



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.



Funded by
the European Union



All LINEup resources are licensed under [CC BY 4.0 LEGAL CODE Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/)

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 authors 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>

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.