| INTRODUCTION O | THEORETICAL FRAMEWORK | Methodological Issues | Results 000000 | Reference | | |
|--|------------------------------|----------------------------------|-------------------|-----------|--|--|
| | | | | | | |
| | Teacher Asses Performance | sment and Stud in OCSE-PISA 2 | .ents' 2009 | | | |
| INVALSI 2012 International Conference October 3 - 5, 2012 Rome (Italy) | | | | | | |

Brunella Fiore¹ and Isabella Romeo²

¹ University of Milano Bicocca, Italy, e-mail: brunella.fiore@unimib.it
 ² University of Milano Bicocca, Italy, e-mail: isabella.romeo@unimib.it

3^{*th*} October, 2012

イロト イポト イヨト イヨト

| INTRODUCTION 0 | THEORETICAL FRAMEWORK | Methodological Issues | Results 000000 | REFERENCES |
|----------------|-----------------------|-----------------------|-------------------|------------|
| Agenda | | | | |



INTRODUCTION

- Educational accountability
- 2

THEORETICAL FRAMEWORK

- Standardized Test and Teachers'Judgment
- Teachers' judgment in literature
- Aim of the study
- 3 METHODOLOGICAL ISSUES
 - OCSE-PISA Data-set
 - Data Methodology
 - Multilevel approach

4 RESULTS

- Correlations
- Results



 Introduction
 Theoretical Framework 000
 Methodological Issues 000000000
 Results 000000000
 References

 EDUCATIONAL ACCOUNTABILITY

In recent years a growing interest in **educational accountability** has been developed.

- more reliable comparison of outcomes across schools, regions or countries
- measurement of the impact of various educational entities on student progress

Focus on Teachers' Judgment : The objective is to analyze whether teachers overestimate or underestimate their students exploiting OCSE-PISA 2009.

 \Rightarrow *A* comparison between **Teachers' Mark** and **Standardized Test** have been made.

Attention on External Examinators and Standardized Test on determining students' ability level:

- High in Anglo-Saxon Countries;
- Not High in other countries as Italy.

But teacher's judgment could not be an objective evaluation:

- Inot considering a national standardized scale;
- could be influenced by a variety of sources (i.e student's behavior).

Since the end of the 1970's a great interest on the relation between different measures by teacher's judgment and standardized test:

- a mean correlation of 0.67 in 16 previous studies (Hoge,Coladarci 1989);
- a range correlation between 0.28 and 0.92 in all studies considered.

The high variability underlines the need for a definition of a more robust method.

Aim of the study:

to examine in upper secondary schools the variability of teachers' judgment on reading marks compared to the OCSE-PISA 2009 standardized test.



Which student and school characteristics have an impact on both overestimation and underestimation of students

Where?:

Lombardy region.

- a more homogeneous context;
- data on ability level.

OCSE-PISA DATA-SET

The key data-set used in the research is provided by the OCSE-PISA (Programme for International Student Assessment).

The survey

- involves 15-years old students (cross-sectional survey);
- takes places every three years with a different major subject area: reading (2000, 2009), mathematics (2003), science (2006).

This research contribution focus on reading given that it represent the focus of the last survey.

Results 000000

SAMPLE DESIGN AND WEIGHTS

In order to understand the methodology used is necessary to consider the theoretical and methodological framework of such data.

Sample design

PISA samples students in **two stages**: schools are first sampled and then students are sampled in the participating schools (hierarchical data structure). A two stage sampling is chosen given that students within a school usually have more common characteristics than students from different schools.

Weights are associated to each student and to each school

Students and schools did not necessarily have the same probability of selection.

STANDARD ERROR: REPLICATES AND PVS

Standard Error = Sampling Error + Imputation Error

Sampling Error: The replicates

Sampling error estimates are obtained through the **the Fay's variant of the Balanced Repeated Replication (BRR)** technique given the complex sample design.

Imputation Error: Plausible Values (Pvs)

- *WHAT are PVS*?: Instead of directly estimating the student's skill from the sample data, a probability distribution is estimated and PVs are random draws from this distribution.
- *HOW* using them?: It is necessary to run each analysis with each PV. The **final estimate** is obtained as mean of each PV, while the **imputation error** is obtained as root square of the variability of such PVs.

INTRODUCTION

Results 000000 References

THE OBTAINED VARIABLE

Two student evaluations from the same period are necessary:

- From the teacher ⇒ Teacher's mark reported in the second year of upper secondary school ;
- If From a standardized evaluation ⇒ Reading score in OCSE-PISA 2009.



Given the 5 PVs...

... five differences between the teacher's mark and each PVs are calculated in order to identify:

- Underestimated students ;
- Overestimated students;
- Students with teacher's judgment coherence with the PISA data;

INTRODUCTION THEORETICAL FRAMEWORK METHODOLOGICAL ISSUES Res 0 000 000 000 000 000 000

TABLE OF DIFFERENCES: UNDERESTIMATED

| D | oiff. PV1 | Diff. PV2 | Diff.PV3 | Diff.PV4 | Diff.PV5 | | |
|---|-----------|-----------|----------|----------|----------|---|----------------|
| | - | - | - | - | - | ⇒ | Underestimated |
| | - | - | + | + | - | | Coherent |
| | + | + | + | + | + | | Overestimated |
| | - | + | - | + | - | | Coherent |
| | - | - | + | - | - | | Coherent |
| | + | + | + | + | + | | Overestimated |

| Diff. PV1 | Diff. PV2 | Diff.PV3 | Diff.PV4 | Diff.PV5 | | |
|-----------|-----------|----------|----------|----------|---|----------------|
| - | - | - | - | - | ⇒ | Underestimated |
| - | - | + | + | - | | Coherent |
| + | + | + | + | + | ⇒ | Overestimated |
| - | + | - | + | - | | Coherent |
| - | - | + | - | - | | Coherent |
| + | + | + | + | + | | Overestimated |

RODUCTION THEORETICAL FRAMEWORK METHODOLOGICAL ISSUES RI

SULTS

References

TABLE OF DIFFERENCES: COHERENT

| Di | ff. PV1 | Diff. PV2 | Diff.PV3 | Diff.PV4 | Diff.PV5 | | |
|----|---------|-----------|----------|----------|----------|---|----------------|
| [| - | - | - | - | - | ⇒ | Underestimated |
| | - | - | + | + | - | | Coherent |
| | + | + | + | + | + | | Overestimated |
| | - | + | - | + | - | | Coherent |
| | - | - | + | - | - | | Coherent |
| | + | + | + | + | + | | Overestimated |

INTRODUCTION

THEORETICAL FRAMEWORK

METHODOLOGICAL ISSUES

Results 000000 REFERENCES

MULTILEVEL APPROACH

A multilevel approach has been employed given the hierarchical nature of the data.

Student characteristics

- Gender
- Immigration status
- Cultural and socio-economic status (ESCS)
- Student repeating a year
- Ability level

School characteristics

- All variables considered at the student level have been aggregated and included at the school level
- Type of secondary school
- School size
- School location (big or small city)
- Teachers expectations

• • • • • • • • •

ESCS

Variable created by OCSE on the basis of the occupational and educational level of student' parents, family wealth, home educational and cultural resources

The Ability Level

- The Final Evaluation of the lower secondary education;
- Mark reported in the first year of upper secondary school.

The ability level is found by a Partial Credit Model (Rasch Analysis).

| INTRODUCTION O | THEORETICAL FRAMEWORK | Methodological Issues | Results | REFERENCES |
|-------------------|-----------------------|-----------------------|---------|------------|
| CORRELA | ATIONS | | | |

Correlations:

- 0.62 = Underestimated & Coherent marks
- 0.65 = Overestimated & Coherent marks
- 0.38 = General level of assessment and OCSE-PISA student score

Ranges from 0.28 to 0.92 in the literature the correlation found between these measures.

| INTRODUCTION | THEORETICAL FRAMEWORK | METHODOLOGICAL ISSUES | RESULTS | References |
|--------------|-----------------------|-----------------------|---------|------------|
| 0 | 000 | 00000000 | 000000 | |
| Datura | | | | |
| BOXPLO | T | | | |



◆□▶ ◆□▶ ◆ □▶ ◆ □▶ ● □ ● ● ● ●

| INTRODUCTION 0 | THEORETICAL FRAMEWORK | METHODOLOGICAL ISSUES | Results 000000 | REFERENCES |
|-------------------|-----------------------|-----------------------|-------------------|------------|
| TABLE | | | | |

| | Variables | Overestimation | Underestimation |
|--------|--|----------------|-----------------|
| | Intercept | 0.842 | -0.595 |
| | Ability | 0.084** | -0.059 |
| ÷ | Escs | -0.107 | -0.217** |
| studen | Student repeating the year (ref. Student not repeating the year) | 0.486* | -0.402 |
| | Female (ref. Male) | -0.424* | -0.297 |
| | Immigrant (ref. Italian) | 0.004 | -0.049 |
| | Mean Ability | -0.147* | -0.081 |
| | Mean ESCS | -0.734* | 1.294*** |
| | Percentage of students not repeating the year | -1.802** | -1.401 |
| _ | Percentage of girls | 0.000 | 0.010** |
| school | Percent of immigrants | 3.193*** | -1.800 |
| 01 | Vocational studies | 0.094 | 1.165* |
| | Technical institute | 0.213 | 0.263 |
| | School size | -0.310*** | 0.305*** |
| | Teacher Behavior | 0.342** | 0.033 |
| | Variance | 0.70 | 0.39 |

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 - のへで

| INTRODUCTION 0 | THEORETICAL FRAMEWORK | Methodological Issues | Results | References |
|-------------------|-----------------------|-----------------------|---------|------------|
| RESULTS | | | | |

Overestimated student

At student level

- Ability
- Being a student repeating a year
- Female (ref.male)

At school level

- Mean Ability
- ESCS index
- Percentage of student repeating a year
- Percentage of immigrant
- School dimension
- Teacher' behavior

Underestimated student

At student level

Escs

At school level

- Escs index
- Percentage of girls
- School dimension

| INTRODUCTION O | THEORETICAL FRAMEWORK | Methodological Issues | Results | References |
|-------------------|-----------------------|-----------------------|---------|------------|
| FINAL R | ESULTS | | | |

From both underestimation and overestimation results, it is possible to make some observations:

Escs

- Students attending technical and especially vocational schools (lower mean ESCS) have both less chances of being underestimated and more chance of being overestimated than students attending "liceo" higher mean ESCS.
- Students attending "liceo" are the ones most likely to be underestimated and the less likely of being overestimated.

School size

The high significance of school size both on underestimation and overestimation needs a closer examination.

THIS WORK SUGGESTS:

- Overestimation appear to be a more complex phenomena than underestimation: a higher number of significant variables are implied;
- Teachers have a tendency to overestimate students with high abilities and more disadvantaged conditions.

| INTRODUCTION 0 | THEORETICAL FRAMEWORK | Methodological Issues | Results 000000 | REFERENCES |
|-------------------|-----------------------|-----------------------|-------------------|------------|
| Some Re | FERENCES | | | |

- Bond, T.G., Fox, C.M.: Applying the Rasch Model: Fundamental Measurement in the Human Sciences. London: Lawrence Erlbaum Associates Publishers (2001).
- Cipollone, P., Sestito, P., *Il capitale umano. Come far fruttare i talenti*, Il Mulino, Bologna (2010).
- Feinberg, A.B., Shapiro, E.S., *Teacher accuracy: An examination of teacher-based judgments of students reading with differing achievement levels*, Journal of Educational Research, 102, 453-462 (2009).

Hoge, R.D., Coladarci, T., *Teacher-based judgments of academic achievement: A review of literature*, Review of Educational Research, 59, 297-313 (1989).

- OCSE-PISA: PISA 2009 Technical Report (Preliminary version), available at http://www.oecd.org/dataoecd/13/34/48578536.pdf.
 - Snijders, T., Bosker., R.: Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling. Sage Publications Ltd (1999)