the selected cohorts: an overview

This subsection presents the core information for each dataset included in the harmonisation process. For each selected dataset, the following information is provided:

- Name of the dataset and country covered.
- A brief description of its characteristics.
- Objectives of the dataset,
- Type of study (pure longitudinal, repeated cross-sectional, pseudo-panel),
- Sample size,
- Geographical coverage,
- Start and end years of the dataset,
- Educational outcomes included.

As several data sources provide information on multiple cohorts, and in some cases—especially for administrative data sources—a new cohort is available each year, two additional pieces of information are also reported:

- Selected cohorts included in the harmonisation,
- Sample size of the selected dataset or cohort.

Table 2 provides an overview of the main characteristics of the selected datasets, described in detail, one by one, in the present section.

NEPS (DE)

The National Educational Panel Study (NEPS) is the largest long-term educational study in Germany, and it is conducted by the *Leibniz Institute for Educational Trajectories*. Its main objective is to describe and explain competence development, educational processes, and their conditions, outcomes, and consequences across the life course. NEPS is a pure longitudinal study with seven samples of altogether over 120,000 participants, including approximately 70,000 core respondents and 50,000 relevant individuals for pupils, such as parents and educational professionals. The study has followed 7 representative cohorts so far. The study covers the entire country and has been

ongoing since 2009. Educational outcomes are measured through standardized tests and teacher-assigned grades in literacy, numeracy, science, foreign languages, and other competencies.

The selected cohorts for the harmonisation are:

- 1) The Kindergarten Starting Cohort, launched in 2010 and still ongoing, following children from early childhood education into adulthood, hereafter referred to as SC2 (Starting Cohort 2, as it is the second cohort developed by NEPS).
- 2) The Grade 5 Starting Cohort, also launched in 2010, following students from early adolescence (Grade 5) into adulthood, from now on SC3 (third Starting Cohort).

It is important to underline that specific NEPS SC2 / SC3 waves were chosen, not the entire NEPS SC2 / NEPS SC3 datasets (e.g. the first two waves of NEPS SC2 were dedicated to students during kindergarten). The sample size of these two cohorts is around 6,000 students for SC2 and 5,000 students for SC3.

MySchool (EL)

MySchool is a unified information system designed by the Ministry of Education, Religious Affairs and Sports (MoERAS) to support the digital management of school units and educational administrative structures across Greece. Its main objectives are to manage digital data on students, teachers, and schools; to support daily educational processes, such as timetabling, assessment, and communication; and to provide a unified digital environment for primary and secondary education. The system covers the entire student population nationwide and resembles a longitudinal administrative register, tracking students throughout their educational trajectories. Data collection has been ongoing since 2014. Educational outcomes included in the database comprise attendance records (adequacy of attendance, total excused and unexcused absences, and attendance status), grade results in secondary education (term grades, final examination results where applicable, final course grades, annual averages, and re-examination scores), and final annual outcomes (promotion, dismissal, repetition, or referral to re-examination). For upper secondary education, qualification data include General and Vocational Lyceum diplomas, dates of issue, and years of graduation. The database also contains information on student conduct and disciplinary actions, as well as detailed data on teachers and school units.

For the harmonisation process, the chosen cohort follows students in Grades 7-12 from 2019/2020 to 2024/2025. In 2019/2020, the sample was composed of around 100.000, while in 2024/2025 it is composed of around 66.000 students. The reduction in sample size is due to the use of mathematics grades as the dependent variable, since mathematics is not taught to all students in the final year of upper secondary education, but also to the exclusion of certain types of school from the sample and to dropouts, as compulsory education in Greece ends at Grade 9.

RALC (ES – Catalonia)

The Catalan Department of Education records and maintains administrative education data, primarily collected for policy purposes. Their student registry, RALC (Catalan: *Registre d'alumnes de Catalunya*), covers the full student population of the Autonomous Community of Catalonia, approximately 80,000 students per grade. It includes yearly registrations in education and basic background information for each student and therefore acts as an administrative register dataset that can be used to reconstruct longitudinal student trajectories. We combined the information in the RALC with other datasets provided by the Department on school characteristics and test scores.

Other datasets available at the Department on additional test scores and teacher information were considered for the LINEup project but not included as they cover information for publicly funded schools only.

The final grade of students finishing lower secondary education (Grade 10) and academic or vocational upper secondary education (comparable to Grade 12) is recorded. In addition, the dataset includes for all students in the final year of primary education (Grade 6) and lower secondary education (Grade 10) test scores in literacy (Catalan and Spanish), numeracy, science, and foreign languages (English). These standardized assessments are externally graded and serve to monitor overall skill levels.

For the harmonisation process, we focus on the cohort of students who first entered Grade 6, the final year of primary education, in the 2017/2018 academic year. These students are followed for five years, until 2021/2022, by which they are expected to have completed Grade 10. The sample size is equal to 82,276 students.

DEPP Panels (FR)

The DEPP datasets (Panels) are produced by the *Direction de l'évaluation, de la prospective et de la performance* (DEPP) under the French Ministry of Education for national public statistics. Their main objectives are to support policy evaluation and to collect comprehensive data on education. The data follow a pure longitudinal design. DEPP Panels range from the first starting in 1973 to the most recent in 2021. The geographical coverage is nationwide. Educational outcomes consist of standardized tests in literacy and numeracy.

For the harmonisation process, two cohorts were selected.

- 1) The DEPP Panel 2007, following a sample of students from entry into Grade 6 in 2007 through all secondary education and beyond, including higher education and early adulthood.
- 2) The DEPP Panel 2011 follows another sample of 15,200 students who entered Grade 1 (CP) in 2011 and is still ongoing.

The selected dataset includes approximately 35,000 students at baseline for Panel 2007 and 15,200 for Panel 2011.

INVALSI-SNV (IT)

The INVALSI dataset is provided by the *National Institute for the Evaluation of the Education System in Italy* and is designed to monitor student learning outcomes through periodic and systematic assessments of students' knowledge and skills. Although the dataset follows a repeated cross-sectional design, but students can be linked across grades and academic years through a unique identifier. This allows longitudinal student trajectories to be constructed for specific cohorts. The dataset includes around 500,000 students each year. Geographical coverage is nationwide at the population level, and data collection has been ongoing since 2011. Educational outcomes consist of standardised test scores in literacy (Italian), numeracy (mathematics), and foreign languages (English) for students in the following grades: Grade 2 and 5 in primary education, Grade 8 in lower secondary education, and Grade 10 and 13 in upper secondary education. Information on test performances is linked with administrative student information provided by schools. For specific grades (grade 5 and grade 10), a student questionnaire containing more detailed information on students' attitudes, motivation, and study habits is administered.

Three cohorts were selected for the harmonisation process:

- 1) Students followed from Grade 5 (2013/2014) to Grade 13 (2021/2022).
- 2) Students followed from Grade 2 (2015/2016) to Grade 10 (2023/2024).
- 3) Students followed from Grade 2 (2014/2015) to Grade 10 (2022/2023).

These three cohorts were selected because data was not collected in 2020 and 2021 due to the COVID pandemic. Each selected cohort includes approximately 470,000 students per wave.

Netherlands Cohort Study on Education (NL)

The NCO dataset is a collection of administrative education registries from the Netherlands covering multiple cohorts and years. Its main objective is to provide schools, researchers, and policymakers with relevant information for educational practice and national trends monitoring. This study is instigated and financed by the Netherlands Organization for Scientific Research (NWO), specifically by the Dutch National Education Institute (NKO) and accessed to the microdata environment of Statistics Netherlands. The data follow a pure longitudinal design and cover the entire student population nationwide. Data collection has been ongoing since 2007, and since 2010 for primary education. The main NCO dataset includes results from national end-of-primary tests (Grade 6), and final examination scores in upper secondary education (Grades 10, 11, or 12, depending on track) across all subjects. In addition, the NCO-LVS dataset includes for students in about 30-50% of primary schools' biannual assessments in numeracy, reading, and spelling throughout Grade 1 to Grade 6.

For the harmonisation process, the selected cohort follows students who entered primary school in 2013/2014 through the completion of primary education in 2018/2019 (Grade 6). The above-mentioned selected cohort includes 160,000 to 170,000 students, representing about 30–50% of the full population.

Portuguese Student Register - DGEEC (PT)

The Portuguese dataset derives from the information extracted from 3 Portuguese specific databases, each one having specific aims:

- **MISI** – MISI is the Central Platform of data in Education in Portugal, which is owned and monitored by the Direção Geral do Ensino Superior (DGEEC) and collects students' academic grades every year.
- **E360 (Escola 360)** - The E360 is an integrative platform developed by DGEEC, which main objective is to improve the handling and management of the seven main administrative processes related to students, namely: enrolments, renewals, and transfers; management of biographical information; classes; attendance; assessment; certificates; and notifications sent to parents or guardians regarding absences, evaluations, the availability of lesson summaries, and assessment schedules.
- **CIF (Classificações Internas Finais)** - CIF refers to students' final school grades at the end of the academic year, which are recorded and used by DGEEC to monitor educational quality.

When combined, these three administrative sources allow for the construction of a longitudinal dataset that follows students across school years. They cover students from the entire country and

together provide information on students' academic performance in Mathematics and Portuguese Language.

For the present dataset, we extracted specific indicators from E360, MISI and CIF referring to the same students to achieve a true longitudinal design, describing students' grades in Mathematics and Portuguese Language throughout their schooling years.

We defined this specific cohort because it was the most recent one with complete information available for extraction when we began the process with DGEEC (in 2024). At that time, the latest cohort with complete data up to the end of 12th grade was the group that completed 12th grade in the 2021/2022 academic year.

Our aim was to present the longest possible educational trajectory based on the information available from the DGEEC databases (E360, MISI and CIF). For the harmonisation process, we used data from a cohort of approximately 105,000 (one hundred and five thousand) students who were enrolled in 5th grade in the 2014/2015 academic year and followed them until 2021/2022, when they reached Grade 12. This dataset covers, then, 8 school years.

Millennium Cohort Study (UK)

The Millennium Cohort Study is a longitudinal birth cohort study conducted by the University College of London. Its main objective is to follow the lives of children born between 2000 and 2002 in the United Kingdom. The study follows a pure longitudinal design and initially included approximately 19,000 participants. It covers the entire United Kingdom, including England, Wales, Scotland, and Northern Ireland. Data collection began in 2000 and is still ongoing. Educational outcomes are measured through standardized tests in literacy and numeracy, as well as school performance indicators. The only available cohort of the study, relevant for the LINEup harmonisation process, follows about 15,000 children born between 2000 and 2002.

Table 2: Summary of the selected datasets for harmonization. Data in this table refers to the cohorts/waves used for harmonization and do not describe the whole datasets.

Country - Partner	Dataset	Data collection				Content (tests)	Grades	Data level	Access to microdata
		Start End	Frequency	Sample Size (first wave)	Design				
DE - PHKA	NEPS	2011/12-2019/20 (SC2) 2012/13-2022/23 (SC3)	Annual	6.000 (SC2) 5.000 (SC3)	Pure longitudinal	Competency tests, school grades	1-9 (SC2)* 5-13 (SC3)	Individual	Yes
EL - KEPE	MySchool	2019/20-2024/25	Annual	Around 100.000	Pure longitudinal	Attendance records, grade results, final annual outcomes, diploma qualification, conduct	7-12	Individual	Only for KEPE
ES - UPF	RALC	2016/17-2021/22	Annual	80.892	Repeated cross-sectional	Standardized test scores, grades	6-10	Individual	Only for UPF
FR - INED	DEPP	2007/08-2014/15 (Panel 2007) 2011/12-2016/17 (Panel 2011)	Annual	35.000 (2007) 15.200 (2011)	Pure longitudinal	Standardized tests	6-12 (Panel 2007) 1-5 (Panel 2011)	Individual	Yes
IT - FpS	INVALSI	2013/14-2021/22 (W1) 2015/16-2023/24 (W2) 2014/15-2022/23 (W3)	Annual	470.000 x wave	Repeated cross-sectional	Standardized test scores, grades	2-13	Individual	Yes
NL - UPF	NCO	2013/14-2018/19	Biannual	160.000 to 170.000	Repeated cross-sectional	Standardized test scores, final examination scores	1-6	Individual	Only for UPF
PT - UTAD	DGEEC	2014/15-2021/22	Annual	Around 105.000	Pure longitudinal	School grades	5-12	Individual	Only for UTAD
UK - UNIMI	MCS	2008-2018	Every three years	15.000	Pure longitudinal	Competency tests, school results	Age 7-17	Individual	Yes

* Only Grades from 1 to 4 in SC2 were used in the present Feasibility study

3.2 Cohorts selection and access to datasets

As noted in the previous subsection, eight datasets were selected for the feasibility study. For datasets covering multiple cohorts, specific cohorts were selected. In total, **twenty-two cohorts** were included in the analysis:

- Two cohorts from NEPS for DE;
- Eleven cohorts from MySchool for EL;
- One cohort from RALC for Catalonia (ES);
- Two cohorts from DEPP for FR;
- Three cohorts from INVALSI for IT;
- One cohort from NCO for NL;
- One cohort from Portuguese Student Register – DGEEC for PT;
- One cohort from MCS for UK.

After briefly describing the datasets and the selected cohorts, it is important for the purposes of the feasibility study to provide information on access to the microdata for these datasets and cohorts. Describing the access procedures not only offers an overview of how easy or complex—and how time-consuming—the process has been but also makes it possible to assess the actual feasibility of producing a single shared harmonised dataset at the end of the process.

The eight datasets differ with respect to two core aspects of data access:

- a) The relative ease with which external researchers can obtain access to microdata, typically through online requests submitted to the data providers.
- b) The possibility of directly using the microdata and storing them outside the secure environment provided by the data holder.

Based on these two aspects—ease of access and the possibility of directly using the microdata—the datasets can be grouped into three main types of data access:

- 1) **Open scientific access**, where microdata can be obtained easily through standard online procedures.
- 2) **Restricted remote access**, where researchers can access the data only through controlled remote environments.
- 3) **Highly restricted administrative access**, where data use is subject to strict institutional procedures and cannot be managed directly by external researchers.

It should be noted that the high heterogeneity in data sources is a direct consequence of the dataset selection, which reflects both diverse data generation processes and an effort to extend data access beyond current research practices. This has enabled the analysis of information that has so far remained inaccessible.

For each type of access arrangement, it is therefore useful to describe in more detail the procedures required to obtain access to each dataset.

1) Open scientific access datasets. This is the case for:

- **NEPS (DE):** Data from NEPS can be accessed through three methods: download, remote access, and on-site use. These methods vary in terms of data anonymization and data security requirements. Accessing NEPS data requires a Data Use Agreement with the Leibniz Institute for Educational Trajectories (LIfBi). Additional agreements are needed for remote and on-site access. Data access is free of charge. All information regarding the Data Use Agreement and data access instructions are available on the LIfBi website.
- **DEPP (FR):** DEPP data are accessible through the Quetelet-Progedo Diffusion platform, following the standard application and authorisation procedures. Access is granted upon registration and submission of a detailed research proposal, together with the required institutional documentation and confidentiality agreement. The DEPP datasets provide an intermediate level of detail between standard aggregated datasets and highly confidential microdata. Consequently, access is subject to formal approval by INED and the Statistical Secrecy Committee (CSS). Authorised users may access the full range of INED FPR datasets. However, the level of detail and variable availability may vary depending on data sensitivity and disclosure risk, and some highly detailed information remains accessible only under additional restricted conditions.
- **INVALSI (IT):** While some INVALSI data are freely accessible through the website in the form of open aggregated statistics, access to anonymised microdata is restricted and only available upon request. Researchers must register with an official institutional affiliation, and accounts are limited to universities or recognised research institutes. Access to microdata requires the following steps: 1) submission of a formal data request, specifying the research title and description, the datasets needed (years/grades/subjects/), the variables required and the intended research outputs; 2) A data usage agreement signed by all researchers involved in the project; 3) A data usage agreement signed by the institution's DPO certifying compliance with data protection regulations. Once the request is approved, the designated researcher receives a download link providing access to the microdata. The package includes documentations such as codebooks, technical reports, and the datasets in multiple formats.
- **MCS (UK):** Most of the study data are available through public repositories, primarily via the UK Data Service. Interested researchers are required to register with the UK Data Service and specify the research project for which access is requested. The remaining data, classified as having medium to high sensitivity and disclosure risk, are accessible only to researchers involved in UK-based research projects approved by the UKDS/CLS, affiliated with a recognized institution or organization, as well as to CLS data managers and approved UK researchers, depending on the level of data sensitivity.

2) Restricted remote access datasets. This is the case for:

- **NCO (NL):** In this case, as it happens for other datasets in related social science fields (e.g. the Luxemburg Income Study and the remote access datasets of the NEPS), it is not possible to download directly the dataset on the researcher's own computer. Instead, the dataset is hosted on the microdata remote access environment of Statistics Netherlands. Acquiring access rights involves a complex procedure, starting with an institutional agreement between Statistics Netherlands and the organization(s) of the involved researcher(s). Additionally, a project agreement specifies the relevant datasets for the project and related costs. In case of access to the supplementary NCO-LVS dataset, additional approval is required from the Dutch National Education Institute (NKO) that coordinates the dataset. In this feasibility report

we focus on the NCO and NCO-LVS datasets, but it is possible to link these datasets to any of the hundreds of other datasets at Statistics Netherlands containing microdata on a wide range of topics. The transfer of results from the remote environment to the researchers only takes place after an export check on disclosure risks by Statistics Netherlands. Costs are charged at the start-up phase of a project, for each month of data access and for any export request. Documentation of the NCO dataset and all general procedures are provided both in Dutch and English. To arrange access to the data for the LINEup project, the team at UPF spent about two months to arrange the institutional agreement. The project agreement took another month to be finalized, after which access to the data was provided. It was possible to complete all forms and communication in English.

3) Highly restricted administrative access datasets. This is the case for:

- MySchool (EL):** MySchool is a nationwide student registry administered by the Ministry of Education, Religious Affairs and Sports (MoERAS). MySchool primarily contains information on students' academic performance and some of their basic characteristics (e.g., gender, age, absences, place of birth, place of residence, etc.) as well as class/school characteristics (size, gender composition, migration background composition, average grades, etc.). To obtain access to students' data it is mandatory to submit a formal request to the competent department of the MoERAS, as well as to the Minister's Office. The request must clearly specify: the exact data and variables required from MySchool; and the intended use of these data (research purposes, methodology, outputs). Approval from the Ministry's legal services is essential before any data can be released, due to strict personal data protection regulations, particularly because the data concern minors. Certain variables may be restricted, anonymised, aggregated, or deemed impossible to release due to legal, technical, or data reliability constraints. There is no financial cost associated with obtaining access to the data. The timeline for accessing the data depends on administrative processing and legal review. For the LINEup project our research team has made an official request to the MoERAS. The process required ex ante careful planning regarding the variables, students, schools, and school years to be included in the dataset, in close cooperation with the MoERAS team to extract the desirable information. By itself this is a challenging task since there is no manual or list of the available information contained in MySchool. Once the dataset is ready and received, additional time and effort is required to bring the data to a usable format ready for quantitative analysis. In the LINEup case, beyond defining the sampling framework, a detailed and exhaustive list of all required variables was provided. Several meetings were also held with the MySchool team to determine what could be legally shared; what was technically feasible; what met acceptable data quality standards; where constraints existed, alternative solutions were proposed and jointly agreed upon. Despite the excellent cooperation and continuous support from Ministry officials, the data provided are raw administrative data and require substantial processing before becoming ready for research. Overall, the construction of an analytically usable, representative panel dataset required extensive methodological design, repeated technical consultation, and significant data management effort and time. In principle, it would be possible to link MySchool data with other administrative sources, such as social security or tax records. This could provide information on family composition, household income, and the occupations of both parents, which are not available in MySchool. However, this process is time-consuming and requires strict protocols to ensure compliance with personal data protection legislation. Moreover, its outcome is uncertain, as it would involve coordination across three different statistical agencies.

- **RALC (ES - Catalonia):** The administrative data on Catalan students can be obtained through the Catalan Department of Education (Department d'Educació i Formació Professional, Generalitat de Catalunya). A data catalogue is available with a brief description of the different datasets and names of the included variables. Researchers can use a standardized form asking access for information on the involved researchers, a project description, the requested data, and data protection measures. A meeting is organized with the Department officers to discuss the feasibility of the request. After the meeting, the agreed-upon datasets are prepared and shared with the researcher after signing a confidentiality agreement. The time between filling the application and receiving the data depends on the number of concurrent requests that the Department of Education is handling at the time and can take several months (i.e. eight months for data requested for the LINEup project). Communication with the Department takes place over email, and it involves several subdepartments. Working language is Catalan, and the dataset is only available in this language. Because of privacy regulations, researchers need to check with the Department before making their findings public. No costs are charged for using the data.

In principle, it should be possible to link the microdata on education in the RALC with microdata held at the Statistical Institute of Catalonia. In particular, the absence of parental and household characteristics in the RALC could be addressed by integrating information from other administrative sources within the Catalan administration. As part of this feasibility study, the UPF team submitted a request for such a data linkage. However, the process has proven to be time-consuming, and no data has been received twelve months after the initial request. Throughout this period the Statistical Institute of Catalonia has communicated that they must first complete other requests before they can attend to the request of LINEup.

- **Portuguese Student Register (DGEEC) (PT):** The process of accessing datasets from the Portuguese national organisms (DGEEC and IAVE) is complex and can only be conducted in loco (at the DGEEC facilities), within their safe-zone system. Because the information includes student data (including grades and potential identification risk), it is considered sensitive information or personal data, which is regulated by law. Therefore, the national organisms DGEEC and IAVE are the legal guardians of this information, which is why they maintain clear and strict regulations and procedures that aim to reconcile the right and value of access to information with students' right to privacy. Researchers interested in accessing the DGEEC and IAVE databases must contact DGEEC (Direção-Geral de Estatísticas da Educação e Ciência) and explain their objectives and submit a specific request (including objectives, etc.). Requests for access to the datasets should include a brief description identifying the responsible researchers for the request (including institutional affiliation, etc.), the objective of the request, and the intended use of the data. These and other details (including specific terms and conditions) will be specified in a Memorandum of Agreement between the parties. There are no payments attached to data access. However, interested researchers should be aware that, because of the sensitive nature of the information and the presence of personal data, access to the databases is only allowed in loco, at the DGEEC Safe Center, and under the terms and conditions defined by DGEEC.

Overall, it should be emphasised that a key achievement of the present feasibility study—and of the LINEup project more broadly—is that, for the first time, external researchers were able to access, directly or indirectly, the microdata from these three datasets for comparative research on educational inequalities. In all cases, access required between eight and twelve months and, in some instances, remained subject to restrictions and limitations, as is common—albeit to varying degrees—across all datasets.

Beyond the challenges encountered and the time required to obtain access, two key points should be highlighted. First, the harmonisation process proved to be more iterative and complex in practice than initially anticipated. Second, the original plan to construct a single, integrated cross-country dataset for joint analysis was found to be unfeasible.

To ensure the inclusion of all selected countries in the comparative analyses, a more realistic approach is to apply the same analytical models to harmonised national datasets stored separately, and then compare the resulting estimates across countries using a common syntax.

4. Harmonised Variables Definitions: Main Challenges and Solutions⁵

As outlined in Section 2, the selection of variables to be harmonised across datasets was informed by the findings of the mapping report produced during the previous phase of the LINEup project (Pavolini et al., 2025).

Tables 3 and 4 summarise the results of the harmonisation process, presenting respectively the variables that could be harmonised and those that could not. Overall, a substantial set of variables (34) was successfully harmonised (Table 3) through the construction of common target variables derived from the original source variables in each dataset, following the crosswalk procedures described in Section 2. However, as discussed in Sections 5 and 6, their availability and quality (e.g. in terms of missing values) vary across datasets and cohorts.

The harmonised variables cover all four analytical domains considered in the feasibility study: students' individual characteristics, household characteristics, school and teacher variables, and academic performance.

Table 3: Overview of harmonised variables across datasets

Student's individual variables	Student's household variables	School-related variables	Teacher-related variables	Student's academic performance
Sex	Parents' sex	School ownership	Subject - literacy	Literacy Performance - tests
Birth year	Mother/Father birth country	Education level	Subject - numeracy	Numeracy Performance - tests
Birth month	Mother/Father birth region	School track	Teacher experience	Literacy Performance - marks
Country of birth	Mother/Father employment	Share of students with migrant background (school)	Teacher sex	Numeracy Performance - marks
Birth region	Mother/Father occupation	Share of female students (school)	Employment contract	

⁵ The present Section reports the structure of the homogenisation coding book. The codebook, as well as the STATA syntax ("Do" files), are available upon requests.

Migration background	Mother/Father education level	Class size		
	Household size	Share of students with migrant background (class)		
	Number of children	Share of female students (class)		
	Household income quartile			
	Household income decile			
	Household NUTS1 region			

A further set of variables (Table 4), related to students' individual and household characteristics, could not be harmonised. This is due to two main reasons:

- The high degree of heterogeneity in how these variables are defined and measured across datasets (e.g. items capturing attitudes or motivation towards education are present in many datasets but differ in their operationalisation).
- The absence of relevant information in many datasets (e.g. variables on scholarships, student allowances, or physical and learning impairments).

Table 4: Overview of variables that could not be harmonised.

Student's individual variables	Student's household variables	School-related variables	Teacher-related variables
Attitude or motivation toward education/school	Siblings	Geo-referencing of the school	Teachers' educational qualification
Specific birth country	Stepparents		
Learning impairments	Parents' age		
Physical impairments	N° of books in the household		
Belonging to a recognised ethnic minority	Parental level of knowledge of host country's language		
Receiving a scholarship/ student's allowance	N° of digital devices in the household		
Province/County of residence	Ownership of the apartment/house they live in		
Town of residence			

Two additional variables were examined but proved particularly challenging to harmonise due to their strong dependence on national contexts and data availability.

The first concerns students' participation in early childhood education and care (ECEC). While highly relevant for educational trajectories and inequalities, its harmonisation is hindered by substantial cross-country variation in ECEC systems and by differences in how the information is collected (e.g. parental reports, administrative records, or survey questions).

The second variable is grade retention. Comparability is affected by differences in institutional rules governing grade progression and repetition, as well as by variation in the age at which students enter specific grades (e.g. due to anticipatory enrolment or migration-related pathways). A solution is to use students' age as a proxy for irregular educational trajectories, classifying them according to their deviation from the modal age within grade (regular, advanced, or delayed).

The remaining part of this section presents the harmonisation procedures adopted for each variable, detailing the construction of target variables and the crosswalk rules applied across datasets.

4.1 Individual and Household Variables

Student's individual characteristics

Among the many possible students' individual characteristics, the harmonisation process focuses on those for which there is well-established empirical evidence in the literature, namely ascriptive structural characteristics, as also highlighted in the LINEup literature review (Kampylis et al., 2024). It should be noted that, beyond this main body of literature, several studies have also examined students' attitudes, motivations, and indicators of school experience. However, these were not included in the harmonisation process because the preliminary comparison of datasets showed that they were only rarely collected and, more importantly for the purposes of this study, were highly heterogeneous in terms of definitions and data collection instruments.

Student's Sex

Sex is a common variable in educational inequality research and a potentially important indicator of inequality in education. All the selected datasets include the sex of each student, making harmonisation easy and straightforward. Most surveys report sex based on biological and physiological characteristics, meaning two binary categories of male and female.

Some variation in the reporting of sex could exist between surveys that allow self-completion and surveys that instruct the interviewer to provide the answer without asking the interviewee. Overall, the harmonisation of the sex variable is straightforward, and we follow the standardization proposed by the European Commission (2020) where sex "refers to the biological and physiological characteristics that define a person to be either male or female". Using this standardized variable allows for a comprehensive harmonisation across all surveys and facilitates comparisons across studies.

We adopted the following harmonisation for sex: *tg_sex* - Student's sex (0: Male; 1: Female).

Student's age

The main challenge in harmonising the student's age variable is related to the difficulty to attribute a student to a certain school grade as in different countries children start school at different ages and they can repeat the same grade several times, making school grade not a reliable variable indicating students' age. To solve this problem, we chose to harmonise two variables—students' birth year and birth month—that can later be used to calculate students' age if needed. We adopted the following harmonisation respectively for the student's birth year and month variables: *tg_brth_year* - Student

birth year (*Valid calendar year, example: 2000*); and *tg_brth_month* - Student birth month (*Valid Western calendar months (January–December)*).

Student's country of birth

The country of birth of the student allows first-generation migrants to be identified and is a key factor to be investigated, when studying educational inequalities. However, this variable varies significantly across datasets. While some surveys report exact countries of birth, others collect information on student ethnicity, and others report a simple yes/no variable for those born in the country of the survey. Finally, the Netherlands Cohort Study is a special case, given that if a student is born abroad to Dutch parents, the student is classified as Dutch/Born in a survey country.

To keep comparability as homogeneous as possible across datasets, considering that this inequality dimension was too important to be dismissed, we adopted a simplified categorical variable that identifies those who are native-born in the survey country - and those who are not.

We adopted the following harmonisation for the student's country of birth: *tg_brthcntr* - Student's country of birth (*1: Born in the survey country; 2: Born in a different country*).

Region of the world where the student was born

This variable captures the geographical origin of the student, operationalised as broad regional groupings rather than specific countries. It distinguishes between students born in the survey country (native-born), those born in Western/EU countries, and those born in non-Western countries. This variable serves as a proxy for migration background and potential sociocultural distance from the host society. This grouping is used as an analytical proxy for broad migration regions and does not imply a normative classification.

The operational definition of "Western/EU" and "non-Western" follows a harmonised framework. The Western/EU category includes all European Union member states and selected Western countries (e.g., Andorra, Iceland, Liechtenstein, Monaco, Norway, San Marino, Switzerland, United Kingdom, Vatican State), as well as the United States, Canada, Australia, New Zealand, Japan, South Korea, and Taiwan. The non-Western category encompasses countries in Eastern and Southeastern Europe not included in the Western grouping (e.g., Albania, Bosnia and Herzegovina, Belarus, Kosovo, Moldova, Montenegro, North Macedonia, Russia, Serbia, Ukraine, Turkey), all countries in Africa, South and Central America and the Caribbean, most Asian countries (except Japan, South Korea, and Taiwan), and Oceania (except Australia and New Zealand), as well as stateless individuals.

A country-specific exception applies for the Netherlands, where Western origin additionally includes Europe (excluding Turkey), Oceania, Indonesia, and Japan but not South Korea and Taiwan, reflecting national statistical classification standards.

In Greece, the "native" category includes both Greeks and Cypriots, reflecting their common language and identical rules of participation in the Greek educational system.

We adopted the following harmonisation for the student's birth region: *tg_brth_region* – Region of world where student was born (*1: Native (born in the survey country); 2: Born in a Western/EU country; 3: Born in a non-Western country*).

Student's migration background

The variable on the student's migration background classifies students according not only to their own country of birth, but also to that of their parents. This harmonised measure supports analyses

of educational inequality, integration processes, language background, and social stratification. Distinguishing between first- and second-generation migrants is analytically relevant, as these groups may differ in terms of language proficiency, educational pathways, institutional familiarity, and socioeconomic outcomes. This is an oversimplification of migratory pathways and biographies, nonetheless by incorporating parental country of birth, the variable accounts for students who may not have migrated themselves but may experience social, linguistic, or structural conditions associated with a family migration history.

The harmonisation scheme distinguishes between students with no migrant background, first-generation migrants, and second-generation migrants. The harmonised categories are defined as follows:

- No migrant background: The student and both parents were born in the country of study.
- First-generation migrant: The student was born abroad, regardless of the parents' country of birth.
- Second-generation migrant: The student was born in the country of study, but at least one parent was born abroad.

We adopted the following harmonisation for the student's migration background: *tg_stmig* - Harmonised migrant background of the student (1: No migrant background; 2: 1st generation migrant; 3: 2nd generation migrant).

Student's household characteristics

Parent sex/gender indicator

This variable, which is not as widely investigated as the other ones in previous literature, was harmonised to enrich the set of student's household features. It captures the reported or administratively recorded sex of the parent or legal guardian. It is harmonised into a binary classification to ensure comparability across datasets.

In the subsequent sections, harmonised variables are labelled as "mother" and "father". These labels are retained for consistency across datasets, even when the original dataset records parents as parent_1 and parent_2. It is important to bear in mind as a special case the British "Millennium Cohort" dataset. In such dataset, parents are not reported for gender but by the labels "parent_1" and "parent_2". Hence parent_1 is the "main parent", the one answering the survey, and parent_2 is the "partner", the one not directly answering the survey.

We adopted the following harmonisation for parents' sex: *tg_prnt_sex* - Parental sex/gender indicator (0: Male; 1: Female).

Country of Birth – Mother and Father

Similarly to the variable on the students' country of birth, parent migration history is an indicator allowing to identify individuals with migration background. Not all datasets include information about the parent's country of birth, those which include the information, differ significantly in the level of detail. While some surveys report exact countries of birth, others provide multiple categories with the options for the most common countries of birth among their residents and others report a simple dummy variable for those born in the country of the survey. We adopted the following harmonisation for parent's country of birth:

tg_mthr_brthcntr - Country of birth - mother (1: *Born in a survey country*; 2: *Born in a different country*).

tg_fthr_brthcntr - Country of birth - father (1: *Born in a survey country*; 2: *Born in a different country*).

Region of birth – Mother and Father

The harmonisation of the variable on the parents' region of birth followed the same logic applied to the student's region of birth. It extends the binary birthplace indicator by incorporating broader geopolitical groupings frequently used in migration research. Not all datasets include information about the parent's country of birth, and those which include differ significantly in detail. We adopted the following harmonisation for parent's region of birth:

tg_mthr_brtcntr_region - Region of the world where the student's mother was born (1: *Native*; 2: *Born in a Western/EU country*; 3: *Born in a non-Western country*).

tg_fthr_brtcntr_region - Region of the world where the student's father was born (1: *Native*; 2: *Born in a Western/EU country*; 3: *Born in a non-Western country*).

Employment - Mother and Father

This variable captures the labour market status of students' parents (in cases where they are missing – stepparents or legal guardians). It distinguishes between employed, self-employed, unemployed, and inactive individuals. Employment status is a key socioeconomic indicator linked to household resources and educational outcomes.

At the same time, it has been one of the most complicated variables to harmonise given the heterogeneity in the definitions adopted in the datasets. Especially two of them present national specificities that are particularly challenging and deserve to be reported. In the Dutch NCO dataset, the "unemployed" category includes individuals receiving unemployment benefits and those receiving social welfare (if they are not classified as pensioner, student, or receiving disability benefits). In the Portuguese Student Register-DGEEC datasets, it is possible only to distinguish between unemployed/inactive or employed. We adopted the following harmonisation for the parent's employment status:

tg_mthr_empl - Employment status - mother (0: *Unemployed*; 1: *Employed*; 2: *Self-employed (own-account or employer)*; 3: *Inactive (students, pensioners, homemakers, permanently disabled, etc.)*).

tg_fthr_empl - Employment status - father (0: *Unemployed*; 1: *Employed*; 2: *Self-employed (own-account or employer)*; 3: *Inactive (students, pensioners, homemakers, permanently disabled, etc.)*).

In the Portuguese Student Register-DGEEC datasets it is only possible to apply the following codification scheme - 0: Excluded from labour force; 1: Worker; 2: Other. This implies that, for analyses including Portugal, it would be necessary to reduce the harmonised variable to the more synthetic version, allowing comparison.

Occupation – Mother and Father

Another important variable that we have included in the harmonisation is occupation. Occupation differs from employment (that refers to a person's current labour market status, that is, whether and how they participate in paid work) as it reflects the type and skill level of the job performed, usually classified into categories reflecting task content, required qualifications, and social status.

Not surprisingly, the level of detail recorded in occupation differs among the datasets. In some cases, values offer a detailed occupation variable by grouping jobs based on the similarity of the task and

duties undertaken (European Commission 2020). In other datasets, broad categories of occupations are used instead of the in-depth codes for each occupation title (e.g. ISCO).

We have adopted the following harmonisation of the parent's occupation, recalling the most relevant cleavages reflecting the main occupational cleavages commonly used in research on social stratification, and a dimension crucial in studies on educational inequalities:

tg_mthr_occ - Occupation status - mother (0: Unemployed; 1: Manual skilled and unskilled workers; 2: Manual self-employed; 3: White collar workers; 4: Lower professionals; 5: Higher professionals and employers).

tg_fthr_occ - Occupation status - father (0: Unemployed; 1: Manual skilled and unskilled workers; 2: Manual self-employed; 3: White collar workers; 4: Lower professionals; 5: Higher professionals and employers).

The INVALSI dataset for Italy does not allow a specific value for lower professionals. For this dataset, the only option available is to aggregate such occupational group with "higher professionals and employers".

Education – Mother and Father

Parental education is a key variable in educational inequality research. This variable is a core social background variable in standardized surveys and is commonly standardised internationally using the International Standard Classification of Education (ISCED).

Luckily, most of datasets in our sample record parental education levels using some categorization of the ISCED levels. In the German NEPS, the parents themselves report their highest level of education, in the Italian INVALSI, it is school's administrative offices reporting it, while in the other datasets it is the students who provide the information on behalf of their parents. This is clearly a source of heterogeneity in the measurement of this variable, which should be considered when interpreting results from different countries.

We adopted the following harmonisation: *tg_mthr_edu_level* - Highest level of education – mother (0: Up to primary school; 1: Lower secondary school; 2: VET, 3: Upper secondary school; 4: Tertiary education).

tg_fthr_edu_level - Highest level of education – father (0: Up to primary school; 1: Lower secondary school; 2: VET, 3: Upper secondary school; 4: Tertiary education).

Household size

To better capture students' family environment, we recorded and harmonised the variable reflecting the size of the household. In most countries, household size corresponds directly to the number of individuals reported as living in the same dwelling.

We adopted the following harmonisation: *tg_hh_size* – Household size, which takes a valid value from 1 to 23 household members, corresponding to the maximum observed household size across datasets.

Number of Children Younger than 14 in the Household

Another variable related to the student's household is the number of minors within the household. We harmonised a variable accounting for the number of children under the age of 14 residing in the respondent's household.

The variable includes values from 1 to 8.

We have adopted the following harmonisation: *tg_hh_children_minor* - Number of children younger than 14 living in the household.

Household Income

We have harmonised the student's household income, acknowledging that could represent a synthetic and comparable measure of family socioeconomic resources. This type of information is traditionally difficult to collect and harmonise among different datasets for several reasons mostly related to either data availability or significant differences in the reported units of measurement and categorization of income groups. Not surprisingly, only few selected datasets contain information about household income: NEPS, DEPP, MCS, and NCO. Yet even datasets that include income indicators collect such information in significantly different ways. Based on available information we adopted two variables measuring respectively income in deciles and quartiles.

The variable *tg_hh_income_qt* represents the total net household income, equivalised, and categorised into national quartiles within each dataset and year. The measure includes income from all household members and comprises wages and salaries (net of taxes and contributions where available), 75% of positive farm income, 75% of business income from unincorporated enterprises, reported net earnings of the self-employed, and net income of farm or business owner-operators. Income measure was ranked within each country-year and divided into quartiles (Q1–Q4), thereby accounting for national income distributions and preserving relative socioeconomic positioning within countries.

The final harmonised variable is categorical, with four ordered categories: Q1 (lowest quartile) to Q4 (highest quartile). Standardised codes for non-response (997 = "Refused"; 998 = "Don't know"; . = system missing) were applied uniformly and it is necessary to underline that the amount of missing data is not negligible.

Although labelled as "equivalised," the documentation specifies that no additional adjustment for household composition and size was applied at the harmonisation stage. Consequently, the measure reflects income ranking rather than fully standardised equivalised income across countries.

The variable *tg_hh_income_dec* extends the quartile measure by categorising harmonised household equivalised monthly income into national deciles (D1–D10) within each country and year. The income components included mirror those used for the quartile variable: wages and salaries, adjusted farm and business income (75%), and net self-employment income. Harmonisation followed the same procedure as described for quartiles.

As with the quartile variable, the documentation notes that no further adjustment for household composition and size was implemented during harmonisation.

We adopted the following harmonisation: *tg_hh_income_qt* - Household equivalized monthly income - quartiles (1: *Lowest quartile*; 2; 3; 4; 5: *Highest income quartile*).

tg_hh_income_dec - Household equivalized monthly income - deciles (1: *Lowest income decile*; 2; 3; 4; 5; 6; 7; 8; 9; 10: *Highest income decile*).

Household Region

The variable identifies the regional location of the respondent's household based on the 2018 Eurostat NUTS1 (Nomenclature of Territorial Units for Statistics) classification. Harmonisation involved mapping national regional identifiers to the corresponding NUTS1 categories to ensure consistency across countries. The harmonised regional variable is categorical and country-specific in its structure, while adhering to a common NUTS1-based framework where possible. These documented deviations ensure transparency and inform appropriate analytical decisions in cross-

national research. We adopted the following harmonisation: *tg_nuts1* - Regional location of the household (see Table 5).

Table 5: Regional location of the household

Value	Label
DE1	Baden-Württemberg
DE2	Bayern
DE3	Berlin
DE4	Brandenburg
DE5	Bremen
DE6	Hamburg
DE7	Hessen
DE8	Mecklenburg-Vorpommern
DE9	Niedersachsen
DEA	Nordrhein-Westfalen
DEB	Rheinland-Pfalz
DEC	Saarland
DED	Sachsen
DEE	Sachsen-Anhalt
DEF	Schleswig-Holstein
DEG	Thüringen
EL3	Attiki
EL4	Nisia Aigaiou, Kriti
EL5	Voreia Ellada
EL6	Kentriki Ellada
ES5	Catalonia
FR1	Île-de-France
FRB	Centre-val De Loire
FRC	Bourgogne-Franche-Comté
FRD	Normandie
FRE	Nord-Pas-de-Calais Picardie
FRF	Alsace- Champagne-Ardenne-Lorraine
FRG	Pays-de-la-Loire
FRH	Bretagne
FRI	Aquitaine-Limousin-Poitou-Charentes
FRJ	Languedoc-Roussillon-midi-Pyrénées
FRK	Auvergne-Rhône-Alpes
FRL	Provence-Alpes-Côte D’Azur
FRM	Corse

FRY	Rup Fr - Régions Ultrapériphériques Françaises
ITC	Nord-Ovest (IT)
ITF	Sud (IT)
ITG	Isole (IT)
ITH	Nord-Est (IT)
ITI	Centro (IT)
NL1	Noord-Nederland
NL2	Oost-Nederland
NL3	West-Nederland
NL4	Zuid-Nederland
PT1	Continente
PT2	Região Autónoma dos Açores
PT3	Região Autónoma da Madeira
UKC	North East (England)
UKD	North West (England)
UKF	East Midlands (England)
UKG	West Midlands
UKH	East of England
UKI	London
UKJ	South East (England)
UKK	South West (England)
UKL	Wales
UKM	Scotland
UKN	Northern Ireland

4.2 School and Teacher Variables

While it is well established that the main sources of educational inequalities are linked to ascriptive variables such as those described in the previous paragraph, it is also important to underline that collecting variables related to school functioning, including information on teachers, is essential for understanding the institutional context in which learning takes place. Research linking individual outcomes to educational contexts and their functioning is less extensive than research focusing on individual predictors of educational outcomes and related inequalities. Moreover, studies examining the relationship between schools, teachers, and students' longitudinal development in educational achievement have been conducted in Anglo-Saxon countries.

While students' characteristics explain individual variation, schools shape the structural and compositional environment that conditions educational opportunities. Without school-level information, analyses risk attributing differences in achievement solely to individual factors, thereby overlooking the broader institutional framework and the extent to which schools and teachers may influence pupils' performance independently of their socio-demographic characteristics, and thus partly modify educational inequalities.

In particular, the United States has a well-established tradition of studies focusing on school and teacher value added (Rivkin et al., 2005). Although significant definitional and methodological difficulties remain in making causal inferences at the school and teacher levels, this strand of research points, for a long time, to a substantial influence of school functioning on educational outcomes (OECD, 2005; Hattie, 2012). Even when the analysis does not aim to estimate value-added effects, information on schools and teachers linked to student-level data enriches our understanding of educational performance and related inequalities. Moreover, factors acting at school and teachers' level, as stated above, are those more easily addressed by policies.

School variables capture structural dimensions such as governance (public, private, or other), size, and tracking. These characteristics reflect differences in funding models, organisational autonomy, curricular pathways, and stratification mechanisms. For example, academic versus vocational tracking structures may influence peer composition, expectations, and instructional focus. Including harmonised school variables allows researchers to analyse how institutional arrangements contribute to inequalities or performance differences across systems.

In addition, compositional indicators such as the share of migrant students, gender composition, and the socio-economic composition of the school are crucial for examining contextual effects. Students' outcomes are influenced not only by their personal background but also by the social and demographic composition of the school they attend. Without harmonised measures of school (and class) composition, it is difficult to analyse peer effects or contextual disadvantage consistently across datasets.

Because national education systems differ, we sought to harmonise school variables at a conceptual level rather than replicate country-specific categories. Broad but analytically meaningful classifications were developed to maximise comparability across datasets while preserving substantive information. Where possible, continuous measures were retained to allow flexible modelling. This harmonisation effort enables multilevel analyses that integrate institutional structure, school (and class) composition, and individual outcomes within a coherent comparative framework.

Similarly, collecting information on teachers in datasets on school achievement is analytically fundamental. Educational outcomes are produced within a structured institutional environment in which teachers act as the primary agents translating curriculum, school policies, and resources into actual learning processes. Without relevant teacher-level information, the analysis of student achievement risks oversimplifying the way educational production operates and may fail to capture the extent to which educational inequalities are reduced or reinforced by the school system.

This last statement is supported by three arguments.

- First, teachers are central to the mechanisms through which learning occurs. Teaching practices, classroom management, pedagogical expertise, and subject knowledge are all conveyed by teachers. Even when direct measures of instructional quality are unavailable, teacher characteristics such as experience, employment status, and subject taught can function as important proxies. For example, years of teaching experience often correlate with classroom effectiveness, familiarity with the curriculum, and the ability to manage heterogeneous classrooms. Similarly, knowing whether a teacher of mathematics or literacy holds a qualification in the relevant field is important when analysing achievement in those subjects. Without such information, analyses implicitly assume that instructional input is constant across contexts and across students with different backgrounds, which is rarely the case.

- Second, collecting teacher information may reduce omitted variable bias in statistical models. Teacher characteristics are often correlated with school composition and institutional characteristics. Schools with higher proportions of disadvantaged or migrant students may experience higher staff turnover, greater reliance on temporary contracts, or a less experienced teaching workforce. If teacher variables are excluded from the models, estimated effects attributed to school composition (e.g. migrant share) may partly reflect differences in the characteristics of the teaching workforce rather than peer or institutional effects. Including teacher-level data can therefore help researchers disentangle whether achievement differences are associated with student composition, institutional structures, or variation in teacher resources.
- Third, teacher information enhances comparability across datasets and countries. Differences between institutional systems may mask similar underlying mechanisms related to teacher allocation and experience (or the other way around). Harmonised teacher variables create a common analytical dimension that complements school- and class-level context variables and may contribute to a better understanding of how educational inequalities develop over students' educational careers.

As will be discussed in the following sections, and as already observed in the LINEup mapping report (Pavolini et al., 2025), existing datasets show limited and uneven availability of variables at both school and teacher levels. At the school level, this lack of information is related to the difficulty of identifying which institutional resources should be monitored systematically. At the teacher level, the situation is even more complex. Collecting information on teachers is challenging for several reasons: going beyond administrative data on teachers often requires the active participation of the latter; they may have concerns about privacy and about the risk of being individually assessed, on the basis of their students' performance; linking teacher information to student- or class-level data requires an additional layer of data infrastructure (namely at least one additional dataset with a merging key); and the predictors of teacher quality are still only partially understood, hence it becomes difficult even choosing which teachers' features should be collected (and how).

All these difficulties hinder the enrichment of longitudinal data on student achievement with crucial information at school and teacher levels. However, a significant step forward in understanding heterogeneous developments in pupil achievement will be possible once this layer of information is collected more systematically.

School variables

School institutional ownership

It is important to harmonise school (public or private) ownership, as it might capture one of the potential mechanisms reproducing social inequalities through education (due to the sorting of students by type of school based on their socio-economic background). This information is available for all datasets except the Italian INVALSI. We adopted the following harmonisation: *tg_schl_type* - Harmonised school type describes funding and governing models, i.e. state, private, public, or other type of governance (0: *Private*; 1: *Public*; 2: *Other*.) Category "other" relates to independent or publicly funded private schools. Public schools are government-funded, while private schools rely on fees. Publicly funded private schools, common in countries like Belgium, Spain, and the Netherlands, are often religious or follow specific pedagogical methods while receiving state funding.

Education level

Another school variable captures the institutional level that student belongs to: Kindergarten, Primary, Lower Secondary, Upper Secondary. This variable allows researchers to track students' progress in time, and it is an essential variable when using datasets with a longitudinal design. We adopted the following harmonisation: *tg_schl_level* - Harmonised level of education (0: Kindergarten; 1: Primary school; 2: Lower secondary; 3: Upper secondary).

School track

The role of school tracking in reproducing social inequalities is a major theme in the literature (see for example Reichelt, Collischon, Eberl., 2019) on educational inequalities and its mechanisms over time. The harmonised variable allows to distinguish among three main school tracks in secondary education: academic, technical, and vocational tracks. We propose the following harmonisation *tg_schl_track* - Harmonised school track (1: Academic; 2: Technical; 3: Vocational; 4: Other.) Category "other" relates to religious, conservatoires or artistic schools.

School size

The variable on school size refers to the enrolment of students attending a specific institution. In survey datasets this variable is based on an indication by the school principal while in full population administrative datasets it can be aggregated from individual level data. To facilitate the comparison of school size across datasets, we focus on the total enrolment of pupils: the variable simply counts the number of students in each school. We adopted the following harmonisation *tg_schl_size* - Number of students (takes values up to 10,000).

Share of students with a migrant background in the school

As a measure of the students' environment, we included among the harmonised variables, the information about the number of pupils with migrant background in each school. We define migrant background if the student is born abroad, or at least one of their parents is born abroad (see Section 4.1 for definitions). This variable comes in the original datasets in two different forms: as an ordinal scale (e.g. no students with migration background to all students with a migration background); as a numeric value (usually the percentage share of students with migration background on the total number of students in the school). For the harmonisation, we applied a transformation of the variable in datasets with ordinal scales into a new ordinal scale with 7 values, from 0 to 6, where 0 means that no migrant students are present in the school and 6 all students have a migrant background. Instead, when datasets contain a numeric value, we applied a transformation of the variable using a bell-shaped curve following the scheme: 0 – None, 0.1 to 15% - Almost none, 16 to 44% less than half, 45 to 55% about half, 56 to 84% More than half, 85 to 99% almost all, 100% - all. It should be noted that this is a proxy based on the students that are in the analysed dataset, meaning that we assume that the observed ones are representative of the entire school. An exception is the MySchool dataset, where this information is derived from the entire school population. This assumption is necessary, but it comes with some issues: i.e., consider students not participating to the assessment (hence frequently not recorded in the dataset) and/or a cohort of students in a school may differ from others for incidental reasons.

We adopted the following harmonisation: *tg_schl_mg* - Share of migrant students (0: None; 1: Almost none; 2: Less than half; 3: About half; 4: More than half; 5: Almost all; 6: All).

Share of female students in the school

Another measure of the students' environment reflects the gender distribution in each school. This variable reflects the share of girls (by convention) present in the school. Except for the MySchool

dataset, the harmonisation of the variable followed the same logic applied to variable concerning the share of students with a migration background and is based on the same assumptions, although they are less problematic in this case.

We adopted the following harmonisation: *tg_schl_gd* - Share of girls in school (0: None; 1: Almost none; 2: Less than half; 3: About half; 4: More than half; 5: Almost all; 6: All).

Class environment

Alongside general school characteristics, we also included information on the class environment, using similar variables as the ones applied at the school level. Three variables were created and coded according to the same logic applied at the school level and are assigned the corresponding values:

- class size (*tg_cl_size*);
- share of students with migrant background in class (*tg_cl_mg*);
- percent of migrant students in class (*tg_cl_mg_percent*)
- share of female students in class (*tg_cl_gd*);
- percent of female students in class (*tg_cl_gd_percent*).

<i>tg_cl_mg</i>	Share of migrant students in class
<i>tg_cl_mg_percent</i>	Percent of migrant students in class
<i>tg_cl_gd</i>	Share of female students in class
<i>tg_cl_gd_percent</i>	Percent of female students in class

Teacher characteristics

Teaching literacy-related courses

The variable identifies whether the teacher is assigned to teaching literacy courses (reading/language-related subjects) to the target class or assess students' literacy aptitude. This is an auxiliary variable that captures subject assignment, not competence or student achievement. Recording such information can help researchers link literacy teachers to literacy outcomes; identify subject-specific teacher characteristics; construct subject-level multilevel models (students nested in literacy teachers and nested in schools). We adopted the following harmonisation: *tg_lang* – School subject – literacy (0: No; 1: Yes.)

Teaching mathematics

Similarly to the previous variable, this one identifies whether the teacher is assigned to teaching mathematics to the target class, or students assessed in mathematics. We adopted the following harmonisation: *tg_math* – School subject - numeracy (0: No; 1: Yes).

Teacher's experience

In the research about the role of teachers in influencing students' achievement, the issue of 'teachers' quality' comes into play. As difficult as it is to operationalise such a complex construct, one potential and fruitful indicator is the amount of experience teachers have. Such experience can be (at least partially) captured by a variable measuring the total number of years of teaching experience. We

adopted the following harmonisation: *tg_teach_year* - Harmonised teaching experience (*takes numerical values from 0 to 70 – with 70 as a theoretical maximum*).

Teacher's sex

The variable identifies whether the teacher is recorded as male or female in the dataset. It reflects sex as registered or reported in the source data. The analytical relevance of this variable is controversial, but, at the same time, collecting it is simple and this is the reason it is often recorded in the datasets. We adopted the following harmonisation: *tg_teach_sex* - Teacher's sex (*0: Male; 1: Female*).

Teacher's employment contract

Among the variables used to capture “teachers’ quality”, part of the research links teachers’ performances in the classroom to their employment condition and stability. This variable may indeed capture the relevance of disruption in teachers’ continuity, due to their turnover, a phenomenon linked to school contexts perceived as more challenging, frequently because of their student composition. The harmonised variable refers to the current labour contract that the teacher has – temporary, permanent, other type of employment. We adopted the following harmonisation: *tg_teach_empl* - Harmonised teacher’s employment (0: Temporary; 1: Permanent).

4.3 Academic Performance Measures

The feasibility study focuses on academic performance in literacy and numeracy, as these are the most assessed domains in this field of research (as also highlighted in Deliverable D3.1 of the present Horizon Europe project, *Mapping Longitudinal Datasets on Educational Inequalities in Primary and Secondary Education: Research Findings and Implications*).

Academic performance measures represent one of the most challenging dimensions in the harmonisation process, as the participating datasets differ not only in the scales used to report achievement, but also in the underlying constructs, assessment designs, scopes, and educational logics. Across the datasets considered in this study—including DEPP, INVALSI, NEPS, NCO, RALC, and MCS—student performance is captured through a wide range of instruments, from raw or weighted test scores to psychometrically scaled IRT-based measures, standard scores, percentiles, and teacher-assigned grades.

Moreover, these measures do not always refer to the same type of competence: some are closely aligned with school curricula and national learning standards, while others capture broader cognitive abilities, such as reasoning, reading speed, or general literacy. As a result, harmonisation requires a careful assessment of what each variable measures, how scores are generated, and whether similar indicators can be meaningfully treated as comparable across countries and cohorts.

In addition, it is necessary to define a common measurement scale to enable comparisons across tests and grade levels. This challenge is addressed within the broader target-variable approach adopted in the feasibility study, which combines conceptual alignment, transparent coding decisions, and detailed documentation of dataset-specific differences.

Methodological Issues and Choices

The first methodological issue concerns conceptual comparability. Before any transformation of scores can be considered, it is necessary to identify whether national indicators refer to sufficiently similar domains of academic performance. The datasets include measures in mathematics, reading, language, science, and broader cognitive skills, but these domains are not always operationalised in the same way. For example, INVALSI and NEPS rely on well-developed psychometric frameworks and clearly defined competence models, while MCS includes a broader mix of cognitive assessments that are not always strictly curricular. Similarly, some datasets provide overall achievement scores, whereas others include even more detailed sub-dimensions or domain-specific sub scores. For this reason, the harmonisation strategy is based first on a conceptual and semantic alignment of variables, aimed at grouping measures into broad comparable domains, while documenting differences in construct definition, level of abstraction, and assessment purpose.

A second issue concerns score metrics. The available measures originate from highly heterogeneous scoring systems: DEPP combines raw scores and IRT-based scores; INVALSI uses Rasch-scaled achievement measures anchored to a national standard metric to make scores comparable across cohorts; NEPS provides plausible values derived from IRT models and weighted likelihood estimation, together with their weight standard errors; RALC reports on weighted test scores; NCO is IRT-based; MCS includes standard scores, percentiles, or raw scores with different ranges depending on the instrument. Given this heterogeneity, a direct comparison of original values is not possible or meaningful since models' coefficients would require transformation to ensure comparability across waves and datasets.

A third issue concerns the longitudinal nature of the data. Since the project aims to explore trajectories over time, harmonisation must preserve not only cross-sectional comparability, but also the interpretability of change. In practice, this means that score differences across waves should reflect comparable learning progress rather than artefacts of scale construction or changes in the assessment difficulty. This is particularly relevant both for datasets where psychometric scaling supports temporal comparability, and for datasets where test instruments differ across ages. For this reason, the proposed approach should not focus exclusively on harmonised levels of performance but should also consider the construction of standardised change indicators, so that growth patterns can be analysed in a more coherent and interpretable cross-national framework.

Overall, from a measurement perspective, the most feasible harmonisation strategy is to transform original scores onto a common standardised metric (e.g. z-scores) within each dataset and subject domain. This approach does not eliminate substantive differences across assessments in terms of content and construct definition, but it allows for consistent comparison of students' relative positions within national distributions over time.

This makes it possible to compare both trajectories over time and the relative magnitude of the associations between student outcomes and their individual and household characteristics across datasets. Differences in the specific content and constructs underlying each assessment are documented in detail (see Section 5.3).

The methodological choice adopted in this feasibility study is deliberately cautious. Rather than forcing all academic performance measures into a single fully equivalent construct, the aim is to create a transparent harmonised framework that preserves national specificity while identifying a minimum set of comparable domains and a shared metric for analysis. This implies accompanying each harmonised variable with detailed metadata on domain, scale type, measurement design, and interpretive limitations, in line with the report's broader emphasis on transparency, codebook construction, and reproducible harmonisation syntax.

Open Issues

Despite the progress made in identifying a workable harmonisation strategy, several open issues remain. Score standardisation can improve formal comparability, but it cannot fully solve differences in substantive meaning. A standardised mathematics score in one country may reflect mastery of curriculum-based content, while in another it may capture broader problem-solving ability or general numeracy. Similar concerns apply to literacy (reading and language) measures, especially where tests differ in the balance between speed, comprehension, inference, and vocabulary. This means that harmonised variables should be interpreted as approximate functional equivalents rather than strictly identical measures. Moreover, some datasets, such as INVALSI, NCO and NEPS, are based on calibrated IRT models and provide robust latent measures of competence, while others rely on

raw marks, weighted scores, or less explicitly documented scoring or grading procedures. This creates an asymmetry in measurement quality, scale properties and even their capacity to capture differences among students and capture variation with sufficient granularity. Bringing all variables onto a common standardised metric improves usability, but may also mask differences in precision, reliability, and sensitivity across instruments. The extent to which these differences affect substantive results remains a critical issue that requires further assessment and robustness checks, to be conducted on the statistical models, changing the metrics of the dependent variables. In other words, once estimated a longitudinal inequality pattern in a subject within a dataset in terms of changes in relative positioning, it will be necessary to test whether it remains stable also replicating the analyses using the different scales available for the same dependent variable.

Finally, a broader question concerns the balance between simplification and analytical richness. A highly reduced harmonised indicator may facilitate cross-country analysis, but it risks losing important information contained in domain-specific structures, sub-scores, and qualitative descriptors. Conversely, preserving too much complexity may weaken comparability and reduce the practical usability of the harmonised dataset. The main challenge, therefore, is to strike a balance between methodological rigor, substantive validity, and operational feasibility. For this reason, the harmonisation of academic performance measures should be treated as an incremental process, to be refined through further analysis as statistical results emerge. These should not only be tested in terms of robustness checks and sensitivity analyses, but also interpreted considering test documentation, distribution checks, and the analysis of missingness, balance, and outlier patterns across datasets, as well as, where possible, the content and cognitive demands of the original assessments.

The harmonised academic performance variables adopted in this study are: *tg_literacy_tests* - Literacy Performance – tests; *tg_numeracy_tests* - Numeracy Performance – tests; *tg_literacy_marks* - Literacy Performance – marks; *tg_numeracy_marks* - Numeracy Performance – marks.

5. Distribution Checks and Data Quality Patterns in the Harmonised Datasets

After harmonising the variables, as described in the previous section, additional steps were undertaken to ensure the quality of the harmonised data. All these steps focused on analysing the distribution of the target variables to assess three key aspects:

- Whether the distribution of the variables in the analysed sample reflects the distribution of the phenomenon in the underlying population (e.g. whether the share of students displaying a given characteristic in the sample broadly corresponds to the same share in the student population; whether competence scores follow an approximately normal distribution; whether there is no abnormal concentration of cases on a small number of values; etc.).
- The extent of missing information due to problems in data collection and/or attrition (e.g. whether the distribution of a given variable changes between the first and last wave because of a decline in cases concentrated among specific groups of students).
- Whether the distributions of the variables show major irregularities, such as outliers or invalid values.

Section 5.1 provides a general overview of the findings by dataset. Section 5.2 presents a set of tables summarising the distribution of each independent variable, grouped by variable domain (students' characteristics, household characteristics, school characteristics, and teacher characteristics), for each dataset by country. For datasets with multiple cohorts, each cohort is described separately. All the tables presented in this section follow the same coding scheme:

- **Column “a”** contains synthetic information on *distribution balance*: whether the variable distribution is balanced with respect to external data sources (when available – reached through Eurostat) or within the dataset (the letter “V” stands for “Verified” and indicates that there are no problems in the variable distribution).
- **Column “b”** contains synthetic information on *missing data*: whether the variable's observations/entries do not present high levels of missing values equal or above 10% of the wave distribution (the letter “V” indicates that there are no problems due to a high level of missing cases).
- **Column “c”** contains synthetic information on *outliers*: whether the variable contains outliers or invalid values (e.g. out of scale) (the letter “V” is used when the presence of outliers is not problematic, with a threshold set below 5%).

Note that the code “ND” is used when the variable is not present in the cohort/dataset.

Section 5.3 focuses on the distributions of the dependent variables, namely student performance, by country and dataset. This section provides a description of the contents assessed for literacy and numeracy among datasets and related procedures/scales, followed by the distributions' analysis (described through the same scheme applied to independent variables).

Finally, section 5.4 summarizes what was observed when inspecting the datasets, providing a general overview of the pattern emerged regarding the selected variables, at the end of the harmonization process.

5.1 An overview by dataset

NEPS (DE): SC2 and SC3

The German National Educational Panel Study (NEPS) provides a particularly strong basis for the feasibility study because it combines rich longitudinal follow-up with extensive methodological documentation, including detailed data manuals, codebooks, competence-test documentation, and wave-specific weighting and nonresponse reports. This documentation framework is an important asset for harmonisation, as it supports the interpretation of source variables and the assessment of data quality over time. Starting Cohort 2 (SC2) shows overall good quality for core student and household variables, with stable basic demographic information and well-documented competence measures. As in many longitudinal studies, however, the progression of the panel is accompanied by some increase in missing information especially in household variables in later waves, which is consistent with the broader nonresponse dynamics explicitly addressed by NEPS through cross-sectional and longitudinal weights and dedicated nonresponse adjustments. At the same time, these weighting procedures are a major strength, since they help researchers account for selective participation and preserve the analytical value of the cohort over time. Starting Cohort 3 (SC3) also stands out for its robust design and high relevance for comparative longitudinal research, as it follows students from Grade 5 through lower secondary education and includes a main sample, a migration-background supplement, and a refreshment sample introduced to address institutional transitions across German federal states. This design improves coverage of heterogeneous educational

trajectories, although it also makes the data structure more complex. In the waves considered for harmonisation, basic student variables such as sex and age remain robust and analytically reliable, while some school- and teacher-related variables are available only in specific waves or through specific instruments. This limited continuity does not reduce the value of SC3, but it does imply that some contextual indicators are better suited for selected-wave analyses than for fully longitudinal harmonisation. Importantly, school and teacher participation, class-level information, and retracking procedures were carefully formalised, and wave-specific weights were constructed to address both school- and student-level participation processes. Overall, SC3 offers high-quality student-level longitudinal information and valuable contextual data, even if some school and teacher measures require cautious use because of uneven wave availability and occasional missingness.

RALC (ES - Catalonia)

The Registre d'Alumnes de Catalunya (RALC) is a particularly valuable source for the feasibility study, being an administrative register embedded in the regular management of the Catalan education system and supported by a clear institutional framework. RALC, as the student register used for planning, management, and evaluation purposes, relying on schools as responsible for keeping students' data and academic records updated. This institutional integration is a major strength from the perspective of data quality, as it supports high coverage and continuity over time. However, the current version of the dataset available for harmonisation is more limited in scope than some of the survey-based datasets included in LINEup. Variables on household background, parental characteristics, and broader migration background are not available in the harmonised extract used for the project. This does not reduce the quality of the available data; rather, it reflects the fact that RALC is primarily designed as an administrative student register rather than as a social survey explicitly focused on educational inequalities. For this reason, a fuller analysis of social inequalities requires integration with richer socio-economic information. From a feasibility perspective, RALC offers a backbone for student-level demographic and educational trajectories, even if its analytical scope for inequality research remains stronger when complemented by other datasets containing more detailed household-level information.

MySchool (EL)

The Greek MySchool dataset has a distinctive profile within the feasibility study, reflecting its nature as a central administrative information system designed to support the day-to-day operation of schools and educational administrative structures across the country. MySchool, as an integrated national infrastructure for the management of primary and secondary education, has a clear institutional role that reflects an important strength from the perspective of data quality, as it supports the regular updating of school- and student-level records and provides a solid basis for the reliability of core administrative information. This strength is especially visible in the quality of basic individual variables such as sex, which are captured consistently, and in the generally good quality of school-level characteristics such as ownership, school size, and class composition. These features make MySchool particularly useful for analysing the institutional and organisational context of learning. However, variables such as country of birth, and parental employment show some limitations in the harmonised extract, especially regarding balanced distributions and, in some cases, missing values beyond the preferred threshold. These patterns should not be interpreted as weaknesses of the system as such, but rather as a reflection of the different purposes served by administrative school databases: MySchool is primarily designed to support school administration and system management, and is therefore particularly strong for institutional variables, while family-background

information may be less systematically collected or standardised for comparative research purposes. Importantly, the dataset does not show major problems related to outliers or invalid values, which confirms the overall internal consistency of the administrative records. From the perspective of the feasibility study, MySchool therefore emerges as a valuable and innovative source: it offers robust information on school organisation and reliable core student records, while its contribution to the study of educational inequalities could become even stronger if complemented by other datasets containing richer socio-economic information at the household level.

DEPP Panels (FR): Cohorts 2007 and 2011

The French DEPP Panels are a major asset for the feasibility study because they combine a strong longitudinal design with an exceptionally well-developed statistical and methodological tradition. Both Panels are considered longitudinal devices specifically designed to follow students' educational trajectories over time and to link achievement, school careers, and family background information. This design is an important strength for harmonisation purposes since it provides a coherent analytical framework and well-documented variables for comparative research. This strength is clearly reflected in the robust quality of core individual variables: in the 2007 cohort, variables such as sex and country of birth remain stable and analytically reliable across the waves considered, and the competence information collected by DEPP is embedded in a broader system of official educational evaluation and statistical production. At the same time, some household variables present the kinds of limitations that are common in longitudinal studies, collecting detailed family information. Parental employment, occupation, and education can show non-negligible levels of missingness in the harmonised material, which should be considered in comparative analyses. However, this needs to be balanced by two positive considerations. First, the inclusion of these variables at all is a significant advantage of the DEPP Panels, since they make it possible to analyse educational inequalities with much greater depth than datasets limited to institutional or demographic information alone. Second, DEPP's documentation highlights continuing efforts to improve the coding and usability of social background variables. Overall, the two DEPP cohorts provide high-quality student-level longitudinal information and a rich basis for inequality research, with the main caution concerning some household variables that remain analytically valuable but require careful treatment because of missing data.

INVALSI (IT): Cohorts 1, 2, and 3

The Italian INVALSI datasets are a major strength of the feasibility study, combining broad coverage with a highly standardised assessment framework and a well-developed system of contextual questionnaires. A student questionnaire is administered to all students in the relevant grades to collect contextual information that supports a more analytical interpretation of test results. Additional teacher questionnaires are also available for sample classes and can be linked to student data via class identifiers. INVALSI data users can rely on the strong standardisation of the testing process and the collection of socio-demographic information through both student questionnaires and school administrative records. They can also be confident that procedures are in place to improve data quality, including controlling coding errors and correcting the effects of cheating. These features are clearly reflected in the harmonised material used in LINEup: across the three selected cohorts, core individual variables such as sex, age, and country of birth are stable, well balanced, and analytically reliable over time across the three selected cohorts. However, some caution is needed regarding household, school, and teacher variables. While household variables such as parental occupation and parental education remain analytically valuable and are well structured, with balanced

distributions and no major outlier problems, but they show some non-negligible missingness in the harmonised cohorts. This is partly due to having richer background information than is available in many purely administrative datasets. School-level information is also an important INVALSI asset, but some variables, such as school size, may require careful handling due to outliers or extreme values in specific cohorts. Teacher-level information is particularly useful where available, but it is less consistent across grades and waves. Teacher questionnaires are linked to sample classes, and the questionnaires and grades may vary across survey years. This is consistent with the limited continuity observed in the harmonised material, particularly in upper secondary education, where the teacher's questionnaire cannot be linked to classroom identifiers after a specific academic year. Overall, INVALSI offers a very solid foundation for comparative analysis thanks to its scale, standardisation, and linkage between achievement and contextual data. Its main limitations are the missing family background information and the uneven availability of teacher information, but these do not diminish the dataset's high value for studying educational inequalities.

NCO (NL)

NCO provides a wide range of information and is affected only to a limited extent by data issues (e.g. attrition). Overall, the dataset contains valuable variables, which are available consistently across waves.

Portuguese Student Register – DGEEC - (PT)

The Portuguese Student Administrative Register –DGEEC , managed within the official education-statistics infrastructure (Direção-Geral de Estatísticas da Educação e Ciência - DGEEC), is a valuable source for the feasibility study because it provides large-scale administrative information that can be used to follow students across several grades and school years. This register makes it possible to track large cohorts over time and to construct longitudinal educational trajectories, which is a major advantage for comparative work. In the material used in LINEup, this strength is reflected in the generally good quality of the variables that are available: across Grades 5 to 12, the selected target variables—covering core student demographics, some household education information, and school characteristics—tend to show balanced distributions, limited missingness, and few problematic outliers. This makes the Portuguese register a particularly solid source for the harmonisation of core longitudinal indicators and for the reconstruction of student trajectories over time. On the other hand, the extract available for harmonisation is more selective than some of the other datasets included in the project. Several individual- and household-level variables are not available, and teacher-level variables are absent. This limitation should be interpreted considering the administrative nature of the source rather than a weakness in data quality. Administrative student registers are especially strong in producing stable, large-scale, and operationally reliable information on enrolment, progression, and school characteristics, while they are often less suited to collecting the richer socio-economic and pedagogical information that is more typical of dedicated survey instruments. From the perspective of the feasibility study, this means that the Portuguese dataset makes a strong contribution to the comparative framework through the quality and longitudinal consistency of the variables it does contain, while its analytical scope for the study of educational inequalities becomes even stronger when complemented by datasets that include more detailed family- and teacher-level information.

MCS (UK)

The MCS is one of the best-established longitudinal cohort studies in Europe, with rich documentation, repeated follow-up, and a wide range of information on children and their families collected over time. Official documentation highlights the methodological sophistication of the study, including extensive derived-variable files, detailed user guides, and explicit strategies to support researchers in dealing with non-response and attrition. This is a major strength for harmonisation, as it provides a robust framework for interpreting source variables and assessing data quality across waves. In the harmonised material used in LINEup, this strength is clearly reflected in the quality of core individual characteristics such as sex and age, which remain stable and analytically reliable. Household variables—including parental country of birth, employment, and education—are also substantively rich and well structured, which is a significant advantage for analysing educational inequalities in a longitudinal perspective. At the same time, some household variables show non-negligible missingness in later waves, which is again consistent with the well-known dynamics of longitudinal cohort studies. Importantly, MCS documentation explicitly recognises attrition and non-response as central analytical issues and provides dedicated methodological guidance, including weighting and recommendations for robust handling of missing data. This means that the presence of missing information should be seen as a challenge that is both transparent and methodologically addressed within the study infrastructure. Another limitation for the purposes of the present feasibility study is that teacher- and school-level variables are much less systematically available compared to household information. However, this should be balanced by the fact that MCS has developed linked education datasets and, in some waves, teacher-survey components, which confirms the study's broader potential for contextual analysis even if such information is not uniformly available across all waves. Overall, MCS provides exceptionally strong longitudinal information on students and their family background, with the main cautions concerning attrition-related missingness in some household variables and the more limited continuity of school- and teacher-level measures in the harmonised material.

5.2 Distribution of Independent Variables by Country and Dataset

This sub-section presents detailed distribution checks for each independent variable across datasets and cohorts. The tables report the results of the three quality checks described above: distribution balance (column a), missing values (column b), and the presence of outliers or invalid values (column c). Variables are organised into four groups for each dataset: individual student characteristics, household characteristics, school characteristics, and teacher characteristics.

NEPS (DE): SC2

Individual student characteristics

Cohort/Grade	WV3			WV4			WV5			WV6			WV7		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Sex	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Birth year	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Birth month	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Country of birth	V	V	V			V			V			V			V
Birth region	V	V	V			V			V			V			V
Migration background	V	V	V			V			V			V			V

Student household characteristics

Cohort/Grade	WV3			WV4			WV5			WV6			WV7		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Parent Sex	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Country of birth (father)	V	V	V			V			V			V			V
Birth region (father)	V	V	V			V			V			V			V
Country of birth (mother)	V	V	V			V			V			V			V
Birth region (mother)	V	V	V			V			V			V			V
Employment (father)	V	V	V			V			V			V			V
Occupation (father)	V	V	V			V			V			V			V
Employment (mother)	V	V	V			V			V			V			V
Occupation (mother)	V	V	V			V			V			V			V
Education (father)	V	V	V			V			V			V			V
Education (mother)	V	V	V			V			V			V			V
Household size	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
N° Children <1	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Household income	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

School characteristics

Cohort/Grade	WV3			WV4			WV5			WV6			WV7		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
School ownership	V	V	V			V	V		V	V	V	V	V	V	V
Education level	V	V	V			V	V		V	V	V	V	V	V	V
School size	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Students migrant background school	V	V	V	V	V	V	V	V	V	V	V	V	ND	ND	ND
Students migrant background school %	V	V	V	V	V	V	V	V	V	V	V	V	ND	ND	ND
Female students' school	V	V	V	V	V	V	V	V	V	V	V	V	ND	ND	ND
Female students' school %	V	V	V	V	V	V	V	V	V	V	V	V	ND	ND	ND
Class size	V	V	V	V	V	V	V	V	V	V	V	V	ND	ND	ND
Students migrant background class	V	V	V	V	V	V	V	V	V	V	V	V	ND	ND	ND
Students migrant background class %	V	V	V	V	V	V	V	V	V	V	V	V	ND	ND	ND
Female students' class	V	V	V	V	V	V	V	V	V	V	V	V	ND	ND	ND
Female students' class %	V	V	V	V	V	V	V	V	V	V	V	V	ND	ND	ND

NEPS (DE): SC3

Individual student characteristics

Cohort/Grade	WV1			WV2			WV3			WV4			WV5			WV6			WV7			WV8			W9		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Sex	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Birth year	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Birth month	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Birth country	V	V	V			V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Birth region			V			V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Migration background			V			V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND

Student household characteristics

Cohort/Grade	WV1			WV2			WV3			WV4			WV5			WV6			WV7			WV8			W9		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Parent Sex	V	V	V	V	V	V	V	V	V	V	V	V	ND	ND	ND	V	V	V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Country of birth (father)			V			V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Birth region (father)			V			V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Country of birth (mother)			V			V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Birth region (mother)			V			V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Employment (father)			V			V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Occupation (father)			V			V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Employment (mother)			V			V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Occupation (mother)			V			V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Education (father)			V			V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Education (mother)			V			V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Household size	V	V	V	V	V	V	V	V	V	V	V	V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
N° Children <1	V	V	V	V	V	V	V	V	V	V	V	V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Household income	ND	ND	ND	V	V	V	V	V	V	V	V	V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND	ND	ND	ND

School characteristics

Cohort/Grade	WV1			WV2			WV3			WV4			WV5			WV6			WV7			WV8			WV9		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Education level							V	V	V			V			V			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
School track												V			V			V	ND	ND	ND	ND	ND	ND	ND	ND	ND
School size	V	V	V	ND	ND	ND	V	V	V			V			V	ND	ND	ND	ND	ND	ND			V			V
Students migrant background school	V	V	V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND			V			V
Students migrant backgr school %	V	V	V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND			V			V
Female students school	V	V	V	V	V	V	V	V	V			V			V	ND	ND	ND	ND	ND	ND			V			V
Female students school %	V	V	V	V	V	V	V	V	V			V			V	ND	ND	ND	ND	ND	ND			V			V
Class size	V	V	V	V	V	V	V	V	V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND
Students migrant background class	V	V	V	V	V	V	V	V	V	ND	ND	ND			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND
Students migrant background class %	V	V	V	V	V	V	V	V	V	ND	ND	ND			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND
Female students class	V	V	V	V	V	V	V	V	V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND
Female students class %	V	V	V	V	V	V	V	V	V			V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND

Teacher characteristics

Cohort/Grade	WV1			WV2			WV3			WV4			WV5			WV6			WV7			WV8			W9		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Literacy	V	V	V	V	V	V	V	V	V	V	V	V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND
Mathematics	V	V	V	V	V	V	V	V	V	V	V	V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND
Experience (n° years)	V	V	V	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Sex	V	V	V	V	V	V	V	V	V	V	V	V			V	ND	ND	ND			V	ND	ND	ND	ND	ND	ND

MySchool (EL)

Individual student characteristics

Cohort/Grade	2019-20			2020-21			2021-22			2022-23			2023-24			2024-2025		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Sex	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Birth year	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Country of birth			V			V			V			V			V			V
Birth region			V			V			V			V			V			V
Migration background			V			V			V			V			V			V

Student household characteristics

Cohort/Grade	2019-20			2020-21			2021-22			2022-23			2023-24			2024-2025		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Parent Sex			V			V			V			V			V			V
Employment (father)			V			V			V			V			V			V
Occupation (father)			V			V			V			V			V			V

School characteristics

Cohort/Grade	2019-20			2020-21			2021-22			2022-23			2023-24			2024-2025		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
School ownership	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Education level	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
School track	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
School size	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Students migrant background school	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Female students' school	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Class size	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Female students' class	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

Teacher characteristics

Cohort/Grade	2019-20			2020-21			2021-22			2022-23			2023-24			2024-2025		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Sex			V			V			V			V			V			V
Employment			V			V			V			V			V			V

RALC Catalan administrative data (ES – Catalonia)

Individual student characteristics

Cohort/Grade	2017-2018			2018-2019			2019-2020			2020-2021			2021-2022		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Sex	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Birth year	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Birth month	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Country of birth	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Birth region	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

School characteristics

Cohort/Grade	2017-2018			2018-2019			2019-2020			2020-2021			2021-2022		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
School ownership	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
School size	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Students migrant background school %	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Female students' school %	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Students migrant background school	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Female students' school	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

Teacher characteristics: no variables available.

DEPP (FR): Cohort N° 1 (DEPP Panel 2007)

Individual student characteristics

Cohort/Grade	2008			2011		
Variables	a	b	c	a	b	c
Sex	V	V	V	V	V	V
Birth year	V	V	<5%*	V	V	<5%*
Birth month	V	V	V	V	V	V
Country of birth	V	V	V	V	V	V
Birth region	V	V	V	V	V	V
Migration background	V	V	V	V	V	V

Note: *Depp 2007 follows students entering Grade 6 in 2007, canonically in France at the age of 11. Including pupils who repeat the year and pupils with early school entry, the sample should be composed by children born between 1995 and 1997. The 1.24% of the sample is instead composed by children born between 1990 and 1994 or in 1998.

Student household characteristics

Cohort/Grade	2008			2011		
Variables	a	b	c	a	b	c
Country of birth (father)	V		V	V		V
Birth region (father)	V		V	V		V
Country of birth (mother)	V		V	V		V
Birth region (mother)	V		V	V		V
Employment (father)	V		V	V		V
Occupation (father)	V		V	V		V
Employment (mother)	V		V	V		V
Occupation (mother)	V		V	V		V
Education (father)	V		V	V		V
Education (mother)	V		V	V		V
Household size	V		V	V		V
N° Children <14	V	V	V	V		V
Household income	V		V	V		V

School characteristics

Cohort/Grade	2008			2011		
Variables	a	b	c	a	b	c
School ownership	V	V	V	V	V	V
Class size				V	V	V

DEPP (FR): Cohort N° 2 (DEPP Panel 2011)

Individual student characteristics

Cohort/Grade	2012			2016		
Variables	a	b	c	a	b	c
Sex	V	V	V	V	V	V
Birth year	V	V	V	V	V	V
Country of birth	V	V	V	V	V	V

Student household characteristics

Cohort/Grade	2012			2016		
Variables	a	b	c	a	b	c
Country of birth (father)	V		V	V		V
Birth region (father)	V		V	V		V
Country of birth (mother)	V		V	V		V
Birth region (mother)	V		V	V		V
Employment (father)	V		V	V		V
Occupation (father)	V	V	V	V	V	V
Employment (mother)	V		V	V		V
Occupation (mother)	V	V	V	V	V	V
Education (father)	V		V	V		V
Education (mother)	V		V	V		V
Household size	V		V	V		V
N° Children <14	V		V	V		V
Household income	V		V	V		V

School characteristics

Cohort/Grade	2012			2016		
Variables	a	b	c	a	b	c
School ownership	V	V	V	V	V	V
Class size	V	V	V	V	V	V

INVALSI (IT): Cohort 1

Individual student characteristics

Cohort/Grade	Grade 5			Grade 8			Grade 10			Grade 13		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
Sex	V	V	V	V	V	V	V	V	V	V	V	V
Birth year	V	V	V	V	V	V	V	V	V	V	V	V
Birth month	V	V	V	V	V	V	V	V	V	V	V	V
Country of birth	V	V	V	V	V	V	V	V	V	V	V	V
Birth region	V	V	V	V	V	V	V	V	V	V	V	V
Migration background	V	V	V	V	V	V	V		V	V	V	V

Student household characteristics

Cohort/Grade	Grade 5			Grade 8			Grade 10			Grade 13		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
Parent Sex	V	V	V	V	V	V	V	V	V	V	V	V
Country of birth (father)	V	V	V	V	V	V	V		V	V	V	V
Country of birth (mother)	V	V	V	V	V	V	V		V	V	V	V
Employment (father)	V		V	V		V	V		V	V		V
Occupation (father)	V		V	V		V	V		V	V		V
Employment (mother)	V		V	V		V	V		V	V		V
Occupation (mother)	V		V	V		V	V		V	V		V
Education (father)	V		V	V		V	V		V	V		V
Education (mother)	V		V	V		V	V		V	V		V
Household region NUTS1	V	V	V	V	V	V	V		V	V		V

School characteristics

Cohort/Grade	Grade 5			Grade 8			Grade 10			Grade 13		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
Education level	V	V	V	V	V	V	V	V	V	V	V	V
School track	ND	ND	ND	ND	ND	ND	V	V	V	V	V	V
School size			>5%*			>5%*			>5%*			>5%*
Students migrant background school %	V	V	V	V	V	V	V	V	V	V	V	V
Female students school %	V	V	V	V	V	V	V	V	V	V	V	V
Class size	V	V	V	V	V	V	V	V	V	V	V	V
Students migrant background class %	V	V	V	V	V	V	V	V	V	V	V	V
Female students class %	V	V	V	V	V	V	V	V	V	V	V	V

Note: School variables, in all the cohorts, are calculated net of the available information, so they are biased by the observations that are available in the dataset.

Teacher characteristics

Cohort/Grade	Grade 5			Grade 8			Grade 10			Grade 13		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
Literacy	V	V	V	V	V	V	ND	ND	ND	ND	ND	ND

Mathematics	V	V	V	V	V	V	ND	ND	ND	ND	ND	ND
Sex	V		V	V		V	ND	ND	ND	ND	ND	ND
Employment	ND	ND	ND	V		V	ND	ND	ND	ND	ND	ND

INVALSI (IT): Cohort 2

Individual student characteristics

Cohort/Grade	Grade 2			Grade 5			Grade 8			Grade 10		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
Sex	V	V	V	V	V	V	V	V	V	V	V	V
Birth year	V	V	V	V	V	V	V	V	V	V	V	V
Birth month	V	V	V	V	V	V	V	V	V	V	V	V
Country of birth	V	V	V	V	V	V	V	V	V	V	V	V
Birth region	V	V	V	V	V	V	V	V	V	V	V	V
Migration background	V	V	V	V	V	V	V	V	V	V	V	V

Student household characteristics

Cohort/Grade	Grade 5			Grade 8			Grade 10			Grade 13		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
Parent Sex	V	V	V	V	V	V	V	V	V	V	V	V
Country of birth (father)	V	V	V	V	V	V	V	V	V	V	V	V
Country of birth (mother)	V	V	V	V	V	V	V	V	V	V	V	V
Employment (father)	V		V	V		V	V		V	V		V
Occupation (father)	V		V	V		V	V		V	V		V
Employment (mother)	V		V	V		V	V		V	V		V
Occupation (mother)	V		V	V		V	V		V	V		V
Education (father)	V		V	V		V	V		V	V		V
Education (mother)	V		V	V		V	V		V	V		V
Household region NUTS1	V	V	V	V	V	V	V	V	V	V	V	V

School characteristics

Cohort/Grade	Grade 2			Grade 5			Grade 8			Grade 10		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
Education level	V	V	V	V	V	V	V	V	V	V	V	V
School track	ND	ND	ND	ND	ND	ND	V	V	V	V	V	V
School size			>5%						>5%			>5%
Students migrant background school %	V	V	V	V	V	V	V	V	V	V	V	V
Female students' school %	V	V	V	V	V	V	V	V	V	V	V	V
Class size	V	V	V	V	V	V	V	V	V	V	V	V
Students migrant background class %	V	V	V	V	V	V	V	V	V	V	V	V
Female students' class %	V	V	V	V	V	V	V	V	V	V	V	V

Note: School variables, in all the cohorts, are calculated net of the available information, so they are biased by the observations that are available in the dataset.

Teacher characteristics

Cohort/Grade	Grade 2			Grade 5			Grade 8			Grade 10		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
Literacy	V	V	V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Mathematics	V	V	V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Sex	V		V	ND	ND	ND	ND	ND	ND	ND	ND	ND

INVALSI (IT): Cohort 3

Individual student characteristics

Cohort/Grade	Grade 2			Grade 5			Grade 8			Grade 10		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
Sex	V	V	V	V	V	V	V	V	V	V	V	V
Birth year	V	V	V	V	V	V	V	V	V	V	V	V
Birth month	V	V	V	V	V	V	V	V	V	V	V	V
Country of birth	V	V	V	V	V	V	V	V	V	V	V	V
Birth region	V	V	V	V	V	V	V	V	V	V	V	V
Migration background	V	V	V	V	V	V	V	V	V	V		V

Student household characteristics

Cohort/Grade	Grade 2			Grade 5			Grade 8			Grade 10		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
Parent Sex	V	V	V	V	V	V	V	V	V	V	V	V
Country of birth (father)	V	V	V	V	V	V	V	V	V	V	V	V
Birth region (father)	V	V	V	V	V	V	V	V	V	V	V	V
Country of birth (mother)	V	V	V	V	V	V	V	V	V	V	V	V
Birth region (mother)	V	V	V	V	V	V	V	V	V	V	V	V
Employment (father)	V		V	V		V	V		V	V		V
Occupation (father)	V		V	V		V	V		V	V		V
Employment (mother)	V		V	V		V	V		V	V		V
Occupation (mother)	V		V	V		V	V		V	V		V
Education (father)	V		V	V		V	V		V	V		V
Education (mother)	V		V	V		V	V		V	V		V
Household region NUTS1	V	V	V	V	V	V	V	V	V	V	V	V

School characteristics

Cohort/Grade	Grade 2			Grade 5			Grade 8			Grade 10		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
Education level	V	V	V	V	V	V	V	V	V	V	V	V
School track	ND	ND	ND	ND	ND	ND	ND	ND	ND	V	V	V
School size			>5%			>5%			>5%			>5%
Students migrant background school %	V	V	V	V	V	V	V	V	V	V	V	V
Female students' school %	V	V	V	V	V	V	V	V	V	V	V	V
Class size	V	V	V	V	V	V	V	V	V	V	V	V

Students migrant background class %	V	V	V	V	V	V	V	V	V	V	V	V
Female students' class %	V	V	V	V	V	V	V	V	V	V	V	V

Note: School variables, in all the cohorts, are calculated net of the available information, so they are biased by the observations that are available in the dataset.

Teacher characteristics

Cohort/Grade	Grade 5			Grade 8			Grade 10			Grade 13		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
Literacy	V	V	V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Mathematics	V	V	V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Sex	V		V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Employment	V		V	ND	ND	ND	ND	ND	ND	ND	ND	ND

NCO (NL)

Individual student characteristics

Cohort/Grade	2013/2014			2014/2015			2015/2016			2016/2017			2017/2018			2018/2019			2019/2020		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Sex	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Birth year	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Birth month	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Country of birth	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Birth region	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Migration background	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

Student household characteristics

Cohort/Grade	2013/2014			2014/2015			2015/2016			2016/2017			2017/2018			2018/2019			2019/2020		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Employment (father)	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Employment (mother)	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Education (father)			V			V			V			V			V			V			V
Education (mother)			V			V			V			V			V			V			V
Household region NUTS1	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Household income	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

School characteristics

Cohort/Grade	2013/2014			2014/2015			2015/2016			2016/2017			2017/2018			2018/2019			2019/2020		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Education level	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
School size	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Share of migrant students at school %	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Share of female students at school %	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

Portuguese Student Register – DGEEC - (PT)

Individual student characteristics

Cohort/Grade	Grade 5			Grade 6			Grade 7			Grade 8			Grade 9			Grade 10			Grade 11			Grade 12		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Sex	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

Student household characteristics

Cohort/Grade	Grade 5			Grade 6			Grade 7			Grade 8			Grade 9			Grade 10			Grade 11			Grade 12		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Employment (father)				V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Employment (mother)				V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Education (father)	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Education (mother)	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

School characteristics

Cohort/Grade	Grade 5			Grade 6			Grade 7			Grade 8			Grade 9			Grade 10			Grade 11			Grade 12		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
School ownership	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Education level	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
School track	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

Teacher characteristics: no variables available.

Millennium Cohort Study (UK)

Individual student characteristics

Cohort/Grade	Age 7			Age 11			Age 14			Age 17		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
Sex	V	V	V	V		V	V		V	ND	ND	ND
Birth year	V	V	V	V		V	V		V	ND	ND	ND
Birth month	V	V	V	V		V	V		V	ND	ND	ND

Student household characteristics

Cohort/Grade	Age 7			Age 11			Age 14			Age 17		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
Country of birth (father)	V		V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Birth region (father)	V		V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Country of birth (mother)	V		V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Birth region (mother)	V		V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Employment (pt 1)	V		V	V		V	V		V	ND	ND	ND
Occupation (pt 1)	V		V	V		V	V		V	ND	ND	ND
Employment (pt 2)	V		V	V		V	V		V	ND	ND	ND
Occupation (pt 2)	V		V	V		V	V		V	ND	ND	ND
Education (pt 1)	V		V	V		V	V		V	ND	ND	ND
Education (pt 2)	V		V	V		V	V		V	ND	ND	ND
Household size	V		V	V		V	V		V	V		V
N° Children <14	V		V	V		V	V		V	V		V
Household income	V		V	ND	ND	ND	ND	ND	ND	ND	ND	ND
Household region NUTS1	V	V	V	V		V	ND	ND	ND	ND	ND	ND

School characteristics

Cohort/Grade	Age 7			Age 11			Age 14			Age 17		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
School ownership	V		V	V		V	V		V	ND	ND	ND

Teacher characteristics

Cohort/Grade	Age 7			Age 11			Age 14			Age 17		
Variables	a	b	c	a	b	c	a	b	c	a	b	c
Experience (n° years)	V		V	V		V	ND	ND	ND	ND	ND	ND
Sex	V		V	V		V	ND	ND	ND	ND	ND	ND

5.3 Description and Distribution of Dependent Variables by Country and Dataset

This sub-section summarises the main characteristics of the dependent variables available in the harmonised datasets, focusing on the two main outcome domains considered in this study: numeracy and literacy. In the first part, the tables describe the subject content, test format, scaling procedures, and grading systems used in each dataset. In the second part, information is reported on data availability (e.g. missing values) in each dataset. In the final overview, concluding the chapter, we highlight the substantial heterogeneity in assessment design and measurement frameworks across countries, which represents one of the main challenges for cross-country harmonisation.

Description of the dependent variables across datasets

In the following pages, for numeracy and literacy separately, we summarise—for each dataset and each cohort—the main features of the dependent variables. To facilitate comparison across datasets, this information is presented in a set of tables.

More specifically, we begin by briefly describing the subject content assessed, since, as noted above, different sub-domains may be included under the same labels of “numeracy” and “literacy”. In this respect, the case of the Portuguese Student Register – DGEEC is particularly illustrative: its more detailed content descriptions reflect the fact that, for this dataset, performance is not based on a single (necessarily synthetic) test, but on a full set of assessments conducted throughout the school year, covering the entire curriculum for each grade.

The tables then focus on test format—namely how information is collected—and on the scaling procedures used to transform it into a numerical measure. Finally, we describe the key features of the variables: the grading scale and its range; the presence of proficiency levels, which allow a substantive interpretation of students’ results; and the existence of a passing threshold.

Numeracy

Dataset	Description of subject contents
Millennium Cohort	Age 7: Spatial problem-solving, numbers, shapes, measurements. Age 17: Reasoning ability with numbers.
DEPP Panel 2007	Grade 6: Basic numeracy skills. Grade 9: Application of mathematical techniques + problem solving.
DEPP Panel 2011	Grade 1: Early number sense & basic geometry; foundational problem-solving. Grade 6: Curriculum numeracy skills: geometry, measurement, applied problem-solving.
NEPS SC2	Assessment of different competence domains including scientific literacy, mathematical competence as well as procedural metacognition, receptive vocabulary, receptive grammatical competencies, and metacognition.
NEPS SC3	Assessment of mathematical competences in the areas the areas quantity, space and shape, change and relationships, or data and chance.
INVALSI	All grades: Assessment of a) problem solving; b) reasoning; c) knowledge: problem solving (within mathematics itself or in applied contexts) and reasoning.
Portuguese Student Register – DGEEC	Grade 5:

Mathematical competencies: Problem solving, mathematical reasoning, computational thinking, mathematical communication;

Numbers: Natural numbers; fractions, decimals, and percentages;

Algebra: Patterns in sequences; numerical and algebraic relationships;

Data: Statistical questions; data collection and organization; graphical representations; data analysis; communication and presentation of a study; probability;

Geometry and Measurement: Plane figures; three-dimensional figures.

Grade 6:

Mathematical competencies: Problem solving, mathematical reasoning, computational thinking, mathematical communication, representations, and connections;

Numbers: Natural numbers; fractions;

Algebra: Patterns in sequences; direct proportionality; numerical and algebraic relationships;

Data: Statistical questions; data collection and organization; graphical representations; data analysis; communication and presentation of a study; probability;

Geometry and Measurement: Plane figures; three-dimensional figures; operations with figures.

Grade 7:

Mathematical competencies: Problem solving, mathematical reasoning, computational thinking, mathematical communication, representations, and connections;

Numbers: Integers; rational numbers;

Algebra: Patterns, sequences, and series; algebraic expressions and equations; functions;

Data: Statistical questions; data collection and organization; graphical representations; data analysis; communication and presentation of a study; probability;

Geometry: Plane figures; operations with figures; three-dimensional figures.

Grade 8:

Mathematical competencies: Problem solving, mathematical reasoning, computational thinking, mathematical communication, representations, and connections;

Numbers: Rational numbers;

Algebra: Algebraic expressions and equations; functions;

Data: Statistical questions; data collection and organization; graphical representations; data analysis; communication and presentation of a study; probability;

Geometry: Plane figures; operations with figures; three-dimensional figures.

Grade 9:

Mathematical competencies: Problem solving, mathematical reasoning, computational thinking, mathematical communication, representations, and connections;

Numbers: Real numbers;

Algebra: Algebraic expressions, equations, and inequalities; functions;

Data: Statistical questions; data collection and organization; graphical representations; data analysis; communication and presentation of a study; probability;

Geometry: Plane figures.

Grade 10

Mathematical models for citizenship: Elections, sharing, finance;

Statistics: Statistical problems; population and sample; univariate and bivariate data; project work;

Synthetic geometry in the plane: Notable points of the triangle; Euler line; nine-point circle;

Functions: General concepts; polynomial functions of degree up to 2;

Analytic geometry: In the plane and in space.

Grade 11:

Geometry: Trigonometry;

	Discrete mathematics: Counting; sequences; arithmetic and geometric progressions; project work; Functions: Cubic and quartic functions; operations with functions; rational functions; differential calculus. Grade 12: Probability and combinatorics; Functions: Continuity and asymptotes; derivatives, monotonicity, and concavity; exponential and logarithmic functions; trigonometric functions; Complex numbers: Mathematical communication.
MySchool	Grade 7: Numerical fluency; basic algebraic manipulation; simple geometric reasoning; application of arithmetic to problems. Grade 8: Algebraic reasoning; graphical interpretation; proportional reasoning; deductive thinking. Grade 9: Multi step problem solving; representation translation (symbolic/graphical); foundational proof strategies. Grade 10: Functional modelling; analytical reasoning; vector comprehension; trigonometric problem solving. Grade 11: Abstract reasoning; extended functional analysis; problem formulation and modelling; spatial reasoning. Grade 12: Calculus application; optimization and interpretation of quantitative data; preparation for tertiary education mathematics.
NCO	Grade 1-6: fundamental math skills: using math language, execute math operations, and using strategies to solve arithmetic and measurement problems.
Catalan RALC	Grade 6: numeration and calculation: space, measurement, and graphical representation of data; relationships and change. Grade 10: numeration and calculation; space, shape, and measurement; relationships and change; statistics.

Dataset	Test format
Millennium Cohort	Home interview, face to face.
DEPP Panel 2007	Standardized national diagnostic test, paper-based booklet with multiple choice and short constructed responses.
DEPP Panel 2011	Standardized national diagnostic test, paper-based booklet with multiple choice and short constructed responses.
NEPS SC2	In class testing with a battery of multiple choice and complex multiple-choice questions.
NEPS SC3	In class testing with a battery of multiple choice, complex multiple-choice questions, and short constructed response
INVALSI	Paper- and computer-based tests, in class environment.
Portuguese Student Register – DGEEC	In school continuous and systematic evaluation during the academic school year (including written tests and quizzes, oral performance and classroom participation, assignments, or other pedagogical activities).
MySchool	In school written examination or assessment (during the school year) In-school assessment includes: 1. written tests and quizzes, 2. oral performance and classroom participation, 3. assignments or other pedagogical activities.
NCO	Paper- or computer-based test, consisting of a limited number of open questions and mostly multiple-choice questions.
Catalan RALC	Paper based test consisting of multiple-choice and open-end questions.

Dataset	Scaling procedure
Millennium Cohort	Age 7: 1. Raw scores (total number of correct responses). 2. ability adjusted (total number of correct responses, accounting for the sets of items the CM was presented with, which depended on age and successful completion of blocks of items, IRT Rasch Model). 3. standardized scores ability and age adjusted on BAS II age normed data. For Age 17 just raw scores
DEPP Panel 2007	Grade 6: Raw scores Grade 9: Raw exam points according to national Brevet scoring scheme
DEPP Panel 2011	Raw scores
NEPS SC2	Partial credit model. A scoring of 0.5 points was applied to each category of the polytomous items, while MC and SCR items were scored dichotomously as 0 for an incorrect and 1 for a correct response
NEPS SC3	Partial credit model. A scoring of 0.5 points was applied to each category of the polytomous items, while MC and SCR items were scored dichotomously as 0 for an incorrect and 1 for a correct response
INVALSI	Item-response model Rasch-based / IRT-style metric
Portuguese Student Register – DGEEC	Grading results from the integration of different indicators (written exams, oral participation, work assignments) emerging from both formative and summative, continuous, and systematic evaluation throughout the academic year.
MySchool	Student performance is assessed on a 0–20 grading scale. Grading is based on a combination of formative and summative assessment throughout the school year
NCO	Grade 1-6: IRT-model, specifically a One Parameter Logistic Model with Conditional Maximum Likelihood. Grade 6 (end of primary school test): IRT-model
Catalan RALC	Adjusted raw scores. For each question, a student can score several points (e.g. 1 point for answering a multiple-choice question correct, 0-4 points for spelling errors in the writing exercise). First the raw score of a specific domain is calculated, then each domain the raw score is adjusted depending on the population weight of that domain, with the total score ranging from 0-100.

Dataset	Grades scale	Proficiency levels	Passing threshold
Millennium Cohort	Age 7 Pattern Construction: raw scores from 0 to 72 + IRT Rasch scale + ability and age adjusted standardized scores Maths: raw scores from 0 to 15 + IRT Rasch scale + ability and age adjusted standardized scores Age 17: raw scores from 0 to 10	No	No
DEPP Panel 2007	Raw scores from 0 to 100	No	No
DEPP Panel 2011	Raw scores from 0 to 9	No	No
NEPS SC2	WLE-based Rasch scale of competence	No	No

NEPS SC3	WLE-based Rasch scale of competence	No	No
INVALSI	Tests: WLE-based Rasch scale of competence Marks: 0 - 10	Cohort 1: No	No
	Tests: WLE-based Rasch scale of competence Marks: 0 - 10	Cohort 2 and 3: level 1- very low proficiency 2 - below expected. 3 - at expected standard 4 - above expected 5 - high mastery	Level 3 is considered the "sufficient" level.
Portuguese Student Register – DGEEC	From the 5 th to 9 th grades – scale from 1 to 5. From the 10 th to the 12 th – scale from 1 to 20.	From the 5 th to 9 th grades 1-5 quantitative scale (1- very low proficiency; 2 - below expected; 3 - at expected standard; 4 - above expected; 5 – high mastery); From the 10 th to the 12 th – 1-20 quantitative scale;	From 5 th to 9 th grades - Level 3 is the minimum passing level. From 10 th to 12 th grades – 10 is the minimum passing level.
MySchool	Scale from 0 to 20	n/a	A grade of 10 in Maths is sufficient for promotion, contingent upon a GPA of 13 in lower secondary and 10 in upper secondary
NCO	Grade 6 (end of primary) commonly a scale between 501 and 550	Grade 1-6: no Grade 6 (end of primary test): 1f: fundamental level 2f: target level	No
Catalan RALC	Scale from 0 to 100	Satisfactory Notable Excellent	No

Literacy

Dataset	Description of subject contents
Millennium Cohort	Age 7: Word decoding ability - Recognition of printed words - Visual and auditory working memory - Skills in word analysis without additional contextual clues - Vocabulary knowledge (BAS II Word Reading) Age 11: Verbal (reasoning using verbal concepts) Measures: (Acquired) verbal knowledge and verbal reasoning: - Verbal reasoning ability - Expressive language skills, including verbal fluency - Vocabulary knowledge - General knowledge - Abstract and logical thinking

	- Ability to distinguish between essential and superficial features - Level of language stimulation (BAS II Verbal Similarities). Age 14: APU Vocabulary Test (Applied Psychology Unit)
DEPP Panel 2007	Grade 6: - Comprehension observe search; Comprehension interpreting; - Production transforming a text; - Production write a text; - Comprehension analyses; - Comprehension collect connect; - Knowledge word recognition. Grade 9: - Reading comprehension, writing, grammar, dictation, vocabulary; based on Brevet specifications.
DEPP Panel 2011	Grade 1: Oral comprehension, writing, phonology, pre-reading. Grade 6: Written comprehension, oral comprehension, language study, Vocabulary/ lexical knowledge.
NEPS SC2	Reading competency tests involve different text types, such as short stories or factual texts. Different tasks are undertaken, which include finding information, drawing conclusions from related sentences, and reflecting on the text.
NEPS SC3	The reading competence test included five texts and five item sets referring to these texts. Each of these texts represented one text type or text function, namely, a) information, b) commenting or augmenting, c) literary, d) instruction, and e) advertising.
INVALSI	Assessment of: a) reading competence, understood as the comprehension, interpretation, and evaluation of written texts; b) grammatical knowledge and skills; c) semantic–lexical competence (implicit and explicit).
Portuguese Student Register – DGEEC	Grade 5: Oral: Presenting, narrating, and expressing opinions; Reading: Informative and explanatory texts, formal communication, narratives; Literary: Introduction to narrative texts; development of reading habits and aesthetic appreciation; Writing: Description, narrative (with dialogue), reasoned opinion; Grammar: Basic knowledge across phonological, morphological, syntactic, semantic, and discourse levels. Grade 6: Oral: Presenting, informing, describing, narrating, evaluating, and arguing (fact-based); Reading: Informative (news, interviews), persuasive (advertising), and functional texts; Literary: Formal aspects of poetic and dramatic texts; Writing: Expository texts (summaries, syntheses), argumentation, narration; Grammar: Progressive knowledge across key linguistic domains. Grade 7: Oral: Presenting, informing, narrating, describing, expressing feelings, persuading; Reading: Biographies, opinion journalism, advertising texts; Literary: Continued study of poetic and dramatic forms; Writing: Summaries, expository texts, opinion texts, narratives, biographies, interviews, commentaries; Grammar: Consolidation of core linguistic knowledge. Grade 8: Oral: Presenting, explaining, and discussing multiple perspectives; Reading: Autobiographical, journalistic, opinion, and functional texts;

	Literary: Further development of poetic and dramatic analysis; Writing: Diaries, interviews, commentaries, reading responses; Grammar: Continued development across linguistic domains. Grade 9: Oral: Presenting, explaining, and arguing in discussions; Reading: Popular science and argumentative texts (reviews, commentaries); Literary: Consolidation of poetic and dramatic analysis; Writing: Opinion texts, reviews, commentaries, summaries; Grammar: Systematized and consolidated linguistic knowledge. Grade 10: Oral: Informing with a critical perspective on contemporary issues; explaining and arguing in debates and discussions; Reading: Narrative (e.g., travel writing), expository texts, and critical texts (reviews, cartoons); Literary: Study and appreciation of Portuguese works from the 12th to the 16th centuries; development of reading habits; Writing: Syntheses, expository texts, and critical reviews; Grammar: Systematized explicit knowledge across phonological, morphological, syntactic, semantic, and discourse-textual domains. Grade 11: Oral: Presenting and arguing in debates and discussions; Reading: Argumentative texts (political discourse, opinion articles, critical reviews); Literary: Study and appreciation of Portuguese works from the 17th to the 19th centuries; development of reading habits; Writing: Expository and argumentative texts; Grammar: Systematized explicit knowledge across key linguistic domains. Grade 12: Oral: Presenting and arguing in debates and discussions; Reading: Argumentative texts (political discourse, opinion articles, critical reviews); Literary: Study and appreciation of Portuguese works from the 20th century; development of reading habits; Writing: Expository and argumentative texts; Grammar: Systematized explicit knowledge across key linguistic domains.
NCO	Grade 1-6: Spelling, understanding, and interpreting. Grade 6 (end of primary school): Reading: vocabulary, understanding, interpreting, searching, summarizing. Spelling: spelling non-verbs, spelling verbs, punctuation marks
Catalan RALC	Grade 6: Listening, writing, reading skills. Grade 10: Reading and Writing

Dataset	Test format
Millennium Cohort	Home interview, face to face.
DEPP Panel 2007	Standardized national diagnostic test, paper-based booklet with multiple choice and short constructed responses.
DEPP Panel 2011	Standardized national diagnostic test, paper-based booklet with multiple choice and short constructed responses.
NEPS SC2	In class testing of reading competence included three types of response formats: simple multiple choice (MC) items, complex multiple choice (CMC) items, and matching items (MA).

NEPS SC3	In class testing of reading competence included three types of response formats: simple multiple choice (MC) items, complex multiple choice (CMC) items, and matching items (MA).
INVALSI	Paper- and computer-based tests, in class environment.
Portuguese Student Register – DGEEC	In school continuous and systematic evaluation during the academic school year (including written tests and quizzes, oral performance and classroom participation, assignments, or other pedagogical activities).
NCO	Paper- or computer-based test, consisting of a limited number of open questions and mostly multiple-choice questions.
Catalan RALC	Paper based test consisting of multiple-choice and open-end questions.

Dataset	Scaling procedure
Millennium Cohort	Age 7 and 11: 1. Raw score (total number of correct responses). 2. Ability adjusted (total number of correct responses, accounting for the sets of items the CM was presented with, which depended on age and successful completion of blocks of items, IRT Rasch Model). 3. Standardized scores ability and age adjusted on BAS II age normed data. For Age 14 just raw scores
DEPP Panel 2007	Grade 6: Raw scores Grade 9: Raw exam points according to national Brevet scoring scheme
DEPP Panel 2011	Raw scores
NEPS SC2	A scoring of 0.5 points was applied to each category of the polytomous items, while MC and SCR items were scored dichotomously as 0 for an incorrect and 1 for a correct response
NEPS SC3	A scoring of 0.5 points was applied to each category of the polytomous items, while MC and SCR items were scored dichotomously as 0 for an incorrect and 1 for a correct response
INVALSI	Item-response model (Rasch-based/IRT-style metric)
Portuguese Student Register – DGEEC	Grading results from the integration of different indicators (written exams, oral participation, work assignments) emerging from both formative and summative, continuous, and systematic evaluation throughout the academic year.
NCO	Grade 1-6: IRT-model, specifically a One Parameter Logistic Model with Conditional Maximum Likelihood. Grade 6 (end of primary school test): IRT-model
Catalan RALC	Adjusted raw scores. For each question, a student can score several points (e.g. 1 point for answering a multiple-choice question correct, 0-4 points for spelling errors in the writing exercise). First the raw score of a specific domain is calculated, then each domain the raw score is adjusted depending on the population weight of that domain, with the total score ranging from 0-100.

Dataset	Grades scale	Proficiency levels	Passing level
Millennium Cohort	Age 7: raw scores from 0 to 90 + IRT Rasch scale + ability and age adjusted standardized scores Age 11: raw scores from 0 to 22 + IRT Rasch scale + ability and age adjusted standardized scores Age 14: raw scores from 0 to 20	No	n/a
DEPP Panel 2007	Tests: raw scores 0-100 Marks: scale from 0 to 20 in grade 9 (national Brevet exam)	n/a	n/a

DEPP Panel 2011	Raw scores, different scales for the different tests Writing from 0 to 10 Phonology and pre-reading from 0 to 16 Comprehension from 0 to 12	n/a	n/a
NEPS SC2	WLE-based Rasch scale of competence	No	No
NEPS SC3	WLE-based Rasch scale of competence	No	No
INVALSI	Tests: WLE-based Rasch scale of competence Marks: 0 - 10	Cohort 1: No	No
	Tests: WLE-based Rasch scale of competence Marks: 0 - 10	Cohort 2 and 3: level 1- very low proficiency) 2 - below expected 3 - at expected standard 4 - above expected 5 - high mastery	Level 3 is considered the "sufficient" level.
Portuguese Student Register – DGEEC	From the 5 th to 9 th grades – scale from 1 to 5. From the 10 th to the 12 th – scale from 1 to 20.	Mark 1- very low proficiency; 2 - below expected; 3 - at expected standard; 4 - above expected; 5 – high mastery;	From 5 th to 9 th grades - Level 3 is the minimum passing level; From 10 th to 12 th grades – Classification of 10 is the minimum passing level.
NCO	Grade 6 (end of primary) commonly a scale between 501 and 550 From grade 10 to 12 th - scale from 1 to 10	Grade 1-6: no Grade 6 (end of primary test): 1f: fundamental level 1s: target level	No
Catalan RALC	Grade scale from 0 to 100	Satisfactory Notable Excellent	No

Distribution of the dependent variables by Country and datasets

We present now, for each dataset, the main features of the dependent variables' distributions, summarized through tables following the same coding schema applied in par. 5.2. In this case column "a" contains synthetic information on distribution balance, checked here only within the dataset; column "b" refers to missing data; column "c" is about outliers.

NEPS (DE)

NEPS data are highly reliable and do not present major problems in relation to the available measurements of students' performances. NEPS SC2 shows a stable core of respondents across the waves. Attrition patterns indicate a slight increase over time. Several factors

explain this pattern, including participant withdrawal together with their parents from the study and temporary nonresponse due to individual retracking, which is defined as temporary dropout in the NEPS documentation, as discussed in Zinn et al. (2020). Attrition patterns in SC3 are even less pronounced compared with SC2, indicating stable respondent continuity across the intermediate waves. A notable increase in dropout after wave 4 is attributed to the removal of 578 students in special-needs schools from the panel (Zinn et al., 2020).

SC2

Wave/grade	Wave 3			Wave 4			Wave 6			Wave 9			Wave 11		
Competences	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
tg_literacy_tests	ND	ND	ND	ND	ND	ND	V*	V	V	V	V	V	ND	ND	ND
tg_numeracy_tests	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

*The literacy test distribution is slightly skewed on the left indicating that for many task takers test was difficult

SC3

Wave/grade	Wave 1			Wave 3			Wave 6			Wave 9		
Competences	a	b	c	a	b	c	a	b	c	a	b	c
tg_literacy_tests	V	V	V	V	V	V	V	V	V*	V	V	V
tg_numeracy_tests	V	V	V	V	V	V	V	V**	V	V	V	V

*A small percent (about 1%) of outliers in literacy test on the right side of the scale

**The numeracy test distribution is skewed on the right indicating that for many task takers test was too easy

MySchool (EL)

The competences of the MySchool dataset are school marks, averaged across two semesters and a written examination, with a pass/fail threshold of 10 and a range that goes from 0 to 20. Each year, around 20% of the students (less during the pandemic waves) receive a mark below 10, although this share varies by grade. Consequently, approximately 80% of observations fall within the 10-20 range.

Cohort/Grade	2019-20			2020-21			2021-22			2022-23			2023-24			2024-2025		
Competences	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
tg_literacy_marks	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
tg_numeracy_marks	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

In waves 2019-20 and 2020-21 the distributions are marginally skewed to the left, while for the rest of the waves they are marginally skewed to the right. The coefficient of kurtosis varies between 1.8 – 2.2.

Catalonia – RALC (ES)

The dependent variables present only some problems of missing values. In 6th grade tests missing values within the grade are about or more than 10% of students, whereas in 10th grade tests missing values within the grade is about or more than 20% of students.

Cohort/Grade	Grade 6			Grade 10		
Variables	a	b	c	a	b	c
tg_literacy_marks_s - Spanish language	V*	≈10%	V	V*	≈20%	V
tg_literacy_marks_c - Catalan language	V*	≈10%	V	V*	≈20%	V
tg_numeracy_marks	V*	≈10%	V	V*	≈20%	V

* The distribution is skewed to the right side. This could indicate that many students reach the passing mark, or the test scores are not too difficult.

DEPP Panel 2007 (FR)

Grade 9 competence assessment derives from the French Diplôme National du Brevet (DNB), a standardised national examination taken by all students at the end of lower secondary education (Grade 9), covering core subjects such as French, Mathematics, History-Geography, Civic Education, and Sciences. Students are assessed through two equally weighted components: continuous assessment, which reflects performance across all subjects throughout the school year, and a series of written and oral final examinations sat at the end of the year. Grade 6 competence assessment is instead a test specifically designed for DEPP panel members.

Cohort/Grade	Grade 6			Grade 9		
Competence	a	b	c	a	b	c
tg_literacy_tests	V		V	ND	ND	ND
tg_numeracy_tests	V		V	ND	ND	ND
tg_literacy_marks	ND	ND	ND	V		V
tg_numeracy_marks	ND	ND	ND	V		V

DEPP Panel 2011 (FR)

In this panel, competences are assessed solely through standardized tests. For literacy in Grade 1, there is not a single test but four different tests covering various aspects of verbal acquisition: writing, phonology, pre-reading, and comprehension. These tests use different grading scales but show similar distributions, so they are presented together in the following table. The same applies to numeracy in Grade 1, with one test for numbers and one for mathematics. In Grade 6, just one test for subject is presented.

Cohort/Grade	Grade 1			Grade 6		
Competence	a	b	c	a	b	c
tg_literacy_tests	V*	V	V	V*		V
tg_numeracy_tests	V*	V	V	V		V

*Distributions do not seem to be problematic but are skewed on the right side

INVALSI (IT)

The three cohorts of INVALSI data on the tests and marks do not present problems in relation to tests, whereas there are some problems with missing data on literacy and numeracy marks.

INVALSI W1

Cohort/Grade	Grade 5			Grade 8			Grade 10			Grade 13		
Competence	a	b	c	a	b	c	a	b	c	a	b	C
tg_literacy_tests	V		V	V	V	V	V	V	V	V	V	V
tg_numeracy_tests	V		V	V	V	V	V	V	V	V	V	V
tg_literacy_marks	V*		V	V		V	V	V	V	V		V
tg_numeracy_marks	V*		V	V		V	V		V	V		V

*Distributions are fine but skewed on the right side, given that in primary school most of the students get high grades

INVALSI W2

Cohort/Grade	Grade 2			Grade 5			Grade 8			Grade 10		
Competence	a	b	c	a	b	c	a	b	c	a	b	c
tg_literacy_tests	V	V	V	V	V	V	V	V	V	V	V	V
tg_numeracy_tests	V	V	V	V	V	V	V	V	V	V	V	V
tg_literacy_marks	V*		V	V*		V	V		V	V**		V**
tg_numeracy_marks	V*		V	V*		V	V		V	V		V

*Distributions are fine but skewed on the right side, given that in primary school most of the students get high grades

**Distribution is odd, since around the 60% of students present grades equal or below to 4 out of 10, a very low grade under the passing threshold of 6

INVALSI W3

Cohort/Grade	Grade 2			Grade 5			Grade 8			Grade 10		
Competence	a	b	c	a	b	c	a	b	c	a	b	c
tg_literacy_tests	V	V	V	V	V	V	V	V	V	V	V	V
tg_numeracy_tests	V	V	V	V	V	V	V	V	V	V	V	V
tg_literacy_marks	V*		V	V*		V	V	V	V	V		V
tg_numeracy_marks	V*		V	V*		V	V	V	V	V		V

*Distributions are fine, but skewed on the right side, given that in primary school most of the students get high grades

NCO (NL)

The blank cells indicate that this variable is missing for a substantial group of children and that these missing are not at random (so a frequency table on parental education is not matching the total Dutch population).

The mathematics and reading skill tests throughout grade 1 to 6 are only available for a subsample of youth, but that subsample is representative of the total Dutch population in terms of all kinds of background characteristics.

Criteria B is not met because of the excessive presence of missing values. A little number of schools and students participate in the spelling tests, so the test distributions are not representative of the total Dutch population.

The end of primary school test is conducted in Grade 6, which means that the related variables are only available for students in that grade.

The variables `tg_numeracy_tests`, `tg_literacy_tests` – reading and `tg_literacy_tests` – spelling are available for subsample of schools and students during the entirety of primary education (except for reading tests in Grade 1). This subsample of students has been shown to differ little in terms of background characteristics from the full population

Cohort/Grade	Grade 1			Grade 2			Grade 3			Grade 4			Grade 5			Grade 6		
Variables	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
<code>tg_numeracy_marks</code> - end of primary school																V	V	V
<code>tg_literacy_marks</code> - end of primary school reading																V	V	V
<code>tg_literacy_marks</code> - end of primary school spelling																V	V	V
<code>tg_numeracy_tests</code>	V		V	V		V	V		V	V		V	V		V	V		V
<code>tg_literacy_tests</code> - reading				V		V	V		V	V		V	V		V	V		V
<code>tg_literacy_tests</code> - spelling	V		V	V		V	V		V	V		V	V		V	V		V

Portuguese Student Register – DGEEC (PT)

For each of the competences two assessments are made in the Portugal school system depending on the level of education. In the lower secondary school, from 5th to 9th grade, students are assessed in Portugal language and mathematics by teachers using 5-score scale. From grade 10th till graduation, grade 12th, students are assessed using different scale from 1 to 20, again in Portuguese and Mathematics. All the integral assessments represent an integration of different indicators (written exams, oral participation, work assignments) emerging from both formative and summative, continuous, and systematic evaluation throughout the academic year.

For lower secondary school assessments students are assigned scores from 1 to 5, where 1 is a low proficiency, 2 – below expected, 3 – expected standard, 4 – above expected, 5 – high mastery. Hence the passing grade for students is 3. Similarly, on a scale 1 to 20, the passing mark is in the middle – 10 – reflecting the expected standard for a student to be promoted to a next grade.

Missing values in DGEEC dataset are structural missing values because they reflect how data were collected. Some variables were not collected in specific waves due to the administrative data collection protocol, as certain items were no longer applicable at those educational stages. Consequently, these missing values reflect data availability rather than participant non-response or dropout.

Importantly, no increase in missing values was observed during the COVID-19 period. This suggests that there was no interruption of schooling or internal assessment processes during that time. Teaching activities continued (including remote learning when necessary), and final grades continued to be assigned according to national guidelines.

Moreover, there is no evidence of systematic patterns of missingness across waves that would indicate attrition bias. The observed missingness is consistent with the longitudinal design of the dataset and changes in the relevance of specific measures over time.

Cohort/Grade	5			6			7			8			9			10			11			12		
Competences	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
tg_literacy_marks	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
tg_numeracy_marks	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

MCS (UK)

The Millennium Cohort Study's competence results available to researchers not affiliated with a UK organisation are test scores, not school grades. F

or numeracy, two tests have been tested at age 7, Pattern Construction (British Ability Scales) and Progress in Maths (developed by the National Foundation for Educational Research). Since at a descriptive level there are no differences between the two distributions, results are provided together, even if analysis was made separated.

For Pattern Construction raw scores were used, to compare the results to the other tests which do not use IRT models. Progress in Maths is the only tests for which ability scores were used for this descriptive aim, as students completed an initial test and based on their scores, they were routed to an easier, medium, or harder section, so that raw scores could not be used in this case.

Cohort/Grade	Age 7			Age 11			Age 14			Age 17		
Competences	a	b	c	a	b	c	a	b	c	a	b	c
tg_literacy_tests	V*	V	V	V	V	V	V	V	V	ND	ND	ND
tg_numeracy_tests	V+	V	V	ND	ND	ND	ND	ND	ND	V	**	V

*The distribution does not seem to be problematic, but is skewed on the left side, meaning that many more students obtain the lowest score than all the other possible scores. This may be due to the Millennium Cohort testing system, which stops the test if the student fails to answer the first few questions positively, and to the use of raw scores.

+the distribution does not seem to be problematic but is a little skewed on the left side. This may be due to the Millennium Cohort testing system, which stops the test if the student fails to answer the first few questions positively.

** only raw scores for each item are available for the cognitive test at this age. Combining score of all the items to create a total score, a lot of missing values arise, as the Millennium Cohort testing system stops the test if the student fails to answer the first few questions positively.

5.4 Key Patterns Across Datasets

Overall, the datasets included in the feasibility study demonstrate the considerable value of the longitudinal data infrastructures currently available in Europe for the study of educational achievement and inequalities. Although they differ in design, scope, and institutional purpose, all of them make an important contribution to the comparative framework developed in LINEup. Some provide particularly rich information on students and households; others offer highly reliable administrative coverage and continuity over time; and others are distinguished by the quality of their documentation and methodological structure.

Across datasets, core student variables are robust and well suited to harmonisation, while school-level information often represents an additional strength, even if more complex to harmonise. Greater variation is observed for household, teacher, and other contextual variables, whose availability is more uneven across countries and waves, reflecting differences in national data systems rather than shortcomings in data quality. The feasibility study therefore shows that the main challenge lies not in the intrinsic reliability of the datasets, but in aligning sources that were created for different—but equally legitimate—policy, administrative, and research purposes.

Across the harmonised datasets, several key patterns emerge. These can be grouped into two main sets: (i) patterns related to the availability and quality of target variables; and (ii) patterns related to the characteristics of the underlying data sources.

The first set of patterns concerns the availability and quality of information across different groups of variables:

- **Individual Demographics:** Variables such as sex and age are universally available and highly robust, showing the highest compliance with all three quality criteria. Migration background is more complex, with a progressive decline in comparability as one moves from simple dichotomous indicators to more detailed measures involving parental characteristics and country of origin.
- **The Household Information Gap:** Most datasets present some degree of missingness in household and parental variables. While the available information is balanced and free from major outliers, missing data pose challenges for analyses controlling for socio-economic status and require appropriate methodological strategies, such as multiple imputation..
- **The School-Variables Information Issue:** Many datasets lack detailed information on school functioning, including characteristics of schools, classes, and teachers. As a result, comparative analyses are often constrained, either in terms of the variables that can be included or the set of countries that can be analysed jointly.
- **The Dependent Variables Problem:** Regarding educational outcomes, three main issues emerge. First, some datasets rely on standardised assessments, while others provide only teacher-assigned grades. Second, the content of the assessments varies across datasets, even when referring to the same domains (e.g. literacy and numeracy). Third, measurement models and scales are highly heterogeneous. These differences allow for transformations and analyses within datasets, but limit direct comparability across them. As a result, only basic harmonisation strategies (e.g. standardisation or percentile-based approaches) are generally feasible, while more advanced harmonisation is possible only for specific subsets of countries.

A second set of patterns concerns the relationship between the characteristics of the data sources and the quality of the available information. These findings extend what already emerged in the Mapping report (Deliverable D3.1):

- **Administrative Data vs. Survey Data:** Administrative datasets (e.g. Greece, Portugal, Spain) are particularly effective in capturing school-level characteristics and ensuring coverage and continuity over large populations. However, they often lack detailed household information. Conversely, survey-based longitudinal studies (e.g. MCS) provide richer individual and household data but are typically weaker in capturing school-level variables.
- **Longitudinal Attrition:** In longitudinal surveys designed to investigate inequalities (e.g. MCS), data quality—particularly in terms of missing values—may decline in later waves. This is especially relevant given smaller sample sizes compared to administrative datasets, and may reduce statistical power for certain analyses.
- **National perspectives embedded in the data:** Even within a harmonised framework, variables often reflect national institutional contexts and conceptual frameworks. Differences in classification systems, definitions, and data collection practices may require additional assumptions or simplifications. For example, country-of-origin classifications or definitions of school classes may vary in ways that are not immediately comparable across systems.

In conclusion, focusing on the source variables, the datasets are balanced and free from major irregularities. However, two main challenges remain. First, missing data—particularly in later waves—limit the completeness of longitudinal analyses. Second, many variables are available only in a subset of datasets and remain only partially comparable, even after harmonisation. These constraints

represent the main limitations for achieving a fully comprehensive cross-national longitudinal analysis of student populations and the development of educational inequalities over time.

6. Limitations

This section discusses the main limitations emerging from the harmonisation process, with a particular focus on the target variables. These limitations concern three main aspects: the comparability of individual and household variables (Section 6.1), the availability and usability of school and teacher variables (Section 6.2), and the comparability of academic performance measures across national assessments (Section 6.3). Most of these limitations reflect structural characteristics of the original datasets, which made the harmonisation process more complex and challenging.

6.1 Harmonisation and Use of Individual and Household Variables

Regarding the variables included in the harmonisation, individual and household variables are the easiest to process and use. This is due to two main reasons:

- Many of these variables are included in all datasets analysed. For example, students' sex and year of birth constitute basic information on cohort members. Other variables, such as migration background and family background, are also frequently available in datasets designed to analyse educational inequalities, as they capture classic ascriptive characteristics associated with such disparities.
- A large share of individual and household variables is measured in broadly similar ways across datasets, with sex again representing an emblematic example. When differences in coding exist, they can usually be addressed through common recoding procedures, although sometimes at the cost of losing more fine-grained information.

Nevertheless, not all individual and household variables have been harmonised to the same extent across datasets. This is due to differences in objectives, methods, variable definitions, categorisation, and coding, all of which may affect concept validity

Difference in the objectives, methods, and categorisation choices of the datasets

The differences between datasets are not specific to the present project; rather, they represent a common challenge in efforts to harmonise pre-existing instruments developed for different purposes and with different structures.

The main discrepancies identified concern differences in dataset objectives, particularly between administrative datasets designed to track students' trajectories and survey-based longitudinal studies designed to address specific research questions.

These differences are reflected not only in the variables collected, but also in the structure, timing, and level of detail of the information. For instance, the Catalan administrative dataset RALC does not include parental information, while the Greek MySchool dataset includes only partial parental variables but extensive school-level information.

Moreover, administrative datasets typically follow entire populations continuously, whereas cohort studies collect information at specific waves and may vary the content over time. For example, the Millennium Cohort Study collects detailed parental information up to age 17 and then shifts focus to other domains.

Therefore, harmonisation may involve comparing students observed at slightly different ages or stages. Comparative analyses of trajectories therefore require statistical models capable of interpolating observations across discrete time points (e.g. growth curve models). This approach introduces additional assumptions and requires careful robustness checks.

Harmonised individual and household variables

Regarding the specific coding of variables, Section 4 has shown that, for some harmonised variables, special cases exist and not all variables can be coded in the same way. Here, we examine how many datasets comply with the harmonisation scheme for each variable and can therefore provide a harmonised version. A variable is considered present if it appears in at least one wave of a dataset. In several cases, however, variables are collected only at specific points in time, and their availability varies across waves.

Variables concerning students' characteristics can be harmonised for most datasets (Table 6). However, problems arise when a greater level of detail is required. For example, most datasets provide information on year of birth, but only some also report month of birth. Similarly, most datasets indicate whether students were born abroad, but only some provide more detailed information on their specific region of birth.

Table 6: Individual variables

Variable	Presence	Datasets that do not include
Student's sex	8/8	
Student's birth year	8/8	
Student's birth month	6/8	MySchool, Portuguese Student Register-DGEEC
Student's country of birth	6/8	MCS, Portuguese Student Register-DGEEC
Student's birth region	5/8	MCS, INVALSI, Portuguese Student Register-DGEEC
Student's migration background	5/8	MCS, RALC, Portuguese Student Register-DGEEC

Household variables related to parental characteristics were also harmonised for most datasets, with birth region being the least frequently available indicator (Table 7). Excluding the datasets with little or no parental information, most datasets include parental variables that can be harmonised, although some special cases remain, such as those concerning parents' sex in the Millennium Cohort Study.

Table 7: Household's variables related to parental characteristics

Variable	Presence	Datasets that do not include
Parent's sex	4/8	Ralc, MCS, DEPP, INVALSI
Mother birth country	4/8	Ralc, MySchool, NCO, Portuguese Student Register-DGEEC
Father birth country	4/8	Ralc, MySchool, NCO, Portuguese Student Register-DGEEC

Mother birth region	3/8	Ralc, MySchool, INVALSI, NCO, Portuguese Student Register-DGEEC
Father birth region	3/8	Ralc, MySchool, INVALSI, NCO, Portuguese Student Register-DGEEC
Mother employment	6/8	Ralc, MySchool
Father employment	7/8	Ralc
Mother occupation	4/8	Ralc, MySchool, NCO, Portuguese Student Register-DGEEC
Father occupation	5/8	Ralc, NCO, Portuguese Student Register-DGEEC
Mother educational level	6/8	Ralc, MySchool
Father educational level	6/8	Ralc, MySchool

Note: As it is not possible to distinguish parents' sex in Millennium Cohort, variables regarding mother/father are here intended for parent 1/parent 2.

Instead, apart from the household region variable, variables related to household structure are the least frequently available across datasets (Table 8). This may reflect both a more limited focus on collecting detailed household information and the practical difficulties associated with obtaining such data.

Table 8: Household structure variables

Variable	Presence	Datasets that do not include
Household size	3/8	Ralc, INVALSI, MySchool, NCO, Portuguese Student Register-DGEEC
Number of children	3/8	Ralc, INVALSI, MySchool, NCO, Portuguese Student Register-DGEEC
Household income quartile	4/8*	Ralc, INVALSI, MySchool, Portuguese Student Register-DGEEC
Household income decile	4/8*	Ralc, INVALSI, MySchool, Portuguese Student Register-DGEEC
Household NUTS1 region	6/8**	Portuguese Student Register-DGEEC, Depp panel

Note: *Specifically, for NEPS the information is available for NEPS SC2 but not for NEPS SC3;

** Specifically, for Depp panel the information is available for Depp panel 2007 but not Depp 2011

In conclusion, while individual variables can be harmonised successfully in almost all datasets, this is only partly true for parental variables, and the harmonisation of household-structure variables is particularly problematic. Moreover, as discussed in Section 5, the presence of a harmonised variable does not necessarily guarantee its analytical quality or comparability, since the harmonisation

process may involve some loss of information and the original variables may already be affected by issues such as high levels of missing data.

6.2 Harmonisation and Use of School and Teachers Variables

As discussed above, school-, class-, and teacher-level variables are collected only rarely and often unevenly. Therefore, the main issue is not the harmonisation process itself, but rather the limited analytical potential resulting from the effort devoted to harmonising these variables. In other words, comparative analyses are confined to school-level variables and, even in this case, only for a subset of the national datasets included in the harmonisation process.

In addition, it should be noted that the school-level variables collected and harmonised in the project (Table 9) often act as proxies for information that would ideally be available in a more direct form. Composition indicators are based on the subset of students included in the datasets, which is assumed to approximate the broader school population. Consequently, all analyses of student segregation across schools, and of the effects associated with this process, should be interpreted with caution, given the way in which these variables were constructed.

Table 9: School variables

Variable	Presence	Datasets that do not include
School type (public/private)	6/8	INVALSI, NCO
Level of education	8/8	
School track	4/8	RALC, Depp panel, MCS, NCO
School size (n. of students)	5/8	Portuguese Student Register-DGEEC, DEPP, MCS
Share of migrant students	4/8	Portuguese Student Register-DGEEC, DEPP, INVALSI, MCS
Share of migrant students (%)	4/8	Portuguese Student Register-DGEEC, DEPP, MCS, MySchool
Share of girls	4/8	Portuguese Student Register-DGEEC, DEPP, INVALSI, MCS
Share of girls (%)	4/8	Portuguese Student Register-DGEEC, DEPP, MCS, MySchool

Moving from school-level to class-level variables (Table 10), the scope for cross-country comparison becomes even more limited. This is mainly because data are often not collected for entire classes of students, due to individual-level sampling designs, and/or because class identifiers are not available in the data. As a result, the possibility of conducting comparative analyses of within-school segregation and peer effects across countries is constrained by the way in which the data are collected and recorded. This limitation obscures an important part of the processes through which education systems may either mitigate or reinforce the development of students' educational inequalities.

Table 10: Class variables

Variable	Presence	Datasets that do not include
Class size (n. of students)	4/8*	Portuguese Student Register-DGEEC, RALC, MCS, NCO
Share of migrant students	2/8	Portuguese Student Register-DGEEC, RALC, DEPP, MCS, MySchool, NCO
Share of migrant students (%)	2/8	Portuguese Student Register-DGEEC, RALC, DEPP, MCS, MySchool, NCO
Share of girls	3/8	Portuguese Student Register-DGEEC, RALC, DEPP, MCS, NCO
Share of girls (%)	3/8	Portuguese Student Register-DGEEC, RALC, DEPP, MCS, NCO

* Specifically, for Depp panel the information is available for Depp panel 2011 but not Depp 2007.

Finally, as shown in Table 11, when it comes to teacher characteristics, comparative analyses become limited both in terms of the theoretical framework (i.e. the range of factors that can be investigated) and geographical coverage (i.e. the number of countries with available information). Extending analyses beyond the national level currently appears unfeasible and would require new data collection efforts by the institutions managing the datasets considered in this study. As a result, the only source of broad cross-country comparability regarding teachers' influence remains cross-sectional international datasets that collect comparable information on teachers (such as PISA, TIMSS, and PIRLS).

Table 11: Teachers' variables

Variable	Presence	Datasets that do not include
School subject	2/8	Portuguese Student Register-DGEEC, RALC, DEPP, MCS, NCO, NEPS
Teaches literacy	1/8	Variable created specifically for INVALSI*
Teaches numeracy	1/8	Variable created specifically for INVALSI*
Teaching experience (years)	3/8	Portuguese Student Register-DGEEC, RALC, DEPP, INVALSI, NCO
Teacher sex	3/8	Portuguese Student Register-DGEEC, RALC, DEPP, NCO, NEPS
Teacher type of contract	2/8	Portuguese Student Register-DGEEC, RALC, DEPP, MCS, NCO, NEPS

Note: * in other datasets there is no need to identify the teacher's subject.

6.3 Harmonisation of Academic Performance Measures

Performance variables are presented in quite different ways, as they are derived from tests or grades intended to measure specific competences. In educational measurement and comparative research, meaningful comparison of test scores requires that they be placed on a common measurement

scale. This implies that the same score value should represent the same level of competence, regardless of where or how it was measured.

This requires at least one of the following condition

- Common test items used across countries.
- Common groups of students taking multiple assessments.
- A shared external reference assessment with psychometrically linkable data.
- Coordinated test design with stable constructs and comparable measurement approaches.

The information collected on national mathematics and literacy assessments across the datasets considered in this report reveals substantial differences between countries. The lack of a shared assessment infrastructure means that numerical score comparability cannot be established in a valid and reliable way. National assessments across EU countries are developed independently to serve national priorities. As a result, they differ in several important respects, including:

- Alignment with national curricula.
- Definitions and emphasis of competencies.
- Test formats and difficulty ranges.
- Test administration (i.e. Paper and pencil with a unique test towards Computer based adaptive tests; the presence or not of an external observer; etc.)
- Scoring and scaling methods.
- Intended purposes (e.g. diagnostic use, monitoring, certification).

Even after harmonisation, important limitations and challenges remain in the interpretation and comparability of academic performance measures across the selected datasets.

A first difficulty is that the available dependent variables are not fully equivalent, even when they appear to refer to similar domains. Across datasets, academic performance may reflect curriculum-based achievement, diagnostic assessment, general cognitive ability, or broader competence constructs. Harmonisation can therefore improve comparability at an operational level, but it cannot fully eliminate differences in the substantive meaning of the original measures.

In addition, harmonisation does not remove the original measurement properties embedded in highly heterogeneous scoring systems. The datasets include raw scores, weighted scores, standard scores, percentiles, and IRT-based measures, sometimes within the same source. These metrics differ not only in scale range, but also in their underlying logic. In some cases, scores are derived from calibrated latent-variable models designed to support comparability across test forms and over time. In others, the available indicators are closer to traditional achievement scores or weighted marks, with less explicit information on their psychometric properties. As a result, transformation to a common metric can facilitate comparison, but cannot fully compensate for differences in reliability, precision, and interpretability.

Some limitations also concern the internal structure of the measures. Some datasets provide only an overall score, while others include subdomains, qualitative competence levels, or multidimensional frameworks that distinguish between different cognitive components. For instance, some sources separate procedural and conceptual dimensions, while others combine them into a single outcome. Harmonisation therefore requires a partial reduction of complexity, which inevitably entails some loss of information. The more the original variables are simplified to fit a common

framework, the greater the risk of weakening the link between the harmonised indicator and the educational meaning of the source measure.

Longitudinal comparability represents an additional challenge. Since the project focuses on educational trajectories, it is not sufficient to align variables cross-sectionally; it is also necessary to ensure that differences over time retain a comparable meaning across countries. This is difficult because cohorts are not always aligned by age, school grade, or stage of the educational pathway, and because repeated assessments may rely on instruments that differ in content, difficulty, or measurement design. Under these conditions, changes observed over time may reflect not only learning progression, but also differences in scale behaviour or test architecture.

Finally, the harmonisation process raises a broader methodological issue: comparability should not be confused with equivalence. A common metric may be useful for descriptive and analytical purposes, but it does not guarantee that the underlying variables represent the same educational reality in each national context.

Harmonised academic performance measures should therefore be interpreted as analytically useful approximations. They are primarily suited to supporting within-country analyses of students' relative positions over time and to enabling cautious cross-country comparisons of patterns and associations, rather than direct comparisons of score levels.

Cross-national comparability remains inherently partial, and interpretation must remain sensitive to the national meaning and measurement properties of the original variables.

7. Lesson Learned: Conclusion and Discussion

Importance of the Feasibility study

The present feasibility study integrates and expands knowledge on what can be achieved in terms of comparative analysis relying on a large set of datasets across European countries.

The study shows that ex-post harmonisation of existing longitudinal education datasets across countries is viable, but it comes with a cost. The ex-post harmonisation process can support the construction of a common evidence base for selected descriptive and within-country analyses framed in a comparative perspective, although this comes with several limitations and caveats. At the same time, the study shows that it is not possible to deliver a fully comparable cross-national longitudinal dataset suitable for robust analyses of trends and mechanisms in educational inequality across several countries. This limitation is not the consequence of a single technical obstacle, but reflects structural heterogeneity in data access, transferability, design, timing, measurement, and institutional context.

Despite these limitations, the harmonisation exercise provides an important contribution by systematically documenting the structure, availability, and comparability of national longitudinal education datasets across several countries. This represents a valuable resource for researchers and policymakers interested in studying educational inequalities, allowing them to design ad hoc comparisons based on subsets of countries and harmonised variables.

Why this matters

The practical implication is that the analytical value of the final harmonised dataset is inversely related to the breadth of the ex-post harmonisation ambition: the more countries, cohorts, and domains are included, the fewer variables can be treated as genuinely comparable, and the more aggregation is required, often implying additional assumptions. This trade-off is documented in the harmonisation literature and in prior cross-national harmonisation projects not necessarily related to education. However, it appears particularly pronounced in the domain addressed by the LINEup project, due to the wide range of measurement issues involved in the study of educational inequalities.

In other words, the present study, combined with the previous mapping, supports researchers and policymakers interested in comparing a limited number of countries on specific issues. The choice of the research focus determines which datasets (and countries) are most suitable for analysis. Available syntaxes and harmonisation codes support researchers in navigating this process.

However, the goal of achieving, in the field of educational performance and inequalities, an infrastructure comparable to the Luxembourg Income Study (LIS) could not be met. LIS is the largest ex-post harmonised income microdata database and covers around 50 countries, many of them in Europe. The attempt to harmonise ex-post educational performance data, together with individual and household variables, has proven more complex than harmonising income- and wealth-related variables. Similar considerations apply to other international harmonisation initiatives based on longitudinal datasets, such as the Cross-National Equivalent File (CNEF), which contains data from population-based household panel surveys.

Limits of ex-post harmonisation: a comparative perspective

The feasibility study highlights a set of structural constraints that limit the potential of ex-post harmonisation in the field of educational achievement and inequalities. These can be summarised as follows:

- **Resources and expertise requirements:** Ex-post harmonisation is costly in terms of time, labour, and expertise. The workload extends well beyond variable recoding to dataset cleaning, wave alignment, documentation review, quality checks, and methodological validation. In practice, this requires a multidisciplinary team including substantive experts (often from the country to which the dataset refers and with prior experience with the dataset itself), survey methodologists, statisticians and, for test-score linkage, psychometric and subject-domain expertise. This exercise is therefore practicable only through a highly coordinated collective effort.
- **Accumulated data inaccuracies:** The feasibility study confirms that harmonisation does not eliminate underlying data weaknesses (when present); rather, it layers them together. Key sources of uncertainty include selective attrition, missing information, non-representative or differently targeted samples, changes in weighting, questionnaire wording, conceptual frameworks, test administration, and data-collection modes. Methodological research shows that such features can materially affect comparability and bias estimates of change.
- **Country-specific institutional heterogeneity and lack of international measurement standards:** Educational trajectories are embedded in the institutional features of education systems. Differences in grade structures, transition points, tracking systems, repetition practices, teacher allocation, and school governance mean that formally similar variables may not capture substantively equivalent processes across countries. Moreover, this type of information is often not collected at all in the datasets. For the limited set of education system variables available, ex-post harmonisation often requires aggregation and simplifying assumptions, reducing the possibility of analysing institutional mechanisms of inequality. In addition, a common framework for measuring key variables—namely students' achievement—is still far from being widely adopted, despite well-established international experiences in the field (such as OECD-PISA).
- **Temporal misalignment:** Longitudinal comparability is further constrained by differences in cohort design, age and grade coverage, and wave frequency. Previous European mapping of educational careers has shown that overlap in variables is not sufficient; alignment in age and time span is also required and can only be partially addressed through statistical models relying on assumptions about the evolution of educational outcomes over time.

7.1 Implications for Cross-National Comparative Research Use

Building on the limitations discussed above, the feasibility study allows us to clarify which types of cross-national analyses can realistically be supported through ex-post harmonisation, and which remain beyond its current scope.

What ex-post harmonization can support:

- Mapping broad patterns of educational inequality over the life course;
- Producing macro-level descriptive indicators;

- Comparing within-country trends or gradients over time across key student characteristics (e.g. gender, socio-economic background);
- Developing more targeted comparisons for tightly defined research questions, based on a limited number of countries and a restricted set of variables..

What ex-post harmonization cannot credibly support:

- Direct cross-country comparisons of trends in test scores or grades when extending the analysis beyond a small number of countries;
- The testing of micro-level mechanisms underlying educational inequalities in a comparative perspective. This is partly due to the limited possibility to harmonise household variables, but especially to the frequent absence or weak measurement of school- and teacher-level variables;
- Analyses that assume full measurement equivalence across countries and waves without additional validation procedures.

What the feasibility study helped us to do:

- Testing the feasibility of defining common concepts and metadata standard.
- Identifying areas where coordinated future data collection would generate the greatest value.
- Highlighting the need for more integrated and coordinated data infrastructures to support robust cross-national longitudinal analyses in the future

In this perspective, a broader question emerges from the harmonisation effort: to what extent would it be preferable to move towards a coordinated European data collection strategy, capable of generating longitudinally comparable data by design? The following section discusses this issue.

7.2 Cost-Effectiveness Reflections: Toward a Coordinated European Data Collection Effort?

The analysis developed in this feasibility study shows that, if the European Union aims to move towards a more integrated and coordinated data collection effort in the field of primary and secondary education, four main strategic options can be considered.

Each option presents a distinct set of strengths and weaknesses, reflecting, on the one hand, the financial and organisational resources required for its implementation and, on the other hand, the type and scope of information and analyses that it can support.

The four options are the following:

1. Ex-post harmonisation of existing national longitudinal studies
2. Incremental coordination of existing national longitudinal studies
3. Transformation of an international cross-sectional survey into a longitudinal panel (integrated with administrative data)
4. Creation of a European longitudinal panel on educational performances.

Option 1: Ex-post harmonisation of existing longitudinal datasets

Ex-post harmonisation, as attempted in this feasibility study, has lower start-up costs than all other options and can generate value from pre-existing investments. However, its marginal costs increase rapidly once researchers move beyond a narrow set of variables, while its returns remain limited. As a result, analytical constraints are substantial, given the high degree of institutional and measurement heterogeneity.

Researchers should treat ex-post harmonised data as analytically useful but epistemically fragile. The main risk is not only reduced statistical power, but also false confidence. Because only a limited subset of information can be harmonised, research questions may become data-driven, and key mechanisms may be omitted—not because they are unimportant, but because they rely on non-comparable measures. In addition, insufficient attention to attrition, measurement changes, and institutional non-equivalence may lead to biased or inaccurate results, or to over-interpreted findings.

An ex-ante coordinated European effort—along the lines of any of the other three options—would be more costly, at least initially, but would address many of the limitations that retrospective harmonisation cannot adequately resolve, or cannot resolve at all: common sampling logic, shared instruments, aligned timing, consistent metadata, and planned comparability of key constructs.

This feasibility study therefore points to a strategic choice for Europe. If the policy goal is limited to descriptive comparability, ex-post harmonisation of selected existing datasets can be justified, also benefiting from complementary evidence provided by cross-national cross-sectional assessments of student achievement (e.g. PISA, PIRLS). If, instead, the goal is to understand how educational inequalities emerge and evolve over the life course, and how they respond to policies and interventions across countries, stronger coordination at the stage of data production is required.

These reflections are consistent with broader discussions on the development of European research infrastructures for social and educational data, which increasingly emphasise the need for coordinated and interoperable data production.

On this basis, three additional strategic directions for Europe can be outlined.

Option 2: Incremental coordination of existing national studies

This option goes in the direction of retaining national datasets but **improving** future comparability through the creation of a network among national dataset producers (or support to emerging ones) that, under the coordination of the European Union, should develop:

- A common minimum set of core variables that each study should provide in a fully harmonised way.
- Shared standards for documenting sampling, weights, and wave changes.
- Common coding for crucial variables, also extended to additional ones, including the adoption of standards already used by Eurostat.
- Common protocols for student assessments or score-linking where feasible.
- Bilingual, or at least English-language, technical documentation.
- Metadata standards for repeated and longitudinal studies.
- Common rules and agreements to facilitate dataset access, circulation, and use among researchers from different countries and institutions.

This feasibility study suggests that data owners can make future ex-post harmonisation easier even without redesigning their surveys from scratch. The most valuable improvements are often mundane: better documentation, version histories for instruments and weights, machine-readable metadata, clearer English-language materials, explicit reporting of mode changes, and the use of international classifications where appropriate. The broader policy lesson is that coordinated international investment yields higher long-run value than repeated one-off retrospective harmonisation.

Among the proposals above, two require further specification: common coding, including the adoption of standards already used by Eurostat; and the development of common protocols for student assessments or score-linking. There are several variables measuring individuals' socio-economic background for which Eurostat adopts consolidated classifications. It would be useful for the common coding exercise to start from these. It would be particularly valuable if an agreement among national dataset producers could be reached regarding the adoption of three variable classifications. It would be useful if all datasets could adopt, for students' parents:

- Occupation: The International Standard Classification of Occupations (ISCO) at a one or two-digits level, which is a classification of occupation groups managed by the International Labour Organisation (ILO). The latest version of ISCO is ISCO-08 (dating from 2008).
- Education level: The international classification for organising education programmes and related qualifications by levels and fields (ISCED). The latest version of ISCED is ISCED 2011.
- Migration background: Eurostat in its "Migrant integration statistics"⁶ provides the following classification: native-born with 2 native-born parents; native-born with one foreign-born parent; native-born with 2 foreign-born parents; and foreign-born persons.

In relation to the development of common protocols for student assessments or score-linking, a potentially sound way forward is to support dataset producers in Member States not to move immediately towards a fully common skill measurement (a challenging task within a coordinated framework). Instead, and more pragmatically, it would be useful to promote the development of a model that places national surveys on a comparable scale at least in terms of broadly defined levels of achievement (e.g. below or above a minimum threshold).

In other words, improving the future comparability of students' achievement scores across European datasets requires introducing a small set of common anchor profiles to be included in each country's data collection framework. The basic idea is not simply to add more test items, but to introduce a limited number of standardised descriptions of student performance that can function as shared interpretive anchors across contexts.

This proposal is broadly consistent with the logic of anchoring vignettes (King et al., 2004), which were developed to reduce problems of interpersonal or cross-group incomparability when the same underlying situation is evaluated through different implicit standards. In the literature, anchoring vignettes are described as brief descriptions of hypothetical persons or situations that help correct differences in the use of response scales across groups.

This approach is relevant to the present case because the harmonisation of achievement variables across countries is not only a technical issue of rescaling scores; it is also a substantive issue of interpretation. Even when two datasets measure similar domains, such as reading or mathematics,

⁶ See https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Migrant_integration_statistics_-_migration_status_and_household_composition#:~:text=The%20article%20presents%20statistics%20on,of%20a%20non%20DEU%20country.

the same level of actual competence may be classified differently due to differences in test design, grading traditions, or national expectations about what counts as basic, intermediate, or advanced performance. As a result, part of the remaining cross-country non-comparability may reflect context-dependent shifts in the meaning attached to achievement levels rather than real differences in students' competences.

The introduction of common anchor profiles would help make these interpretive thresholds more explicit and therefore more comparable across countries. For instance, the Italian INVALSI has developed a variable that classifies students according to a five-level proficiency scale (where values equal to 1 or 2 indicate inadequate competences not in line with expected learning outcomes, and values equal to 5 indicate excellent results).

What makes the INVALSI framework especially useful here is that its proficiency levels are already formulated as ordered descriptions of what students can do. In Italian, the progression moves from the identification of explicit information in short and familiar texts to the integration of distant and implicit information, the reconstruction of global meaning, and the recognition of style, structure, tone, and rhetorical devices in more complex texts. In Mathematics, the progression moves from elementary knowledge and basic procedures to the use of mathematical objects in non-routine situations, the construction of models, the interpretation of data through multiple representations, and the production of appropriate arguments.

This kind of ordered substantive description can support the construction of anchor profiles. For this reason, our proposal is to derive from these proficiency levels a small set of short, standardised performance profiles in reading and mathematics. These profiles could then be used as common anchors in at least two ways.

First, they could support the mapping of national achievement scales onto a more homogeneous interpretive framework. Countries could assess where each common profile would be located on their own national scale, making it easier to detect systematic differences in classification thresholds.

Second, they could serve as a validation device. If the same anchor profile is systematically assigned to different achievement levels in different contexts, this would provide direct evidence that similar performances are being interpreted differently across countries. In this sense, the profiles would not replace existing outcome measures but would complement them by providing an external point of reference for the substantive meaning of the levels assigned.

From an operational point of view, a prudent strategy would be to begin with a limited number of anchor profiles, for example three for each subject. The limits of the approach should be clearly stated. The effectiveness of anchoring depends on certain assumptions, particularly that the criteria used to evaluate the profiles are consistent with those used to evaluate actual performances. For this reason, the construction of the profiles would require careful wording, controlled translation, and pilot testing across countries before being used in full-scale comparative work.

The profiles should also remain focused on observable competences rather than country-specific curricular details; otherwise, they would risk reproducing national differences instead of helping to reduce them. Overall, the proposal is to expand the future set of anchor items not by adding a new cross-national test battery, but by introducing a small number of common interpretive anchors based on ordered descriptions of student performance.

An alternative, more ambitious step would be to include some common items in the tests of different Member States. While this would not be sufficient to fully anchor international performance scales, it would improve comparability by placing samples from different countries on a shared difficulty scale, although this would still require some degree of shared theoretical framework. More

importantly, such an approach would contribute to the gradual development of common frameworks for future work in this direction.

Also in this case, the INVALSI experience is relevant. INVALSI is part of an international consortium of institutions operating in the field of education (FLIP+), committed to sharing assessment tools and best practices and to developing a shared item library for use in testing. It should be underlined that, to maximise the chances of success of such initiatives, it would be advisable to start with numeracy, a domain in which convergence across institutions and countries is easier to achieve.

Such an option would require European Union institutions (e.g. different DGs, Eurostat, etc.) to act as facilitators and promoters of a formal network among the main longitudinal dataset producers in each Member State, along lines similar to those currently used by Eurostat and national statistical offices. One complication, compared to the current Eurostat model, is that not all longitudinal dataset producers in education are public entities. As this feasibility study, as well as Deliverable D3.1, shows, universities and private organisations (e.g. foundations) also invest in the production of longitudinal datasets, and in some countries they are the main data producers.

Therefore, any coordination mechanism established at EU level should include both public and private dataset producers, depending on national contexts. This coordination platform should also address microdata access, which is currently highly heterogeneous across countries and datasets.

Option 3: Transformation of international cross-sectional studies into a longitudinal panel (integrated with administrative data)

Currently, there are two international cross-sectional surveys that assess students' performance in primary education, respectively in literacy and numeracy. The *Progress in International Reading Literacy Study* (PIRLS), conducted by the International Association for the Evaluation of Educational Achievement (IEA), is a cross-sectional study that measures the reading achievement of fourth-grade students. It is carried out every five years. The study collects information at the student, parent, teacher, and school levels, covering topics such as basic demographic characteristics, home environment, school climate for learning, and attitudes toward reading. In addition, the reading assessment focuses on four comprehension processes: retrieving explicitly stated information, making straightforward inferences, interpreting, and integrating ideas and information, and evaluating and critiquing content and textual elements.

The *Trends in International Mathematics and Science Study* (TIMSS), also conducted by the IEA, assesses mathematics and science achievement among fourth- and eighth-grade students. It is implemented every four years, allowing for a quasi-longitudinal design in which the fourth-grade cohort is reassessed in eighth grade. In addition, school, teacher, student, and home questionnaires collect information on curriculum, the education system, and students' home and school environments, as well as their attitudes toward learning.

For the purposes of the present feasibility study—namely, following students' performance from primary to upper secondary education—PIRLS and TIMSS present some advantages compared to the *Programme for International Student Assessment* (PISA) conducted by the Organisation for Economic Co-operation and Development (OECD). Both PIRLS and TIMSS collect data during primary education, whereas PISA is a repeated cross-sectional study that assesses students' reading, mathematics, and science competencies at age 15, i.e. at a later stage in the educational trajectory.

One viable option would be to establish a research agreement between EU institutions and Member States (including Ministries of Education and National Statistical Offices) aimed at transforming these cross-sectional studies into longitudinal ones, by following the same students from fourth grade through to the end of upper secondary education. Each new cohort would be initiated every four years (TIMSS) or five years (PIRLS), and the additional costs associated with the longitudinal component would be shared between European Union institutions and Member States.

This third option could represent a pragmatic and effective solution, positioned between lighter coordination or ex-post harmonisation strategies (as in Options 1 and 2) and the more substantial investment required by Option 4 (see below). At the same time, two potential limitations should be considered.

First, not all EU Member States currently participate in both studies (for example, TIMSS 2023 included 23 EU Member States). In this respect, countries not currently participating should be encouraged to join these international assessments.

Second, both studies are based on sample surveys. As highlighted in the present feasibility study, longitudinal survey designs are affected by attrition over time: the longer the panel duration, the greater the risk of sample loss. TIMSS and PIRLS typically sample at least 4,000 and 4,500 students per grade, respectively, from around 150 schools in each participating country. While these samples are sufficiently large for cross-sectional analyses, they may become more problematic for longitudinal purposes, particularly by the final wave (end of upper secondary education). This is especially the case because tracking students over time would require following them across a much larger number of schools than those originally sampled.

A potential solution could involve, on the one hand, increasing initial sample sizes (including a larger number of schools) from the first wave, and, on the other hand, linking sampled students to administrative datasets (e.g. student registries maintained by education authorities). This would allow the collection of additional information over time and would help mitigate attrition by providing data even when students drop out of the panel (for example due to school dropout or other forms of non-response).

Option 4: Creation of a European longitudinal panel on educational performances

A European panel—especially if it robustly relies on administrative student registry data and large samples (or potentially covers the entire student population)—would involve higher costs and longer lead times (for design, pretesting, and implementation). However, it is the option most likely to support robust cross-country analyses of mechanisms, trends, and policy effects on educational performance, as well as the role played by different sources of education.

As suggested above, starting from numeracy would facilitate convergence across countries on a common theoretical framework and assessment tools. From the outset, this effort should focus not only on student- and household-level information, but also on variables related to school functioning.

Such a data infrastructure would be particularly valuable for addressing research questions that current data cannot adequately answer and for which comparative evidence—accounting for both educational context and country-specific policies—is currently lacking. These include, for instance:

- The role of institutional features of school systems in the development of inequalities over the life course.
- School segregation and its long-term impact on academic performance.

- Teacher allocation and the influence of teacher characteristics and behaviours (such as grading standards and track recommendations).
- The effects of grade repetition and school failure on dropout.
- Track allocation and curriculum differentiation.

Although developing such a dataset would be clearly challenging, it is important to underline that similar approaches have already been successfully implemented in other social domains, both in terms of data infrastructure and scientific output. A prominent example is the *Survey of Health, Ageing and Retirement in Europe* (SHARE), a research infrastructure co-funded by the European Commission, which has been crucial for studying the effects of health, social, economic, and environmental policies over the life course of European citizens and has produced substantial scientific returns over the past decades⁷.

7.3 Conclusions

In conclusion, the feasibility study does not support the creation of a single ex-post harmonised longitudinal dataset that would allow for robust cross-country comparisons of trends in educational inequalities. Instead, it highlights a set of structural obstacles that make such a dataset unlikely to be feasible under current conditions.

On the positive side, the study provides a set of national datasets harmonised to the maximum extent possible within a common framework, enabling national analyses whose results can be meaningfully compared across countries.

The LINEup experience suggests a more limited but still valuable role for harmonisation: identifying comparable cores of information, enabling carefully bounded descriptive and within-country analyses framed in a comparative perspective, and informing the design of future coordinated European data collection efforts.

The main lesson emerging from this phase of the LINEup project is therefore strategic. When the research goal is to explain mechanisms and policy effects across institutional contexts, comparability must be built upstream—into data design, documentation, and governance—rather than reconstructed downstream from heterogeneous national data.

Strengthening coordination in data production is therefore not only a technical issue, but a key condition for developing robust evidence to inform education policy and to better understand the dynamics of educational inequalities across Europe.

Dataset producers and policymakers should carefully assess which of the options outlined in the previous subsection—and other possible alternatives—are both effective and feasible to implement.

⁷ <https://share-eric.eu/impact>

8. References

- 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). Research guidelines for rigorous retrospective data harmonization. *International Journal of Epidemiology*, 46(1), 103–105. <https://doi.org/10.1093/ije/dyw075>
- Hattie, J. (2012). *Visible Learning for Teachers*. Maximizing impact on learning, New York: Routledge.
- Kampylis, P., Fragkiadaki Theodoroulea M., Kandila, M., Cholezas, I., Mobilio, V., Sampson, D., Lievore, I., Mauro, V., Gunzelmann, S., Lanoë, M., Maurya, P., Moulin, L., Ortiz, L., Passaretta, G., & Moreira, P. (2024). *Tracing Educational Inequalities in Primary and Secondary Schools - Insights from a Systematic Review of Longitudinal and Repeated Cross-sectional Studies*. LINEup Project - Deliverable 2.1.
- King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(1), 191–207.
- Kolczyńska, M. (2022). Combining multiple survey sources: A reproducible workflow and toolbox for survey data harmonization. *Methodological Innovations*, 15(1), 62-72. <https://doi.org/10.1177/20597991221077923> (Original work published 2022)
- Kroezen, T., & Alieva, A. (2022). PIONEERED: Data Harmonisation Guidelines. Deliverable No. 4.1. <https://doi.org/10.5281/zenodo.7225243>
- OECD (2005) *Attracting, Developing and Retaining Effective Teachers - Final Report: Teachers Matter*, Paris, OECD.
- Pavolini E., Argentin G., Mobilio V., Lievore I., Kochergina E., Zanga G. (2025). *Mapping Longitudinal Datasets on Educational Inequalities in Primary and Secondary Education: Research Findings and Implications*. Deliverable D3.1, Horizon Europe Project LINEup.
- Reichelt M, Collischon M, Eberl A. School tracking and its role in social reproduction: reinforcing educational inheritance and the direct effects of social origin. *Br J Sociol*. 2019 Sep;70(4):1323-1348. doi: 10.1111/1468-4446.12655. Epub 2019 Mar 29. PMID: 30927272; PMCID: PMC6767437.
- Rivkin, S., Hanushek, E. e Kain, J. (2005). Teachers, Schools and Academic Achievement, *Econometrica*, 73(2), 417-458.
- SDR Team. 2019. “[From Source to Target: Harmonisation Workflow, Procedures and Tools. Part 3: Quality Check and General Variable Report](#)”. Presentation prepared by Weronika Boruc for the SDR Workshop Building Multi-Source Databases for Comparative Analyses, Warsaw, Poland, December 16-20.
- SDR Team 2019. “[Approaches to Ex-post Harmonisation of Rating Scales in the Survey Data Recycling Framework](#)” presentation prepared by Irina Tomescu-Dubrow and Ilona Wysmulek for the Gesis conference “Ex-post Harmonisation of Rating Scales in National and International Surveys”, Mannheim, Germany, May 9-11.
- Slomczynski, K. M., & Tomescu-Dubrow, I. (2018). Basic principles of survey data recycling. *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)*, 937-962.
- StataCorp. 2023. *Stata Statistical Software: Release 18*. College Station, TX: StataCorp LLC.

Zinn, S., Würbach, A., Steinhauer, H.W. *et al.* Attrition and selectivity of the NEPS starting cohorts: an overview of the past 8 years. *AStA Wirtsch Sozialstat Arch* 14, 163–206 (2020).
<https://doi.org/10.1007/s11943-020-00268-7>