Analysis of the OECD's PISA Data

Projects – Applied Statistics a.y. 2022/23

Mathematical Engineering

Chiara Masci
chiara.masci@polimi.it

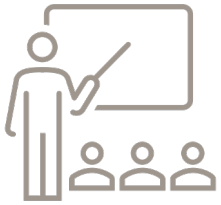
Alessandra Ragni
alessandra.ragni@polimi.it

What is PISA?

PISA

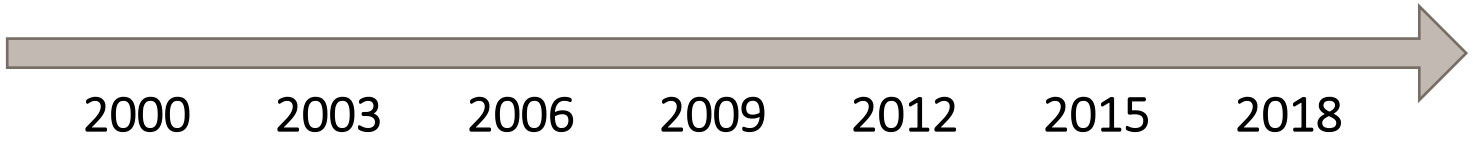
Programme for International Student Assessment

OECD
BETTER POLICIES FOR BETTER LIVES

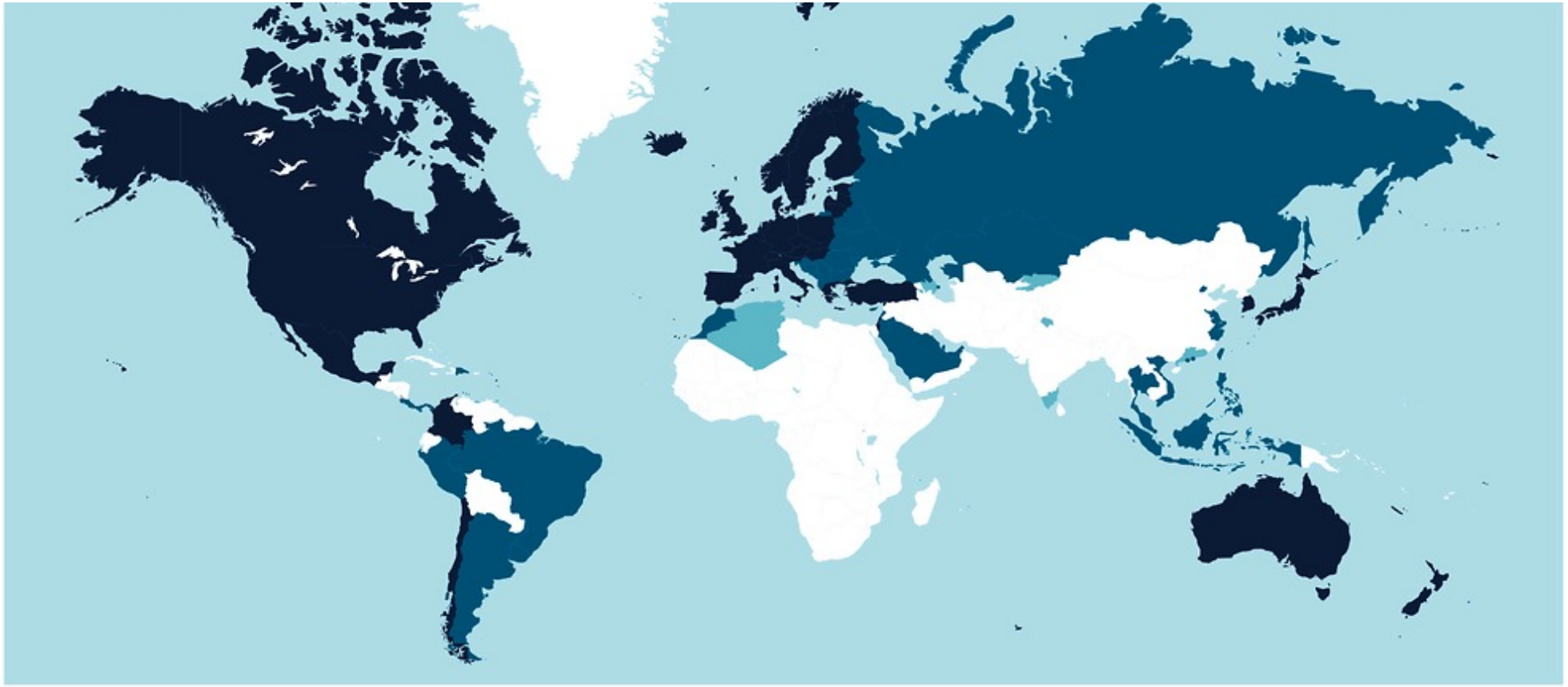


- **PISA** measures **15-year-olds' ability** to use their **reading**, **mathematics** and **science** knowledge and skills to **meet real-life challenges**;

- Multiple databases are available at https://www.oecd.org/pisa/data/;



Countries involved in *PISA 2018*



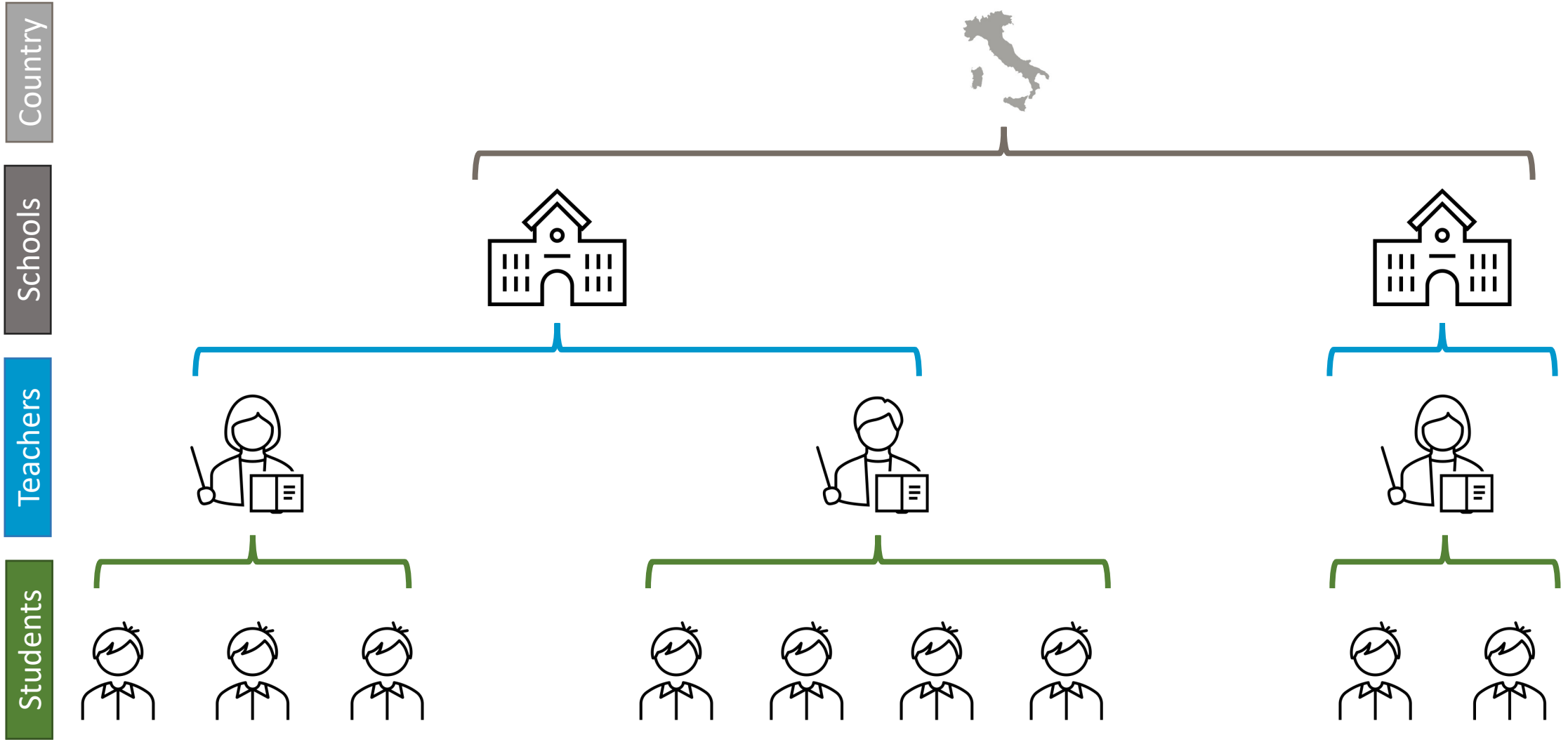
OECD member countries

Australia	Lithuania
Austria	Luxembourg
Belgium	Mexico
Canada	Netherlands
Chile	New Zealand
Colombia	Norway
Czech Republic	Poland
Denmark	Portugal
Estonia	Slovak Republic
Finland	Slovenia
France	Spain
Germany	Sweden
Greece	Switzerland
Hungary	Turkey
Iceland	United Kingdom
Ireland	United States*
Israel	
Italy	
Japan	
Korea	
Latvia	

Partner countries and economies in PISA 2018

Albania	Malaysia
Argentina	Malta
Baku (Azerbaijan)	Republic of Moldova
Belarus	Montenegro
Bosnia and Herzegovina	Morocco
Brazil	Republic of North Macedonia
Brunei Darussalam	Panama
B-S-J-Z (China)**	Peru
Bulgaria	Philippines
Costa Rica	Qatar
Croatia	Romania
Cyprus	Russian Federation
Dominican Republic	Saudi Arabia
Georgia	Serbia
Hong Kong (China)	Singapore
Indonesia	Chinese Taipei
Jordan	Thailand
Kazakhstan	Ukraine
Kosovo	United Arab Emirates
Lebanon	Uruguay
Macao (China)	Viet Nam

The Hierarchical Structure



PISA 2018: the data files

The main data files are:

1. the student-questionnaire; *[very large]*
2. the school-questionnaire;
3. the teacher-questionnaire.

Codebook is a huge excel reporting, for each questionnaire and for each question:

- the codes,
- the explanation
- a small summary on the variable values;

Student questionnaire data files

NAME	VARLABEL	TYPE	FORMAT	VARNUM	MINMAX	VAL	LABEL	COUNT	PERCENT
ESCS	Index of economic, social and cultural status	NUM	7.4	872	-8.1734-4.2051				
						.V/9999995	Valid Skip	0	0.00
						.N/9999997	Not Applicable	0	0.00
						.I/9999998	Invalid	0	0.00
						.M/9999999	No Response	0	0.00
						SYSTEM MISSING	Missing	14379	2.35

How to set up the analysis

1. Read the **questionnaires** and select the most interesting questions;
2. Go to the **codebooks** and check the format of the answers (numerical or categorical);
3. In the Codebook, after all the questions, you can also find some **extra indicators (numeric) computed by OECD** that you can use [e.g.: **ESCS** (Index of economic, social and cultural status), **HISEI** (Index highest parental occupational status), **TMINS** (Total learning time - minutes per week), **WEALTH** (family wealth)].
4. Download the data:

SPSS (TM) Data Files (compressed)

Databases including Global Competence variables were uploaded in October 2020.

- Student questionnaire data file (489 MB)
- School questionnaire data file (3.1 MB)
- Teacher questionnaire data file (12.8 MB)

How to import SPSS data?

https://www.r-bloggers.com/2014/03/how-to-open-an-spss-file-into-r/

```
library(foreign)
schools = read.spss("CY07_MSU_SCH_QQQ.sav", to.data.frame=TRUE)
students = read.spss("CY07_MSU_STU_QQQ.sav", to.data.frame=TRUE)
teachers = read.spss("CY07_MSU_tch_QQQ.sav", to.data.frame=TRUE)
```

Further information avilable at https://www.oecd.org/pisa/data/httpoecdorgpisadatabase-instructions.htm

An example

Variable name	Type	Explanation
GENDER	0/1	0 = male 1 = female
ESCS	Num	Socio-economical status (mean = 0, sd = 1)
IMMIGRANT	Cat	0 = not immigrant student 1 = first generation immigrant 2 = second generation immigrant
TIME HOMEWORK	Int	Number of hours of student homework per week
HISCED	Cat	Highest level of education of parents (levels from 0 to 6)
VIDEO GAME	0/1	Whether the student plays video games or not
SPORT	0/1	Whether the student plays sport or not
DISCIPLIN CLIMATE	Num	How is the disciplinary climate in class
TEACHER SUPPORT	Num	Teacher support in class
MMINS	Num	Hours of mathematics lessons per week
BELONG	Num	Subjective well-being: sense of belonging to school
MOTIVAT	Num	Student attitudes, preferences and self-related beliefs: Achieving motivation
ANXTEST	Num	Personality: test anxiety
COOPERATE	Num	Collaboration and teamwork dispositions: Enjoy cooperation
PARENTS SUPPORT	Num	Parents emotional support
CULTURAL POSSESSION	Num	Cultural possession at home
HOME EDUCAT RESOURC	Num	Home educational resources

Variable name	Type	Explanation
# STUDENTS	Num	Number of students in the school
RATIO-COMPUTER-STUD	Num	Number of available computers per student
MANAGEMENT1	1/6	How much the school principal uses student performance results to develop school's educational goals
MANAGEMENT2	1/6	How much the school principal discusses schools' academic goals with teachers at faculty meetings
STUD-ADMIT-RECORD	0/1	Whether the students are admitted to the school depending on their previous scores or not
PRIVATE	0/1	0 = Public school 1 = Private school
% GOVERN FUNDS	Num	Percentage of school funds given by the government
TEACHERS-INADEQ	1/4	How much the principal thinks that teachers are inadequate (on a 1 to 4 scale)
MATERIALS-INADEQ	1/4	How much the principal thinks that materials are inadequate (on a 1 to 4 scale)
INFRASTRUCT-INADEQ	1/4	How much the principal thinks that infrastructures are inadequate (on a 1 to 4 scale)
RATIO-STUDENTS-TEACHER	Num	Student-teacher ratio
RATIO-STUDENTS-TEACHER5	Num	Student-teacher with level 5 ratio
% STUD SPECIAL NEEDS	Num	Proportion of students with special needs
% DISADVANT STUDENTS	Num	Proportion of disadvantaged students in terms of socio-economical index
STUDENTS TRUANCY	1/4	Students truancy (on a 1 to 4 scale)
STUD-NO-RESPECT-TEACH	1/4	Students lack respect for teachers (on a 1 to 4 scale)
TEACHER ABSENTEEISM	1/4	Teacher absenteeism (on a 1 to 4 scale)
% PARENTS SPEAK TEACHERS	Num	Proportion of students' parents speaking with teachers at the meeting
% PARENTS IN SCHOOL GOVERN	Num	Proportion of students' parents participating at the school government

Project goals

1. **Identify** factors that **explain** the phenomenon you are interested in (e.g., performances in math/italian/science scores among different schools/countries);

2. **Analyse** the effect of the **teacher/school/country** belonging;

3. Look for **differences** between students:
→ having different **teachers** (effect of the teacher);
→ being in different **schools** (effect of the school);
→ being in different **countries** (effect of the country);
using techniques such as **linear** models, **classification** models, **mixed effects** models or **clustering**.

Contacts

Chiara Masci
chiara.masci@polimi.it

Alessandra Ragni
alessandra.ragni@polimi.it