

PROJECT OVERVIEW

The LINEup project conducted **the first EU-wide mapping of existing longitudinal datasets** on student learning outcomes within primary and secondary school education.

104 Datasets Mapped*



24 European Countries (out of 32)



Focus: Student Learning Outcomes and Academic Performance

Data Types: Longitudinal (i.e. panel, repeated cross-sections, pseudo-panels)

*One U.S. dataset is included as a best-practice example, but excluded from the descriptive statistics

RESEARCH POSSIBILITIES

The potential of a dataset depends on the research questions we ask. Most support descriptive research on individual characteristics. But explaining inequalities through school-level or comparative analysis remains difficult. Even when access is granted, poor documentation limits reusability.



WHAT WE CAN DO

Research Possibilities

- Analyse learning trajectories (especially via test scores or grades)
- Study socio-demographic effects (presence of basic ascriptive variables)
- Track grade retention and transitions (coverage of key educational transitions)
- Compare performance by school type/track (availability of basic school-level identifiers)



WHAT WE CANNOT DO

Analytical Limitations

- Study peer/class composition (lack of class-level data)
- Assess impact of teacher quality or turnover (limited variables)
- Analyse school segregation or territorial inequalities (no geo-referencing)
- Compare non-cognitive skills cross-country (heterogeneous or missing data)

METHODOLOGY | A Three-Step Strategy

1. DATASET IDENTIFICATION

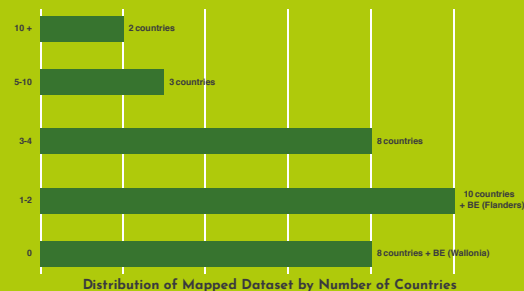
- Literature Review (43)
- Expert Questionnaire sent to 315 Experts from 28 Countries (70)
- Targeted Web-Search & Partner Knowledge (108)
- Collaboration with other projects (8)

2. DATASET SELECTION

- **229** datasets **screened**
- **104** datasets **included**
- **125** datasets **excluded**

3. DATASET ANALYSIS

Focus on Data Structure, Time Span, Indicators & Variables, Quality, Accessibility



KEY FINDINGS | By the numbers

HOW are datasets designed, linkable, and accessible to users?

- 68%** are panel data (individual tracking over time)
- 22%** are repeated cross-sectional (with individual identifiers)
- 10%** can be used as pseudo-panel
- 47%** are survey data
- 19%** are administrative data
- 34%** are hybrid data (survey + admin)
- 55%** allow linkage to other data sources
- 11%** are open access
- 50%** are accessible upon request
- 26%** are highly restricted or not accessible

WHO is included in the datasets?

- 32%** cover the entire student population
- 45%** use random samples
- 23%** use non-random samples
- 50%** with sample < 5,000 students
- 29%** with sample > 10,000

WHAT do the datasets measure?

- 77%** include standardised tests
- 69%** include school grades
- 53%** include both tests and grades
- 64%** cover both numeracy and literacy skills
- 36%** cover foreign language skills
- 30%** cover science skills
- 69%** cover other domains skills
- 95%** include information on students' gender
- 68%** include information on foreign birth country
- 74%** include information on parental education
- 61%** include information on parental occupation
- 30%** include information on teachers
- 23%** include information on classroom composition

WHERE do the datasets apply?

- 62%** have national coverage
 - 38%** are sub-national.
- Most datasets come from Western & Northern Europe

WHEN were data collected and for how long?

- 23%** span 14+ years
 - 20%** cover only 2 years
 - 51%** are still ongoing
 - 47%** cover both primary and secondary education.
- Most frequently: grades 8, 9, 10
- 30%** begin in kindergarten

The value of longitudinal data lies in its responsible, effective use.

LINEup contributes to a more transparent and visible education data landscape in Europe.

