# What we can and cannot learn from international student assessments?

Maciej Jakubowski and Tomasz Gajderowicz

EVIDENCE INSTITUTE

UNIVERSITAS VARSOVIENSIS

UNIWERSYTET WARSZAWSKI
Wydział Nauk Ekonomicznych

EVIDENCE INSTITUTE

# Agenda

- ✓ What are international large-scale assessments? Are they very different?
- ✓ Key results. Can we trust them?
- ✓ Examples of important evidence-based lessons
- ✓ How reliable are non-cognitive data?
- ✓ Can we trust research based on ILSA data?

# ILSA historically

**before 1990:**

FIMS 1964

FISS 1970

SIMS 1980

SISS 1983

**1990s:**

Reading L

TIMSS 1995

IALS

CIVIC

**2000s:**

PIRLS 2001 2006

TIMSS 2003 2007

TALIS 2008

TED S-M

ICCS 2009

PISA 2000 2003 2006 2009

**2010s**

TALIS 2013 2018

ESLC 2012

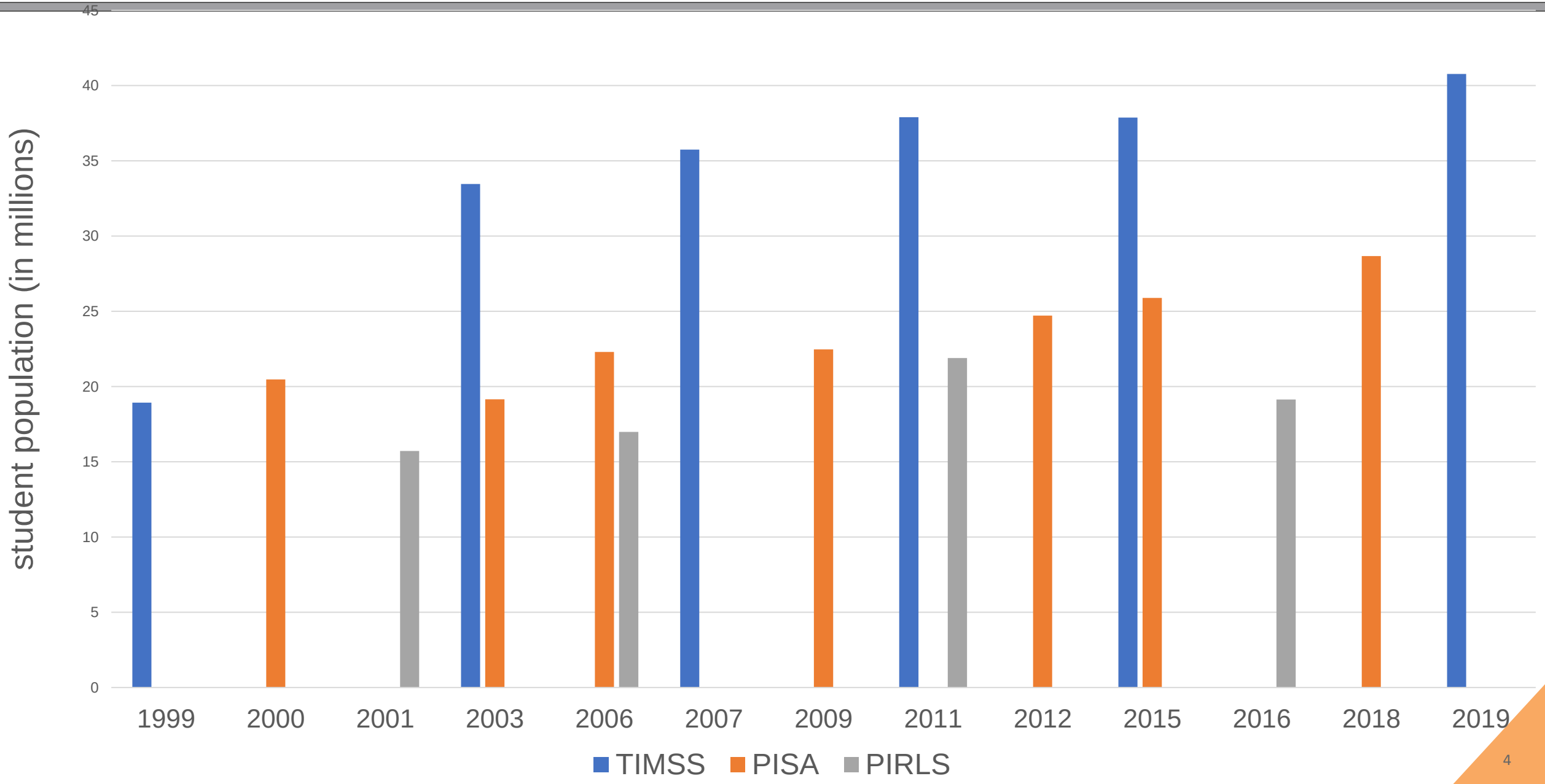
ICILS 2013 2018

TIMSS 2011 2015 2019

PIRLS 2011 2016

PISA 2012 2015 2018

# Student population represented in ILSA



EVIDENCE INSTITUTE

# TIMSS, PISA and PIRLS – are they different

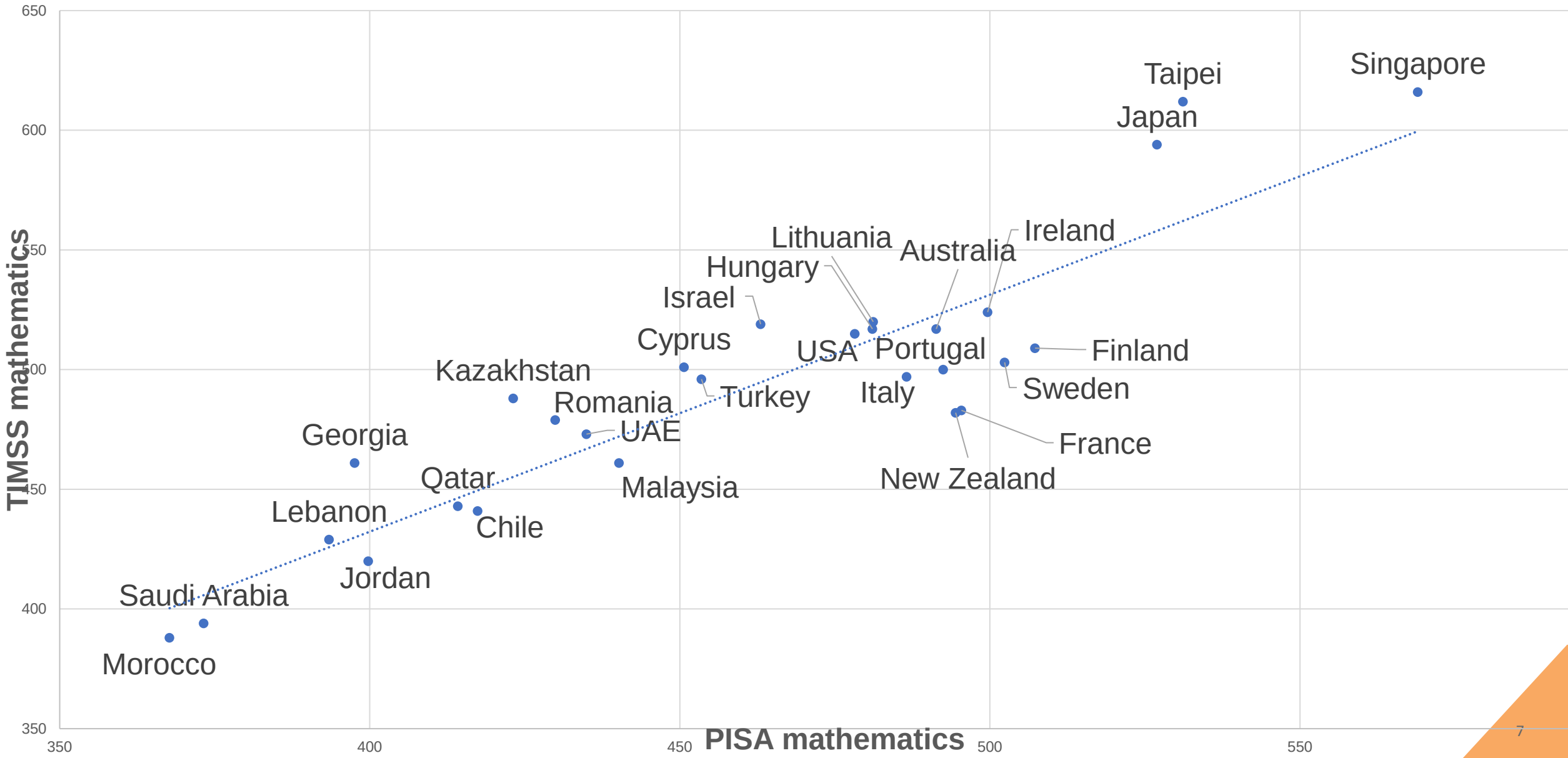
| | TIMSS | PISA | PIRLS |
|---|---|---|---|
| **Subjects/domains** | Mathematics and science | Reading, mathematics, science… but also global comptences, problem solving, financial literacy, creativity | Reading |
| **Assessment framework** | „Internationally agreed curriculum” | „ability to use knowledge and skills usefulto meet real-life challenges” | „broad notion of what an ability to read is” |
| **Age/grade** | 4th – 10-year-olds<br>8th – 14-year-olds | 15-year-olds<br>9th grade modal | 4th – 10-year-olds |
| **Sampling** | School->class->student | School->student | School->class->student |
| **Who is covered?** | Students, parents, teachers, principals, and experts | Students and principals;<br>Parents/teachers (optional) | Students, parents, teachers, and principals |
| **Statistical methods** | CB adaptive/branched test, conditional PVs and 3PL IRT model, replicate weights | | |
| **Student population** | 41 milion (TIMSS 2019) | 29 milion (PISA 2018) | 19 million (PIRLS 2016) |

# TIMSS, PISA and PIRLS – are they very different?

## *Reading assessment framework in PISA and PIRLS*

| ***PISA 2009*** | ***PIRLS 2006*** |
|---|---|
| *Access and retrieve* | *Focus on and retrieve explicitly stated information* |
| *Integrate and interpret* | *Make straightforward inferences* |
| | *Interpret and integrate ideas and information* |
| *Reflect and evaluate* | *Examine and evaluate content, language, and textual elements* |

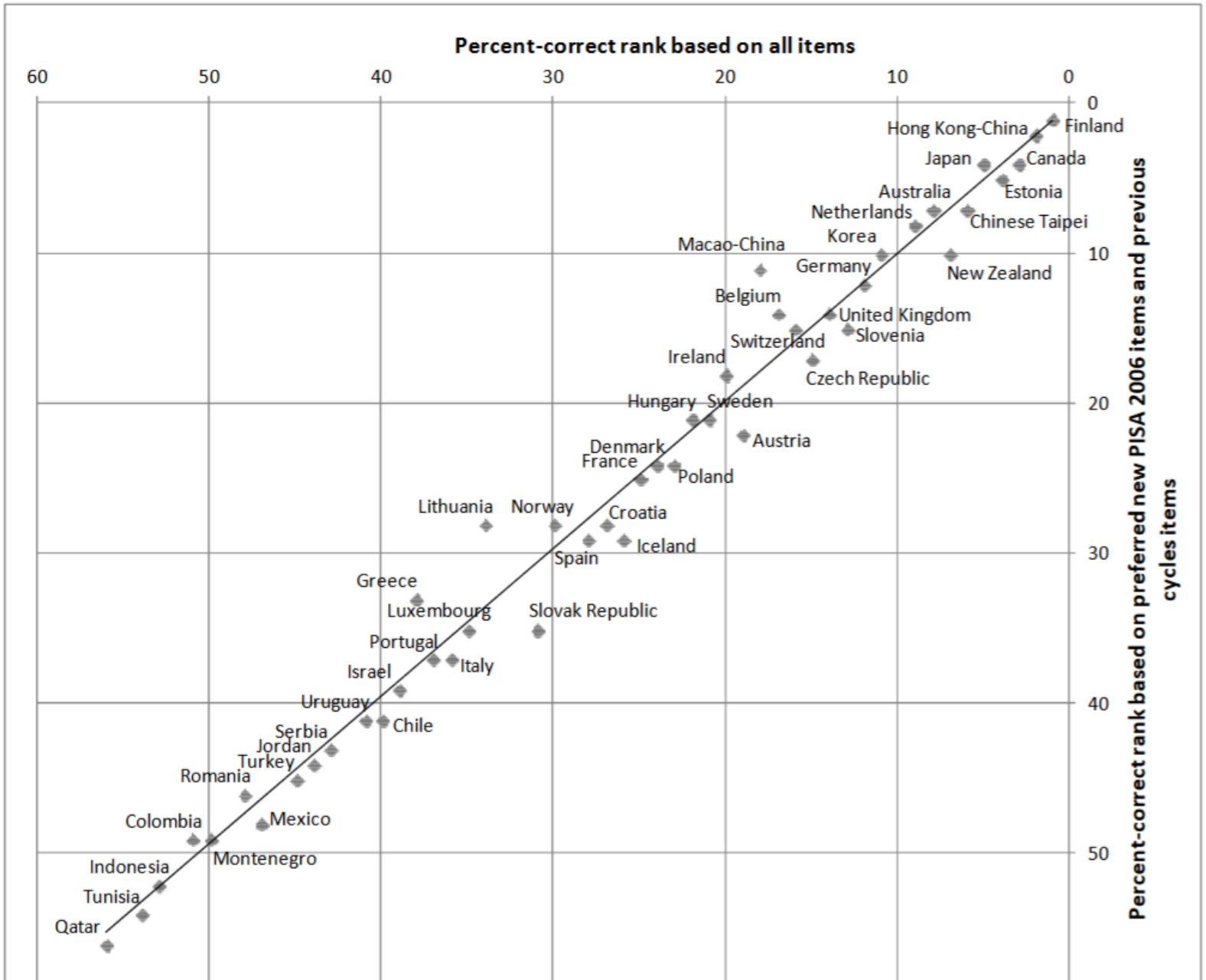
# Correlation of mean achievement in PISA and TIMSS mathematics



# Key results

- ***Achievement comparisons***
- ***Inequality measures***
- ***Achievement trends***
- ***Comparisons of student groups with similar characteristics***
- ***Associations/causal relations between structural choices in education systems and achievement***

**Figure 1. Percent-correct ranking based on all PISA 2006 science items and the percent-correct ranking based on countries' preferred items plus the items kept from the previous PISA cycles**




OECD Education Working Papers
No. 46

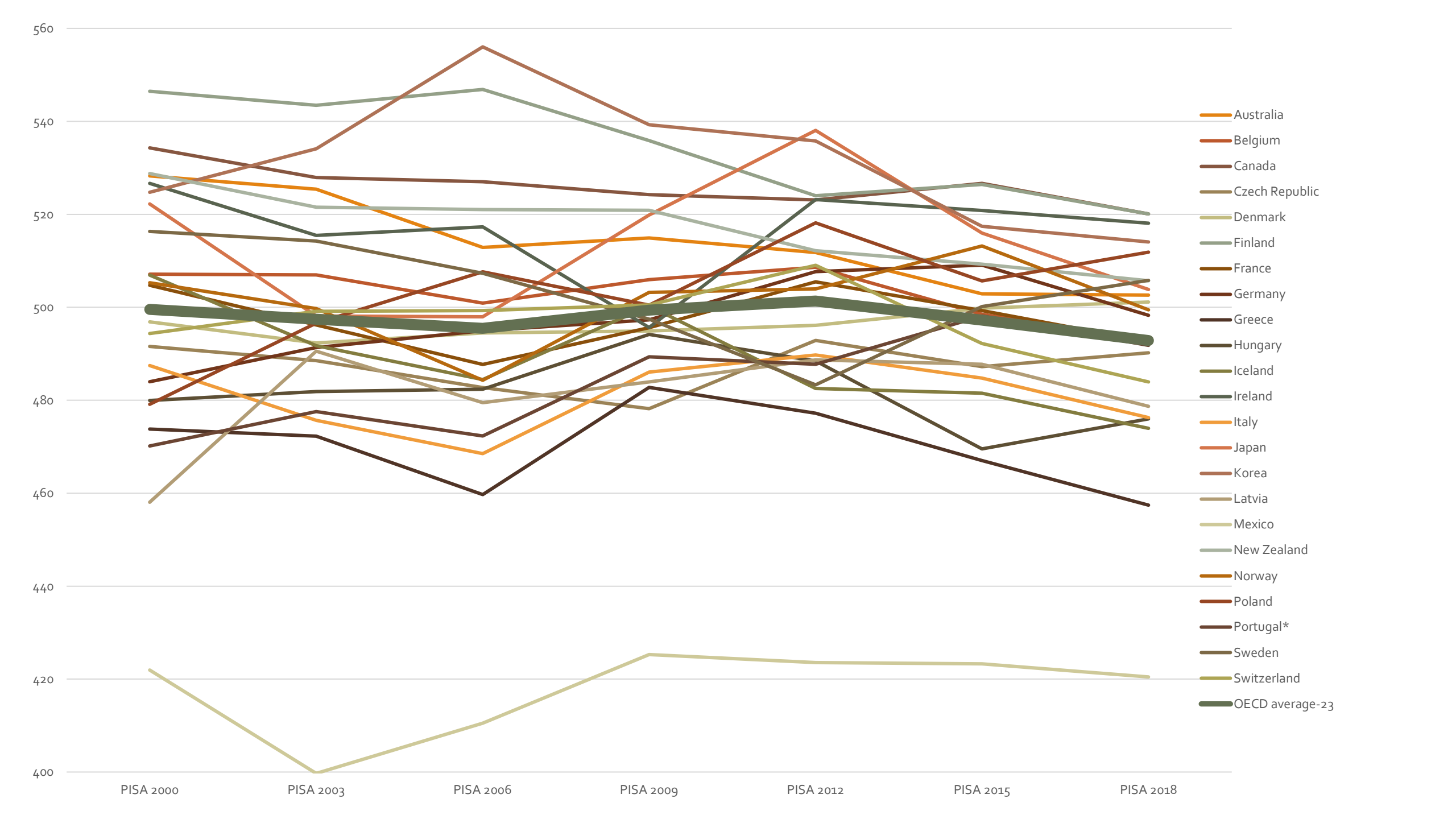
# Analysis of PISA 2006 Preferred Items Ranking Using the Percent-Correct Method

Ray Adams, Alla Berezner,
Maciej Jakubowski

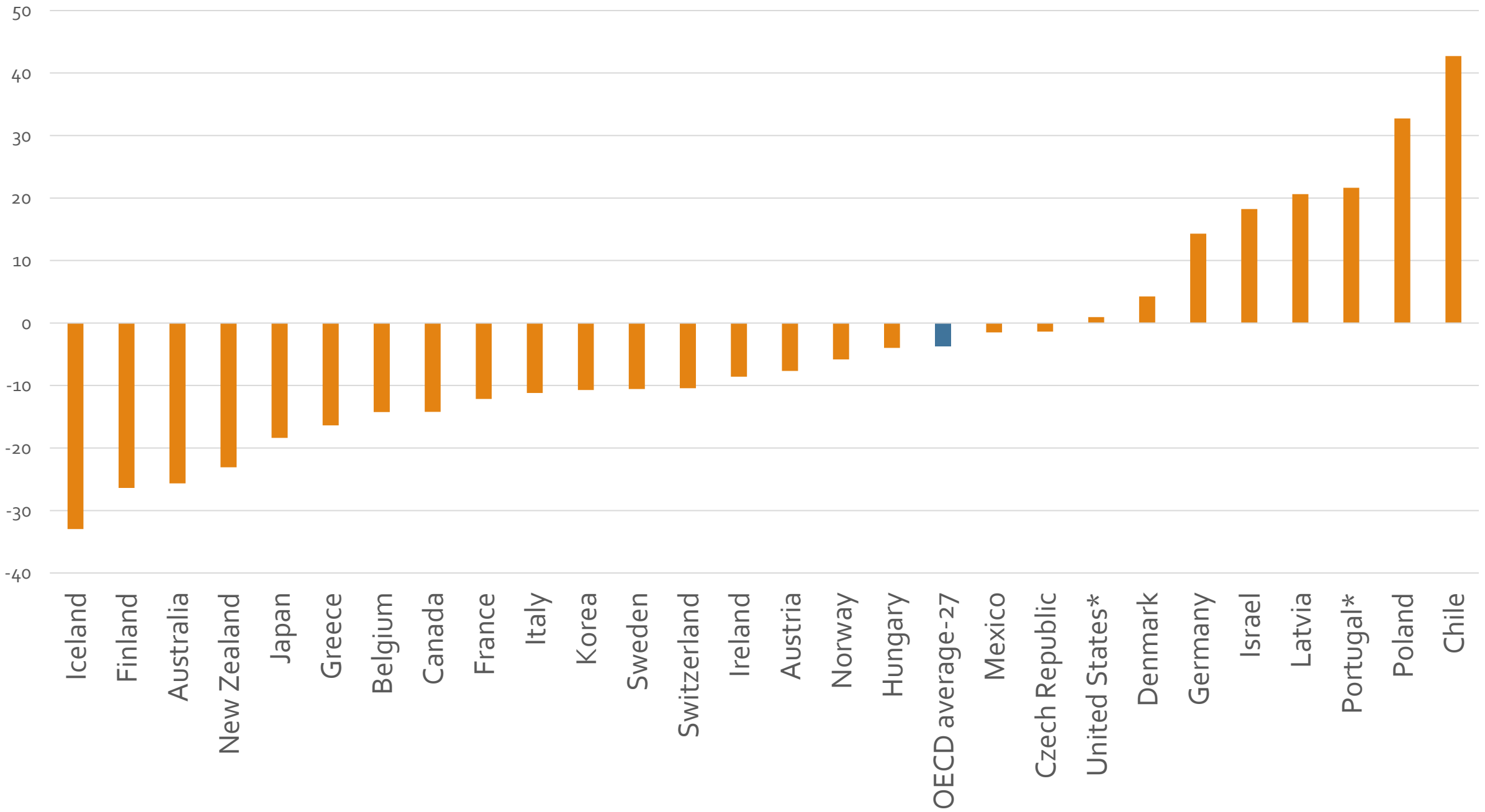
| Country | Average Mathematics Scale Score based on All Items | Norway (5) | Lithuania | Austria | Netherlands | United States | Czech Republic | Belgium (Flemish) | Cyprus | Finland | Denmark |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Singapore | 625 (3.9) | 625 (3.9) | 619 (3.8) | 620 (3.8) | 617 (3.8) | 625 (3.9) | 618 (3.8) | 624 (3.9) | 624 (3.9) | 625 (3.9) | 624 (3.8) |
| Hong Kong SAR | 602 (3.3) | 602 (3.3) | 601 (3.3) | 600 (3.3) | 598 (3.5) | 602 (3.4) | 598 (3.3) | 600 (3.3) | 602 (3.4) | 601 (3.3) | 601 (3.3) |
| Korea, Rep. of | 600 (2.2) | 599 (2.2) | 599 (2.2) | 598 (2.2) | 592 (2.2) | 599 (2.2) | 594 (2.2) | 603 (2.3) | 596 (2.2) | 601 (2.2) | 601 (2.2) |
| Chinese Taipei | 599 (1.9) | 598 (1.9) | 596 (1.9) | 597 (2.0) | 593 (2.0) | 599 (1.9) | 595 (2.0) | 599 (2.0) | 594 (2.0) | 601 (2.0) | 599 (1.9) |
| Japan | 593 (1.8) | 590 (1.8) | 595 (1.7) | 593 (1.8) | 590 (1.7) | 592 (1.8) | 591 (1.7) | 594 (1.7) | 587 (1.8) | 596 (1.8) | 594 (1.8) |
| Russian Federation | 567 (3.3) | 566 (3.4) | 571 (3.4) | 570 (3.4) | 569 (3.3) | 567 (3.4) | 571 (3.4) | 566 (3.3) | 567 (3.4) | 566 (3.3) | 567 (3.4) |
| Northern Ireland | 566 (2.7) | 567 (2.8) | 564 (2.7) | 564 (2.7) | 567 (2.8) | 566 (2.8) | 565 (2.8) | 566 (2.7) | 567 (2.8) | 565 (2.8) | 565 (2.7) |
| England | 556 (3.0) | 556 (3.0) | 553 (3.0) | 554 (2.9) | 555 (3.0) | 556 (3.0) | 552 (3.0) | 556 (3.0) | 555 (3.1) | 555 (3.0) | 556 (3.0) |
| Ireland | 548 (2.5) | 549 (2.5) | 545 (2.4) | 546 (2.4) | 547 (2.5) | 549 (2.5) | 546 (2.5) | 549 (2.5) | 549 (2.5) | 548 (2.5) | 548 (2.5) |
| Latvia | 546 (2.6) | 546 (2.6) | 551 (2.7) | 548 (2.6) | 552 (2.6) | 546 (2.6) | 550 (2.6) | 546 (2.6) | 548 (2.7) | 545 (2.6) | 546 (2.6) |
| Norway (5) | 543 (2.2) | 543 (2.2) | 542 (2.2) | 542 (2.2) | 544 (2.2) | 542 (2.2) | 541 (2.2) | 543 (2.2) | 543 (2.2) | 542 (2.2) | 544 (2.2) |
| Lithuania | 542 (2.8) | 542 (2.8) | 547 (2.8) | 543 (2.8) | 545 (2.9) | 542 (2.8) | 545 (2.8) | 541 (2.7) | 543 (2.8) | 542 (2.8) | 542 (2.8) |
| Austria | 539 (2.0) | 540 (2.0) | 536 (2.1) | 543 (2.1) | 541 (2.1) | 538 (2.0) | 542 (2.1) | 539 (2.0) | 539 (2.0) | 538 (2.1) | 540 (2.0) |
| Netherlands | 538 (2.2) | 537 (2.2) | 541 (2.2) | 539 (2.3) | 544 (2.3) | 537 (2.2) | 538 (2.3) | 537 (2.2) | 537 (2.3) | 539 (2.1) | 539 (2.2) |
| United States | 535 (2.5) | 535 (2.5) | 532 (2.5) | 533 (2.5) | 533 (2.5) | 535 (2.5) | 533 (2.5) | 535 (2.5) | 536 (2.5) | 535 (2.5) | 534 (2.5) |
| Czech Republic | 533 (2.5) | 532 (2.6) | 533 (2.5) | 536 (2.6) | 532 (2.6) | 533 (2.5) | 537 (2.6) | 533 (2.5) | 533 (2.5) | 532 (2.5) | 533 (2.6) |
| Belgium (Flemish) | 532 (1.9) | 531 (1.9) | 529 (1.9) | 529 (1.9) | 528 (1.9) | 532 (1.9) | 529 (1.9) | 533 (1.9) | 531 (1.9) | 532 (1.9) | 533 (1.9) |
| Cyprus | 532 (2.9) | 533 (2.9) | 530 (2.8) | 529 (2.8) | 530 (2.9) | 532 (2.9) | 530 (2.8) | 531 (2.9) | 535 (2.9) | 531 (2.9) | 532 (2.9) |
| Finland | 532 (2.3) | 532 (2.4) | 531 (2.3) | 530 (2.3) | 533 (2.3) | 531 (2.3) | 530 (2.3) | 531 (2.4) | 532 (2.3) | 532 (2.4) | 532 (2.3) |
| Denmark | 525 (1.9) | 525 (2.0) | 524 (1.9) | 525 (1.9) | 527 (1.9) | 524 (1.9) | 522 (2.0) | 524 (1.9) | 524 (1.9) | 525 (1.9) | 526 (1.9) |

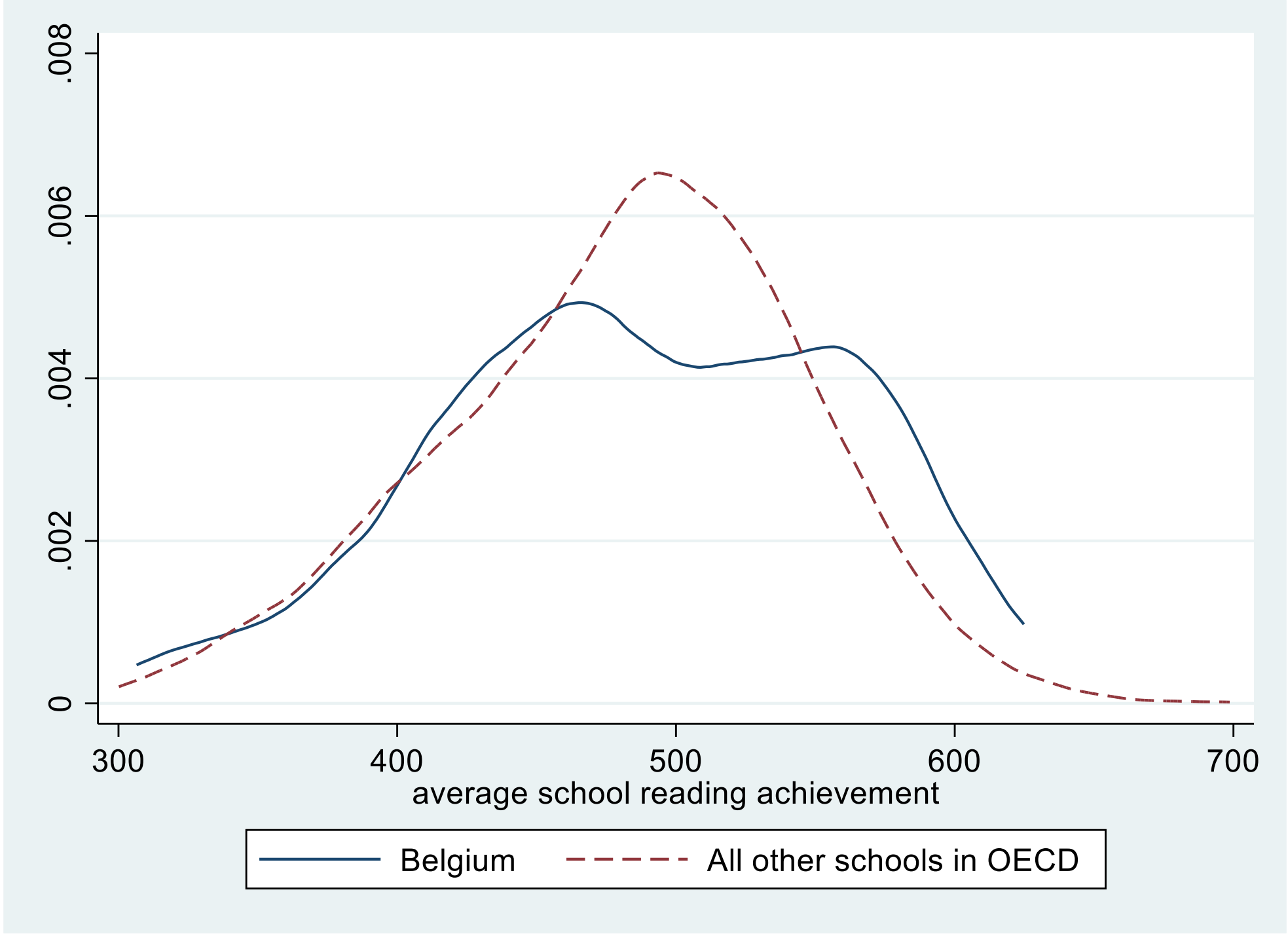
# Research on the comparability of ILSA

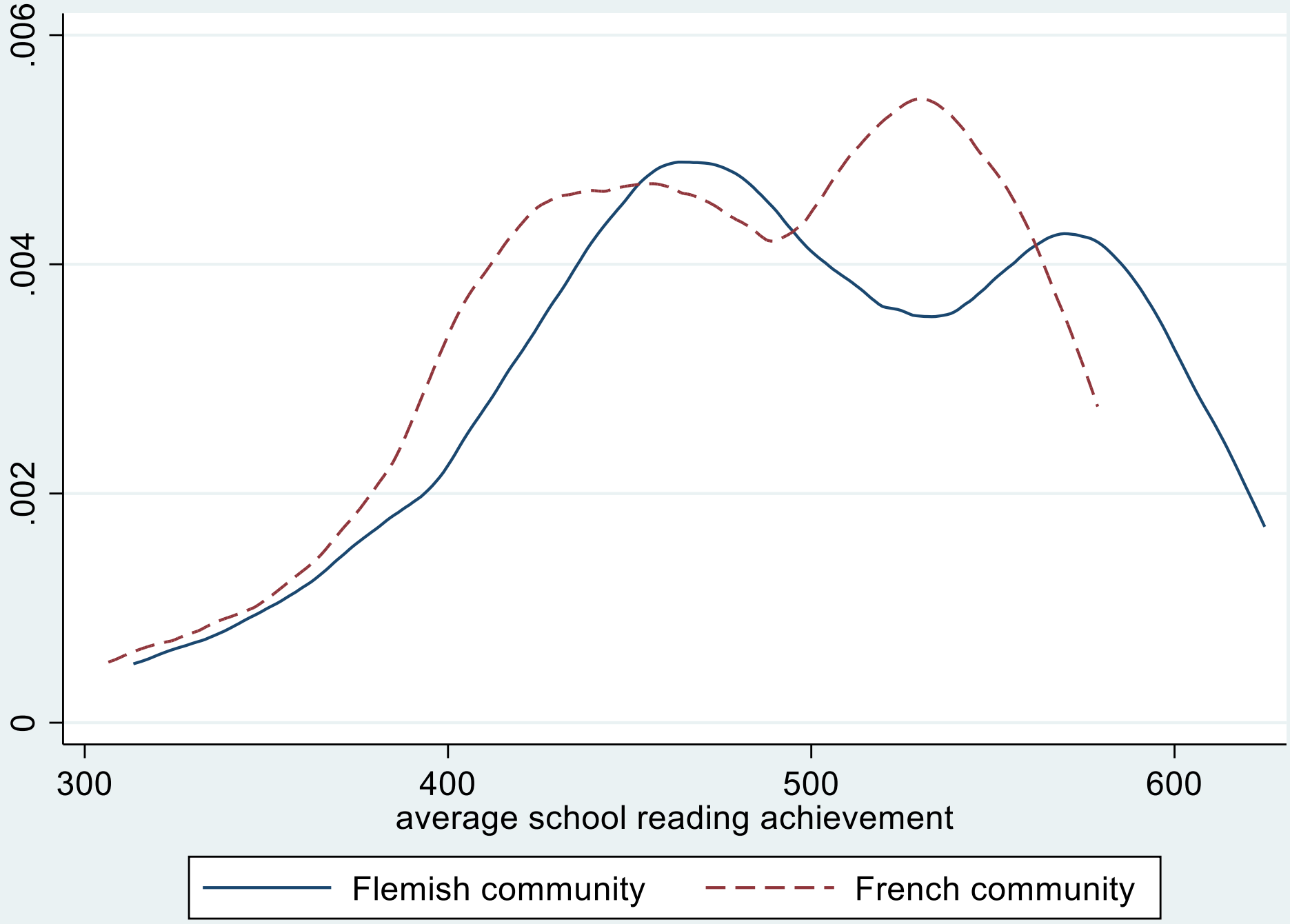
- ***Assessment frameworks***
- ***item selection***
- ***Sampling, coverage, and underlying populations***
- ***IRT scaling and plausible values***
- ***Non-cognitive scales***

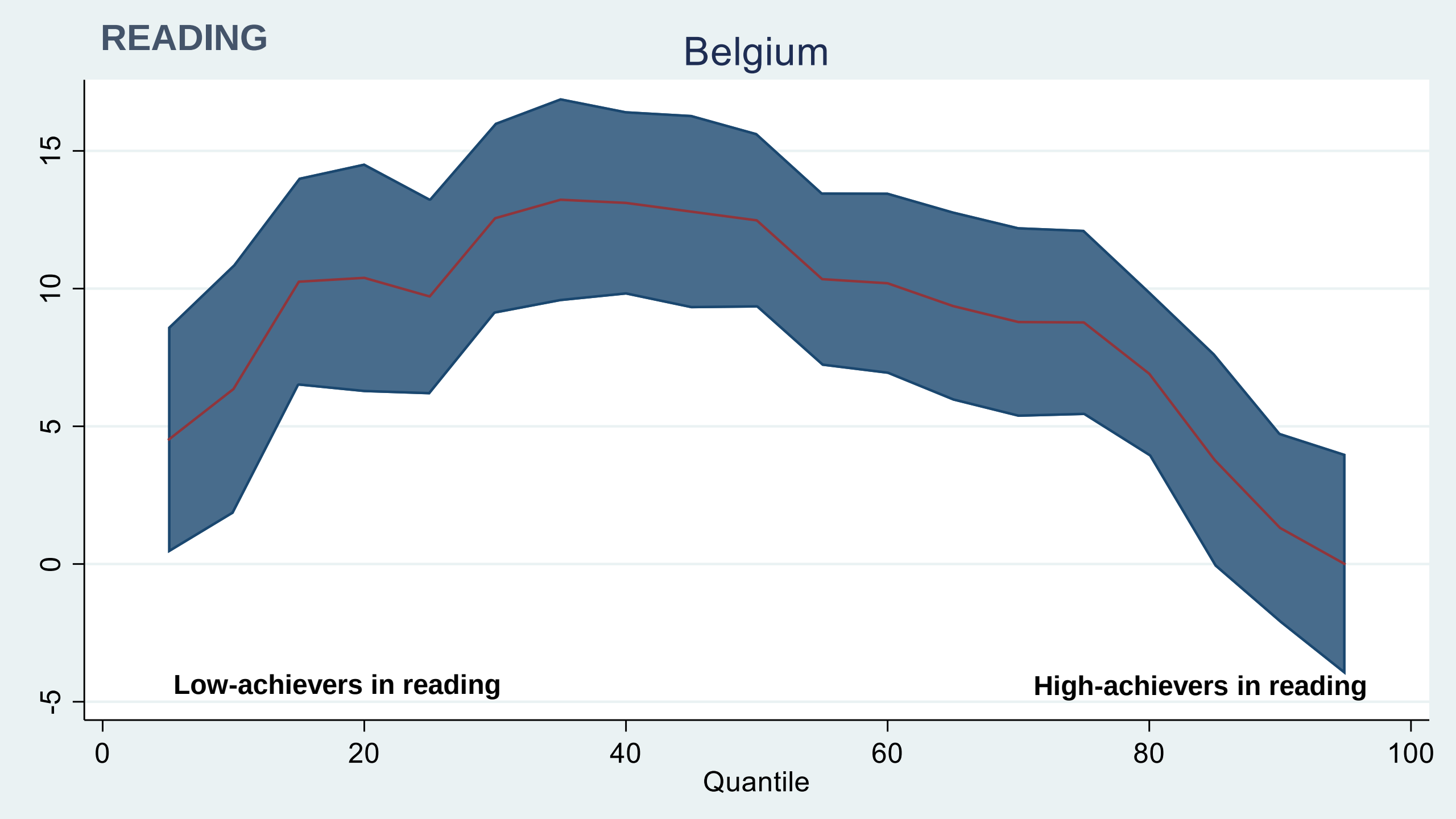

560
540
520
500
480
460
440
420
400
PISA 2000
PISA 2003
PISA 2006
PISA 2009
PISA 2012
PISA 2015
PISA 2018
Australia
Belgium
Canada
Czech Republic
Denmark
Finland
France
Germany
Greece
Hungary
Iceland
Ireland
Italy
Japan
Korea
Latvia
Mexico
New Zealand
Norway
Poland
Portugal*
Sweden
Switzerland
OECD average-23

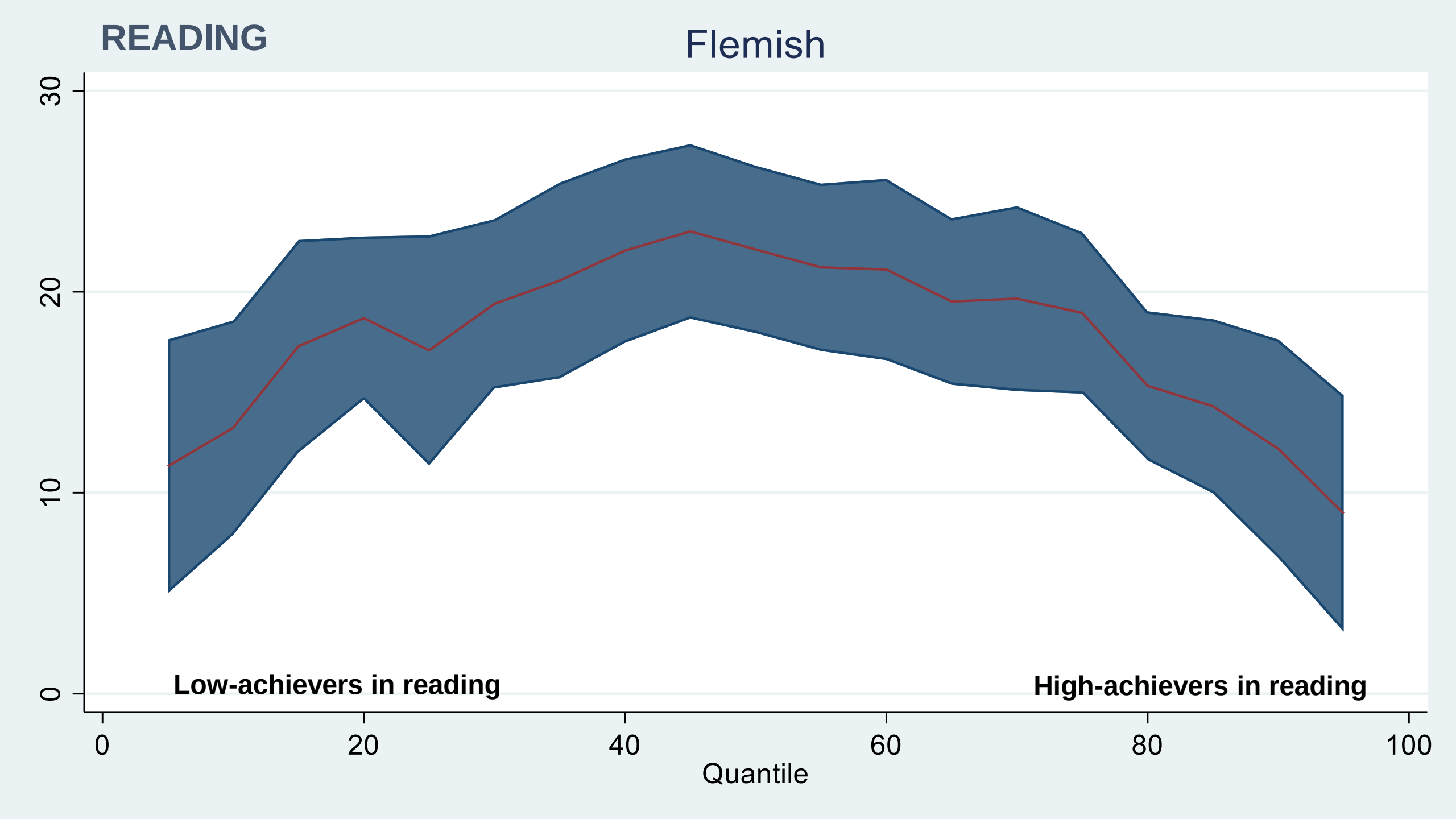
# Change in reading performance: PISA 2018 minus PISA 2000

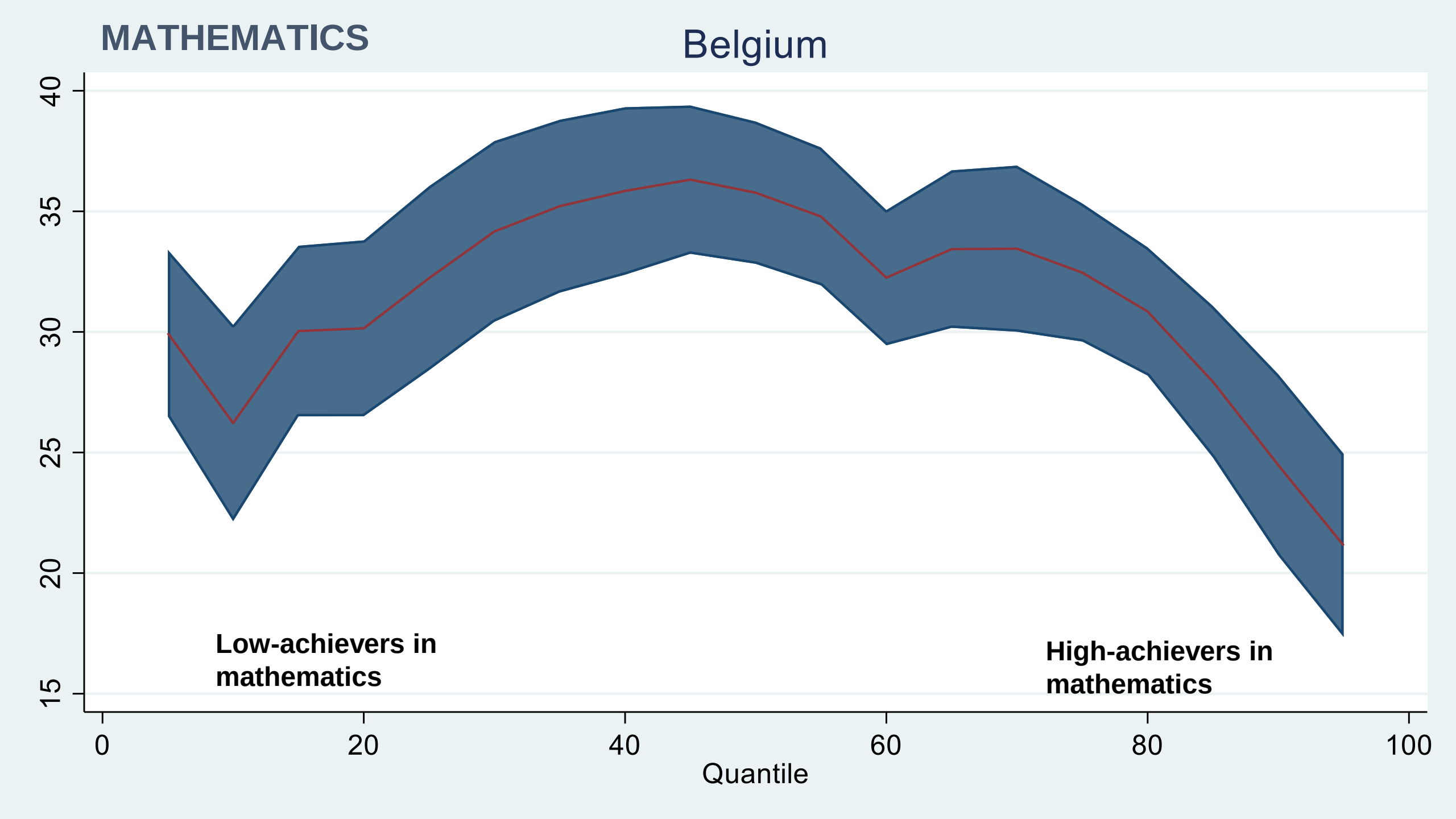


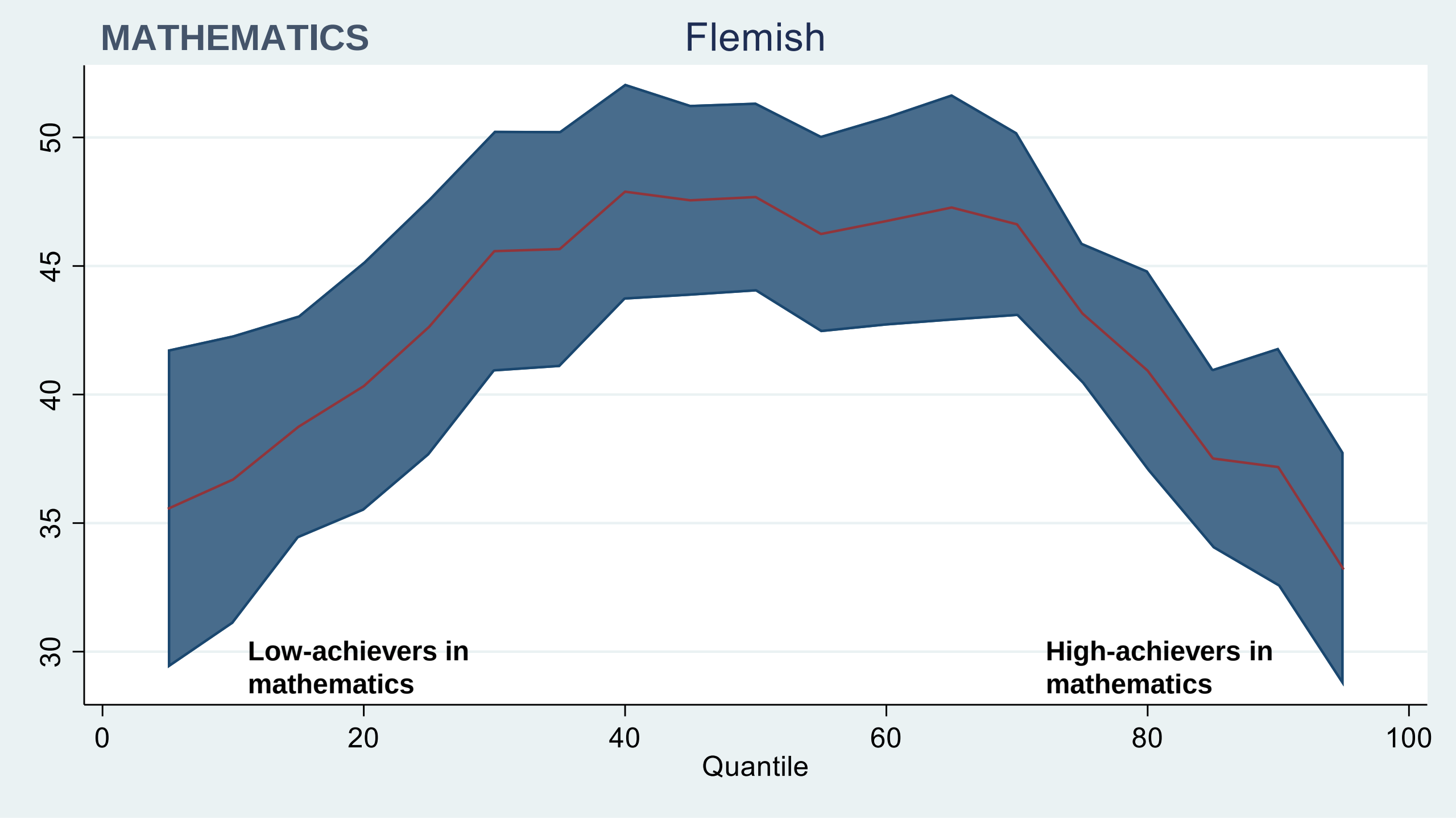

.008
.006
.004
.002
0
300
400
500
600
700
average school reading achievement
Belgium
All other schools in OECD


.006
.004
.002
0
300
400
500
600
average school reading achievement
Flemish community
French community


READING
Belgium
15
10
5
0
-5
Low-achievers in reading
High-achievers in reading
0
20
40
60
80
100
Quantile


READING
Flemish
30
20
10
0
Low-achievers in reading
High-achievers in reading
0
20
40
60
80
100
Quantile


MATHEMATICS
Belgium
40
35
30
25
20
15
Low-achievers in mathematics
High-achievers in mathematics
0
20
40
60
80
100
Quantile


MATHEMATICS
Flemish
50
45
40
35
30
Low-achievers in mathematics
High-achievers in mathematics
0
20
40
60
80
100
Quantile

Contents lists available at ScienceDirect

International Journal of Educational Development

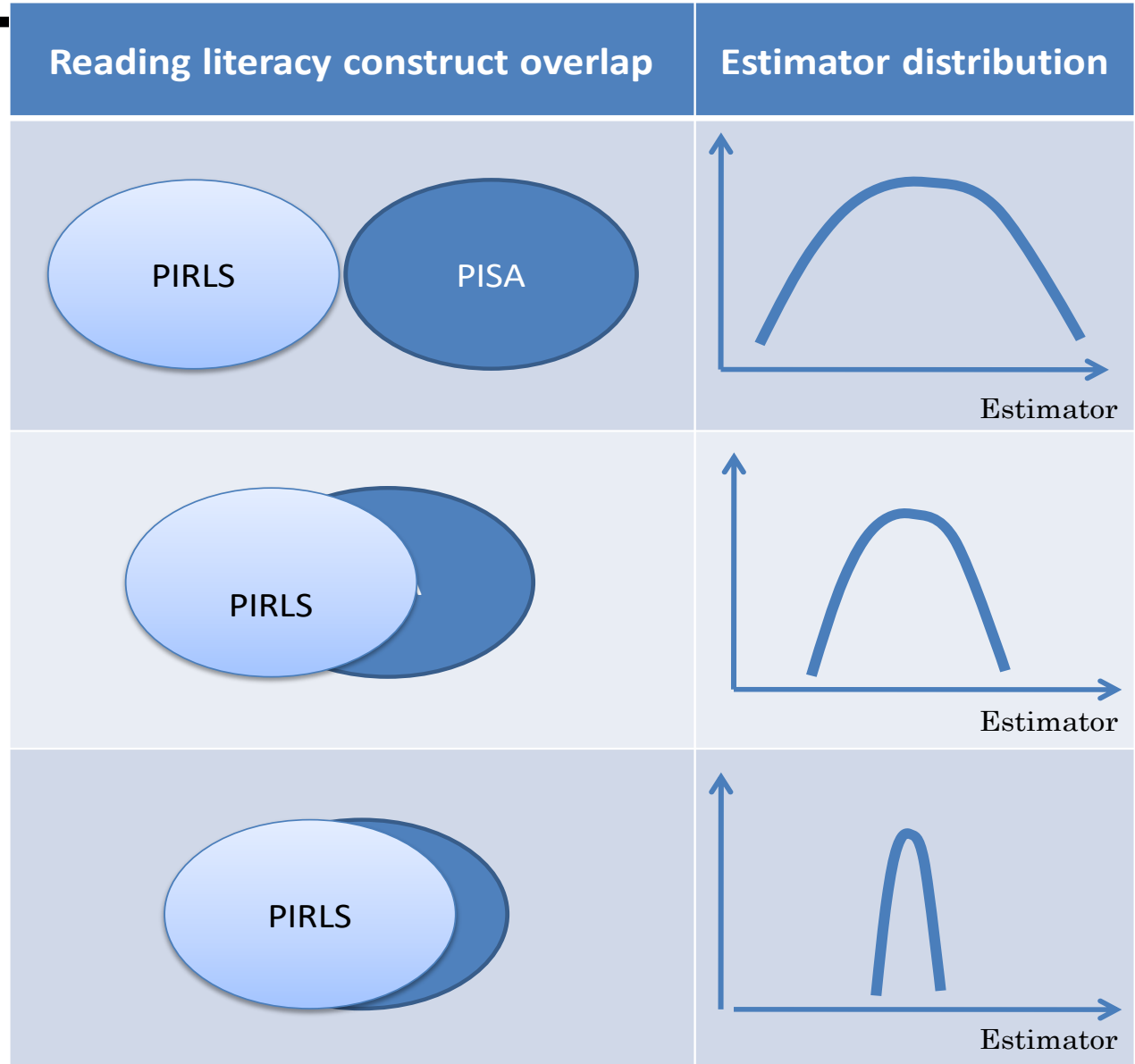
journal homepage: www.elsevier.com/locate/ijedudev

# Reading achievement progress across countries

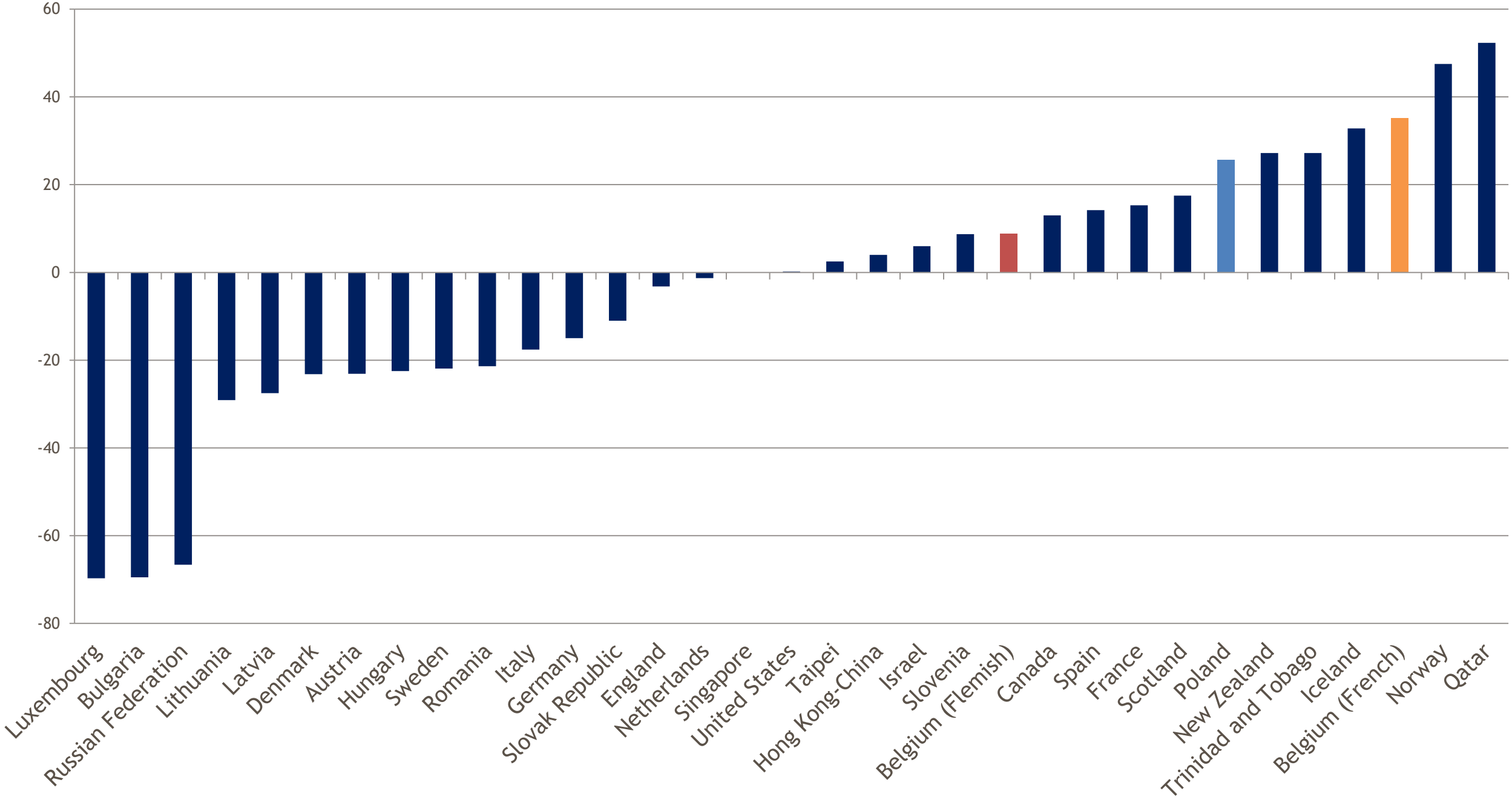
Maciej Jakubowski [a], Artur Pokropek [b,*]

[a] Faculty of Economic Sciences, Warsaw University, Poland
[b] Educational Research Institute (IBE), Poland

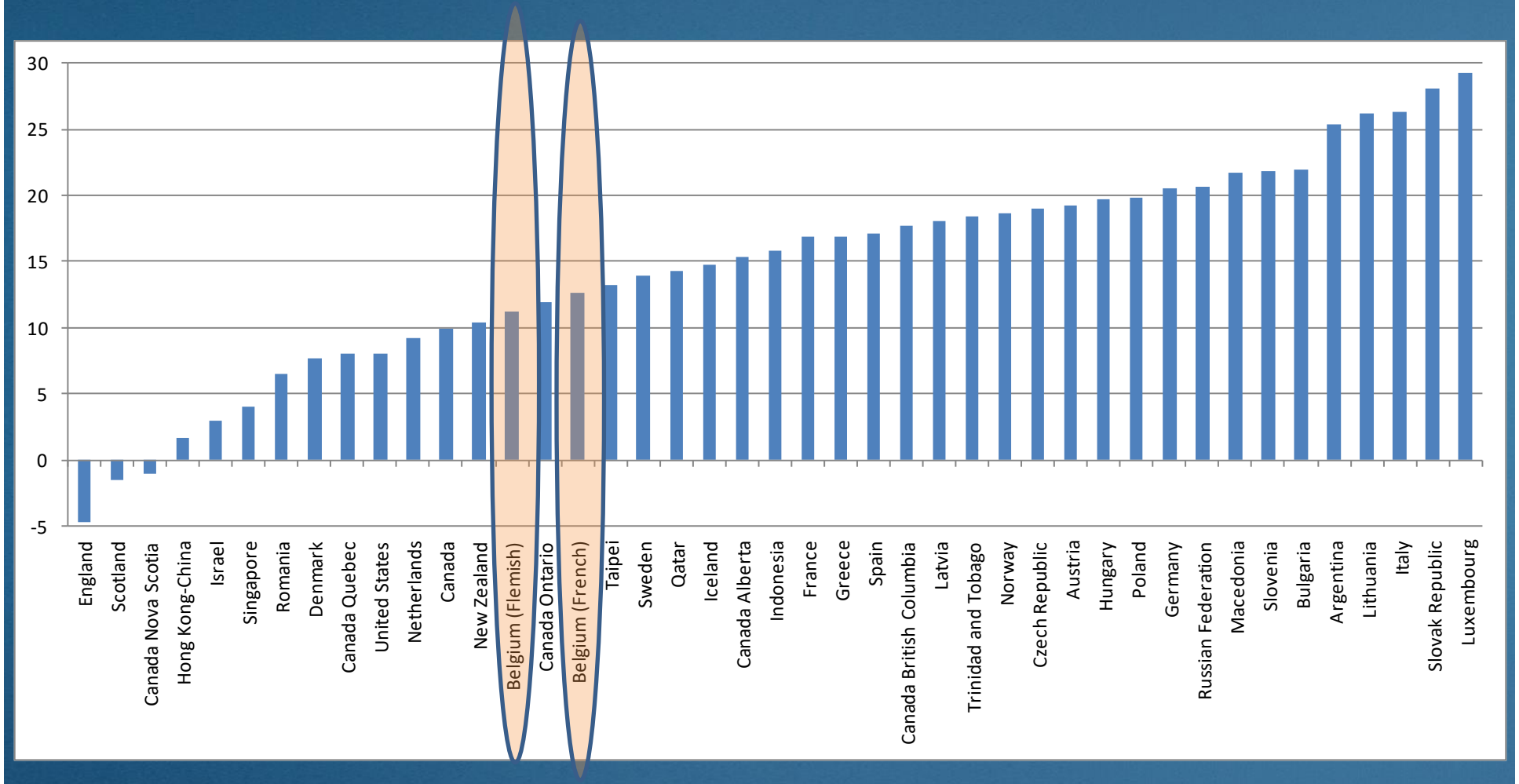
- Comparisons based on randomly taken reading assessment items from PIRLS 2006 and PISA 2009
- The same 3PL IRT model with conditional plausible values
- Re-weighting to adjust for demographic differences
- Regression correction for age effect



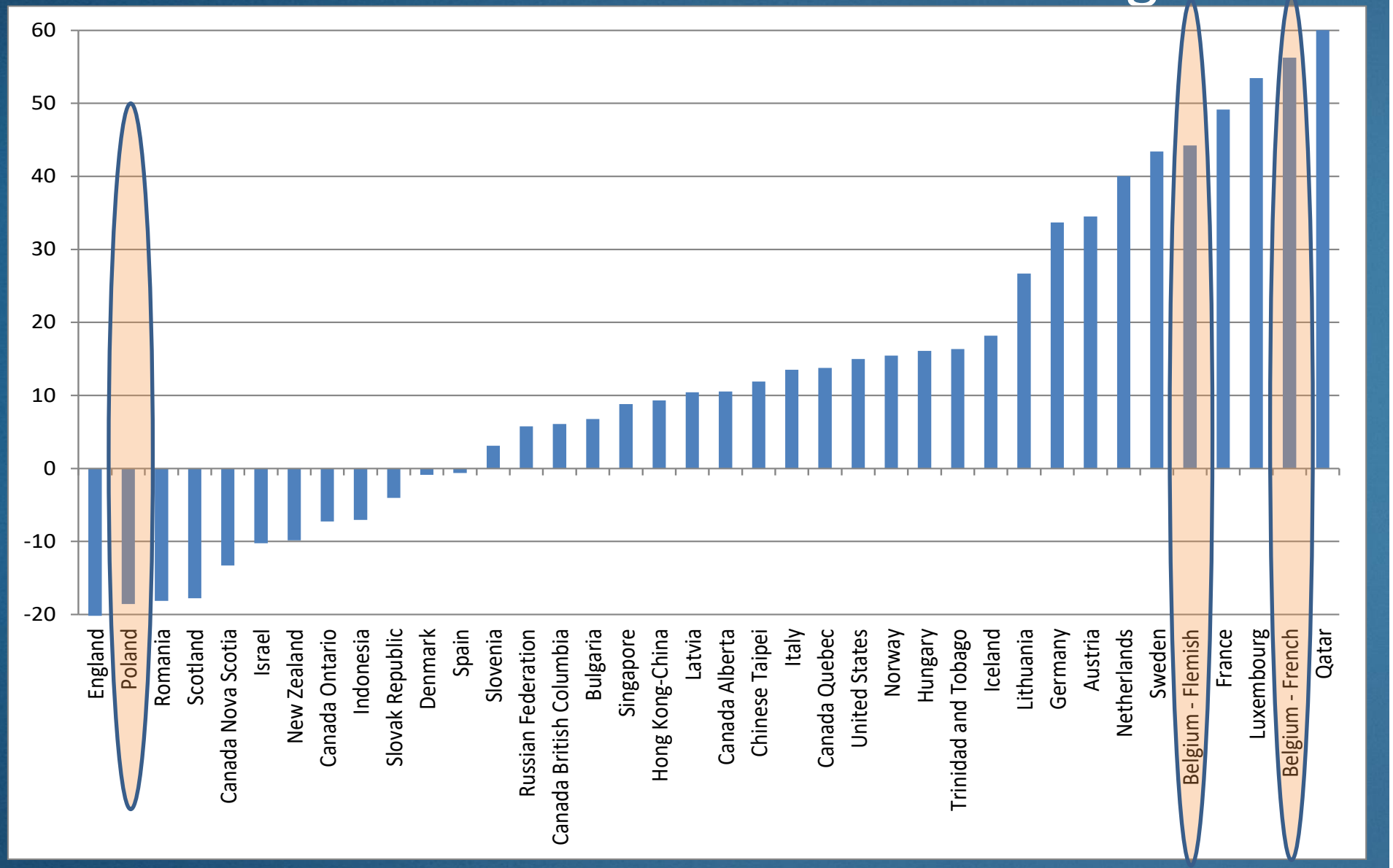
# Average reading achievement progress between 4th (PIRLS) and 9th grade (PISA)

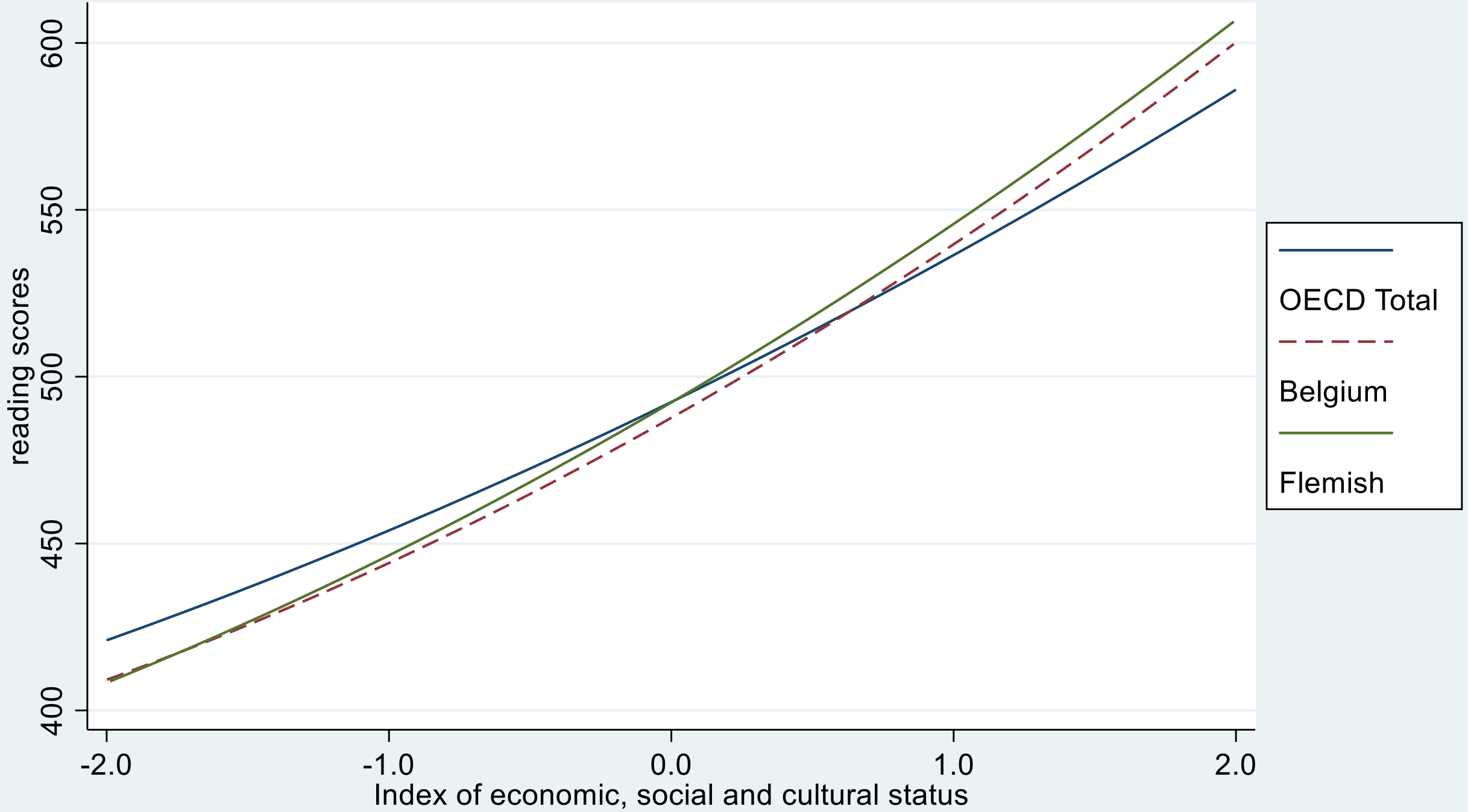


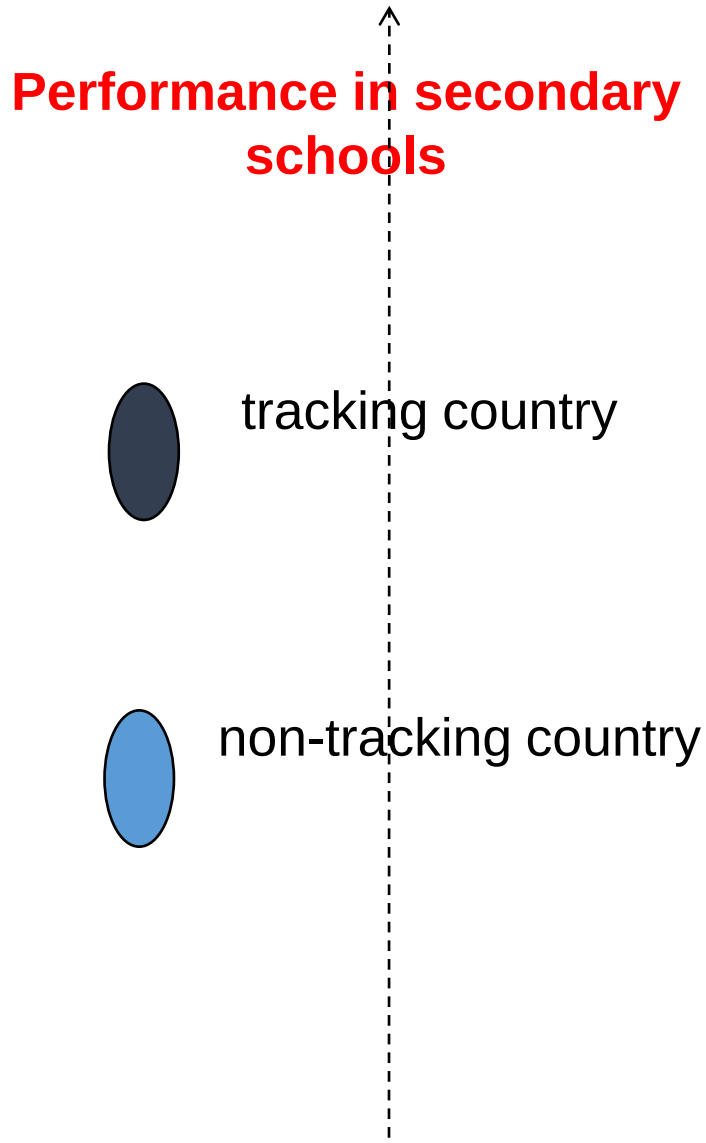
# Girls are progressing much faster

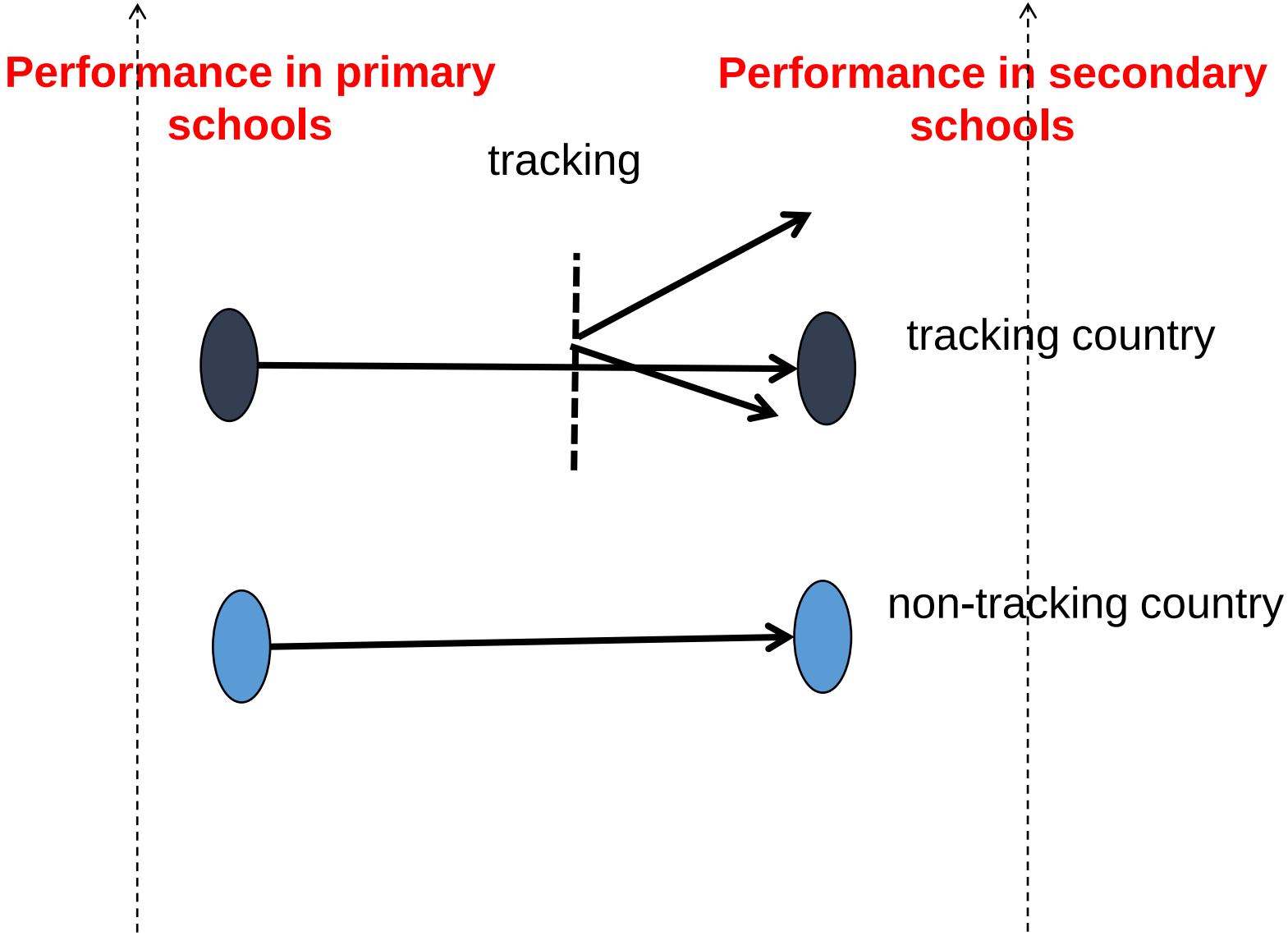


# Change in inequality of student reading achievement between 4th and 9th grade

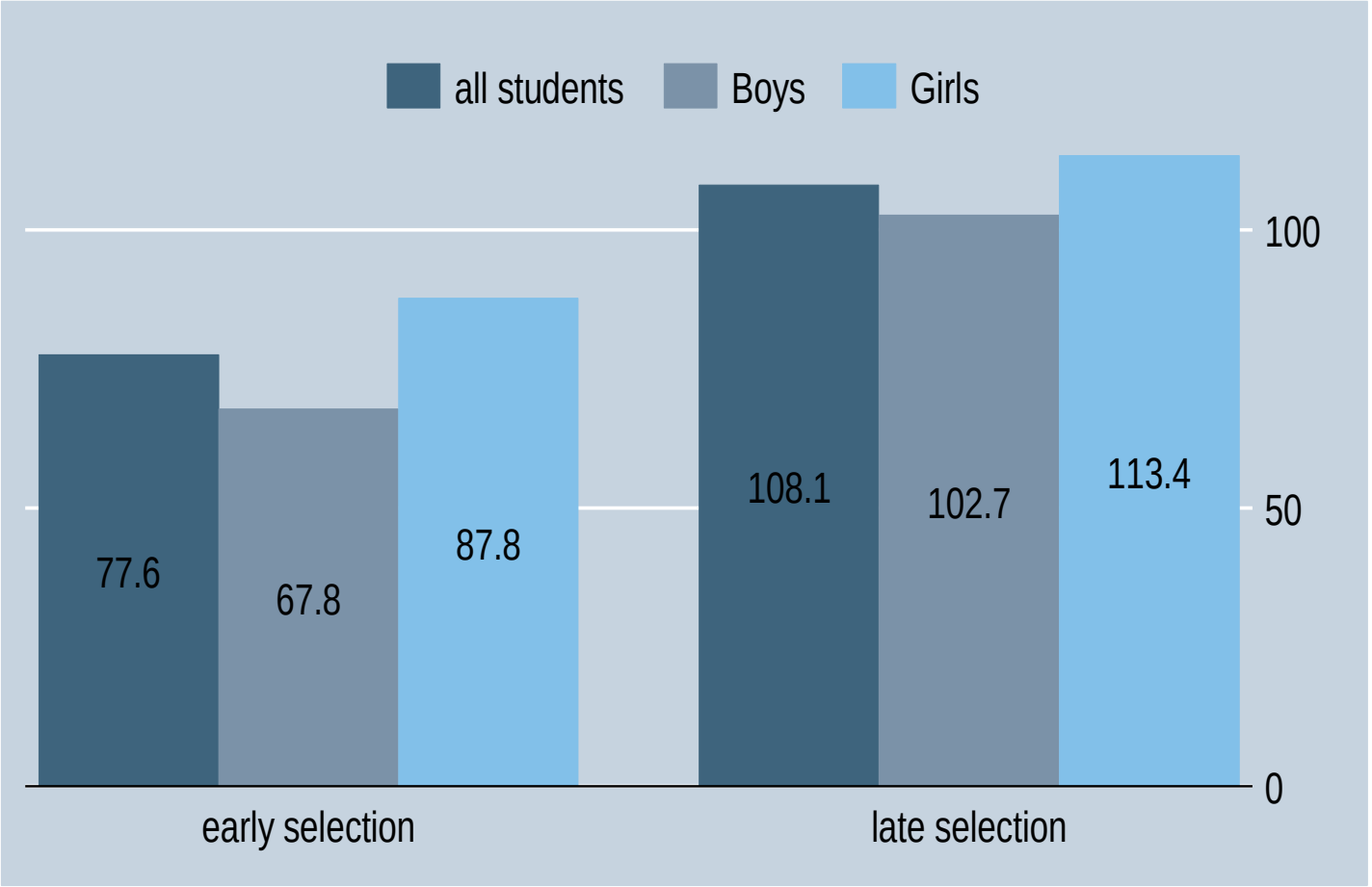



reading scores
400
450
500
550
600
-2.0
-1.0
0.0
1.0
2.0
Index of economic, social and cultural status
OECD Total
Belgium
Flemish


Performance in secondary schools
tracking country
non-tracking country


Performance in primary schools
Performance in secondary schools
tracking
tracking country
non-tracking country

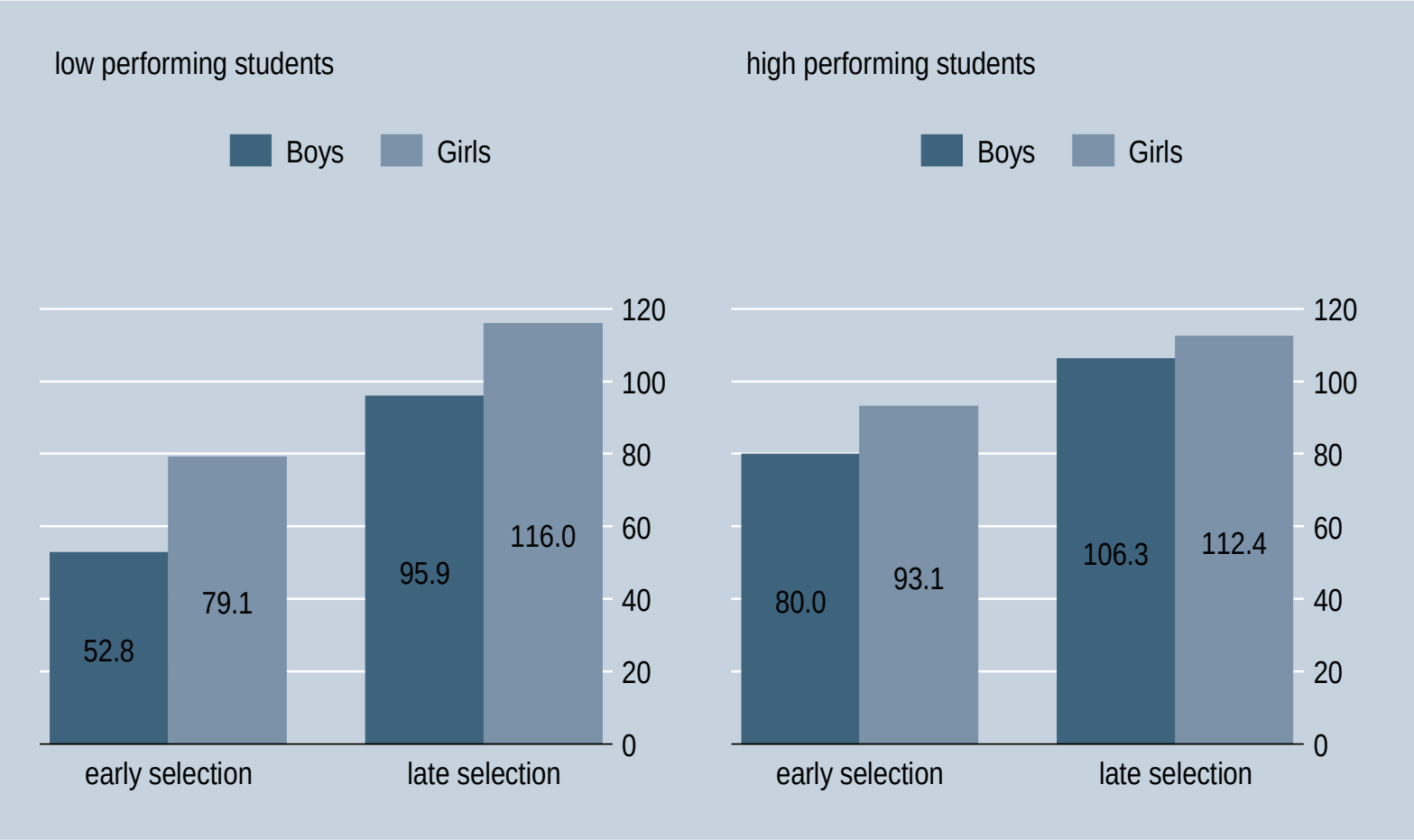
# Progress between primary and secondary education across countries




**adjusted change between PIRLS 2006 and PISA 2009**

Source: Jakubowski, Pokropek, 2015

# Progress between primary and secondary education across countries




**adjusted change between PIRLS 2006 and PISA 2009**

Source: Jakubowski, Pokropek, 2015

# Teacher black box

***„The quality of an education system depends on the quality of its teachers"***

- **Teachers have large and long-term impact on student performance** (see reviews in Hanushek and Rivkin, 2006, 2010, 2012; Chetty, Friedman, Rockoff, 2014; Jackson, Rockoff, Staiger, 2014)

- Mixed evidence on the association between achievement and teacher training, PD, teaching methods, teacher characteristics, etc.

# International evidence

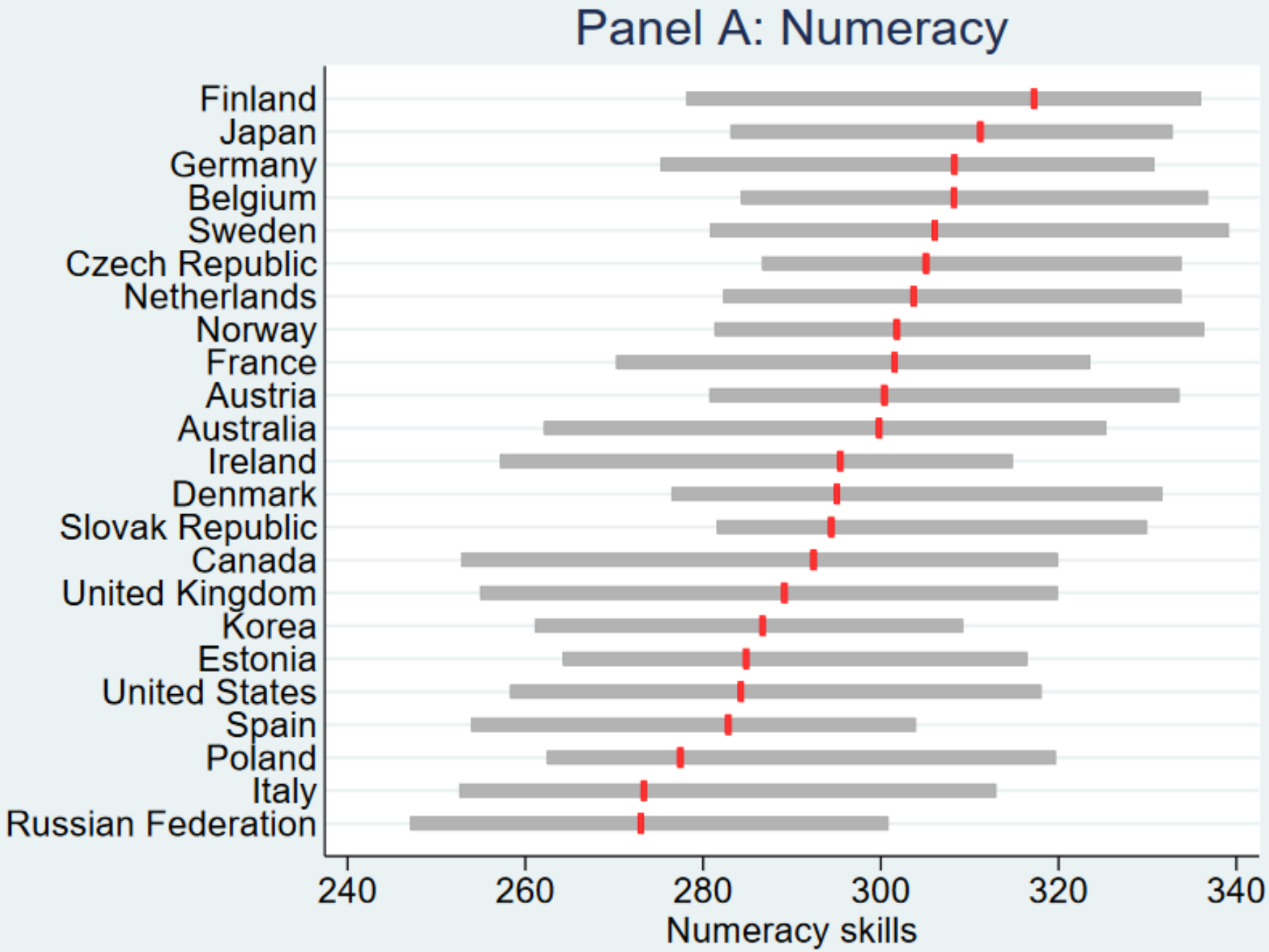
*Hanushek, Piopiunik, Wiederhold, 2018. JHR*

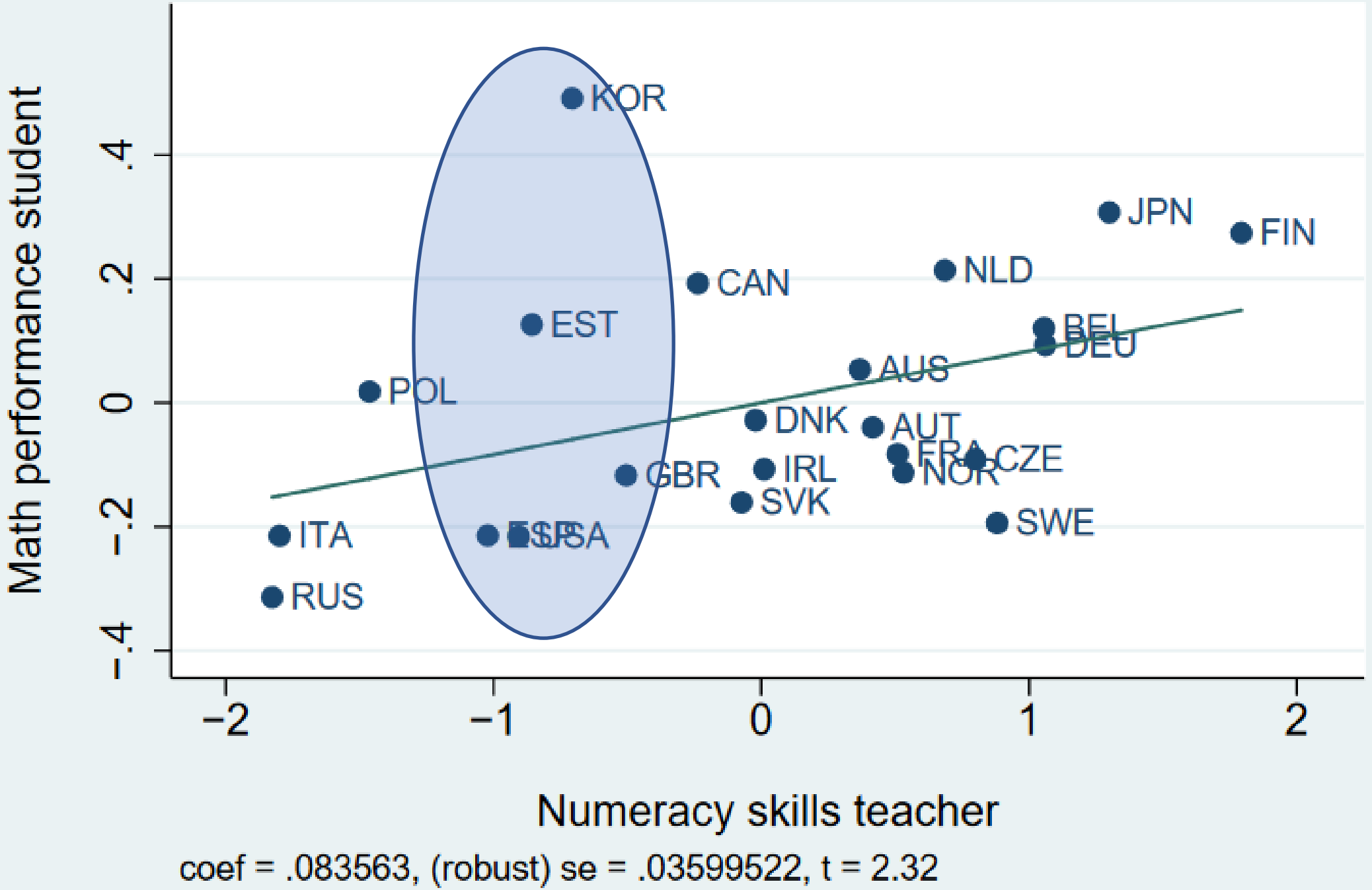
This influential study uses data from 31 countries to analyze relationship between teacher skills and student performance

> *„We find substantial differences in teacher cognitive skills across countries that are strongly related to student performance."*

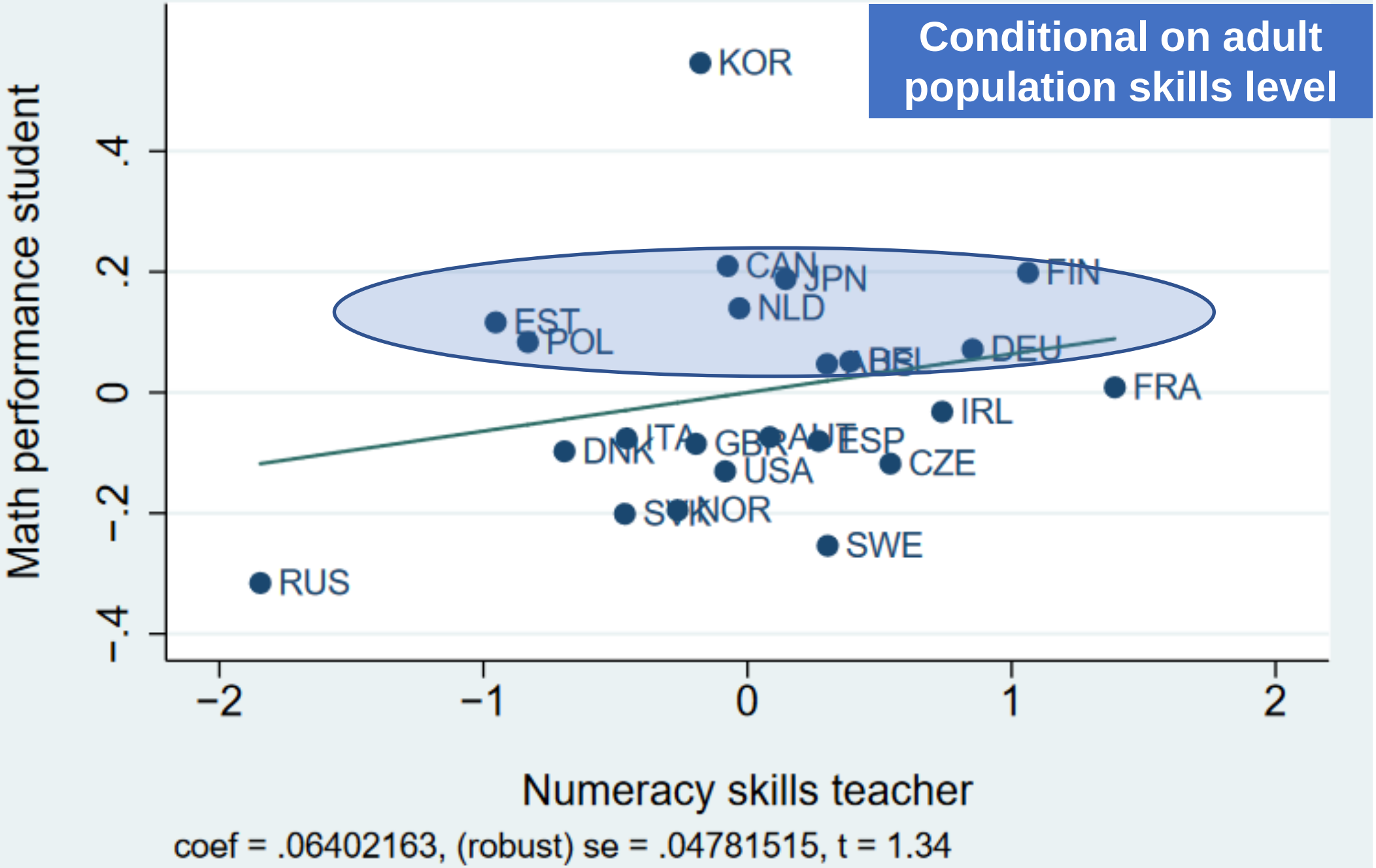
*That would demonstrate the importance of teacher education and selection but how robust are these findings?*

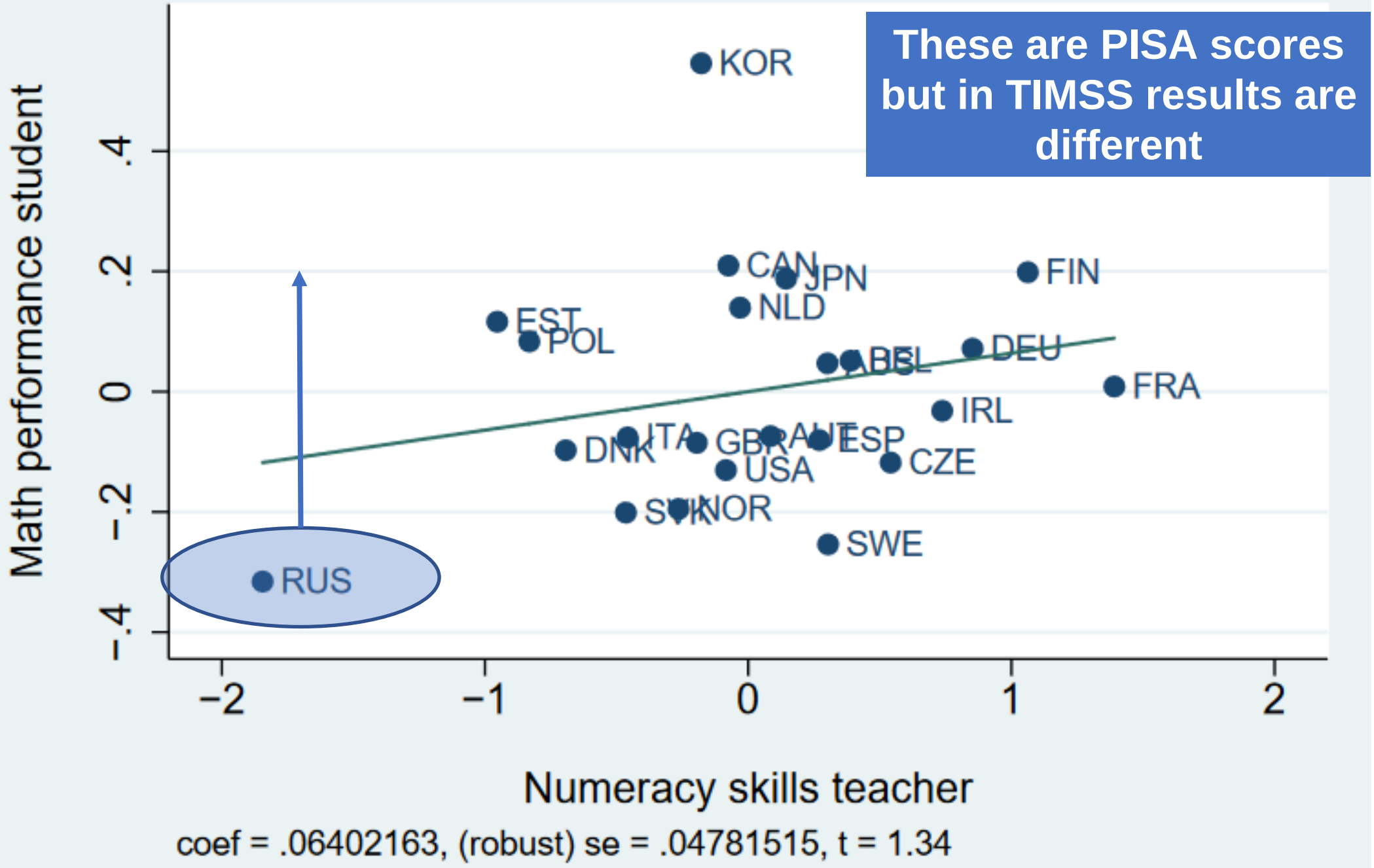
*See also: Meroni et al., 2015.*


Panel A: Numeracy
Finland
Japan
Germany
Belgium
Sweden
Czech Republic
Netherlands
Norway
France
Austria
Australia
Ireland
Denmark
Slovak Republic
Canada
United Kingdom
Korea
Estonia
United States
Spain
Poland
Italy
Russian Federation
240
260
280
300
320
340
Numeracy skills




coef = .083563, (robust) se = .03599522, t = 2.32


Conditional on adult population skills level
Math performance student
.4
.2
0
−.2
−.4
KOR
CAN
JPN
FIN
NLD
EST
POL
DEU
FRA
IRL
DNK
ITA
GBR
ESP
CZE
USA
NOR
SWE
RUS
−2
−1
0
1
2
Numeracy skills teacher
coef = .06402163, (robust) se = .04781515, t = 1.34


These are PISA scores but in TIMSS results are different
Math performance student
.4
.2
0
−.2
−.4
KOR
CAN
JPN
FIN
NLD
EST
POL
DEU
FRA
IRL
ITA
GBR
ESP
DNK
USA
CZE
NOR
SWE
RUS
−2
−1
0
1
2
Numeracy skills teacher
coef = .06402163, (robust) se = .04781515, t = 1.34

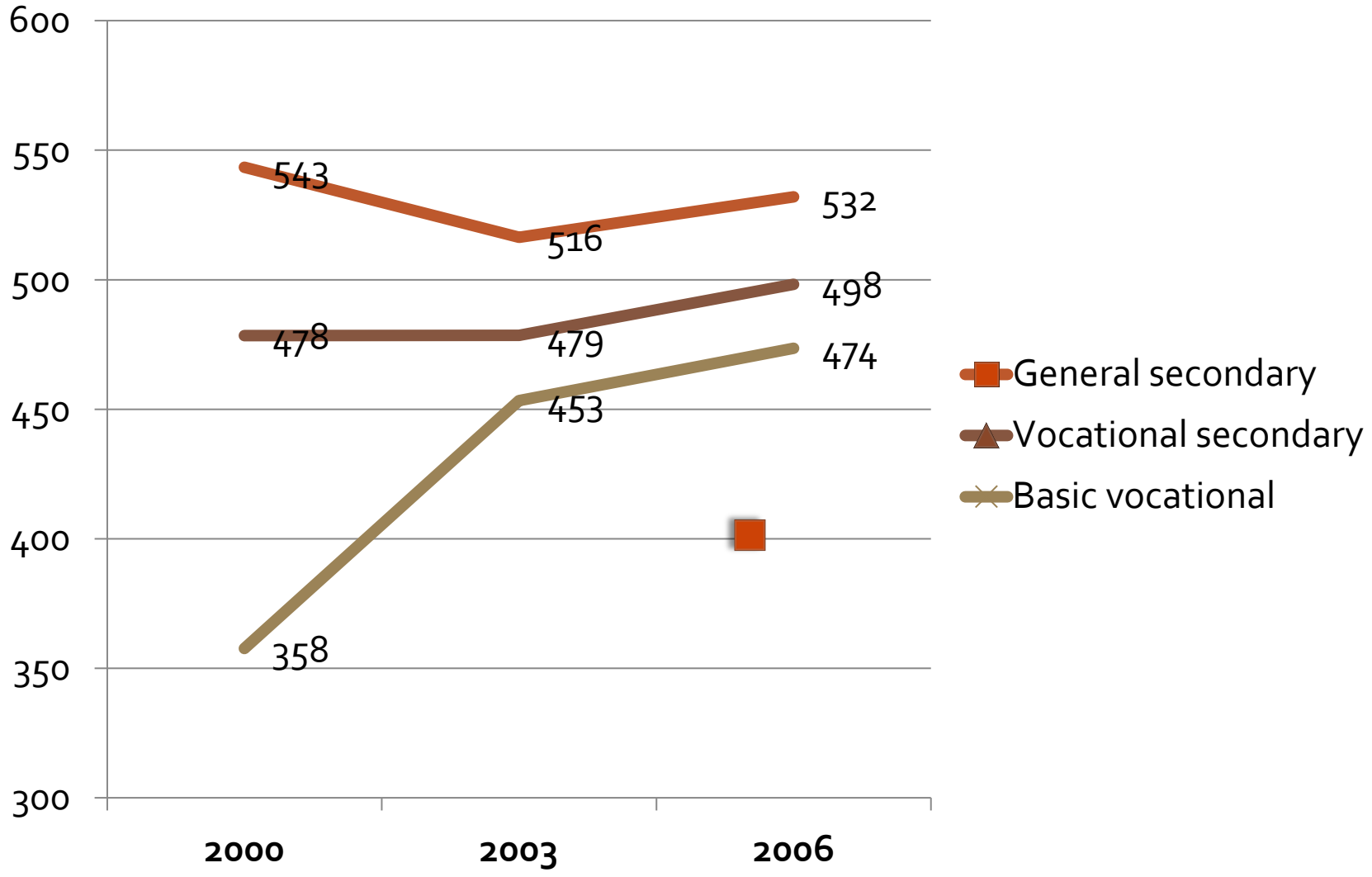
# Estimated impact on student performance

- Hanushek, Piopiunik, Wiederhold, 2018. JHR:  one standard deviation increase in teacher quality would improve student performance by 10%
- Meroni et al., 2015: variation in teacher skills explain 17% of cross-country variation, but the latter explains less than 5% of student performance variation

- **On the PISA scale it gives 10 score points increase, which is equivalent of less than 3-4 months of school education**
- **A moderate improvement in average teacher skills would lead to negligible changes in country average performance**

EVIDENCE INSTITUTE

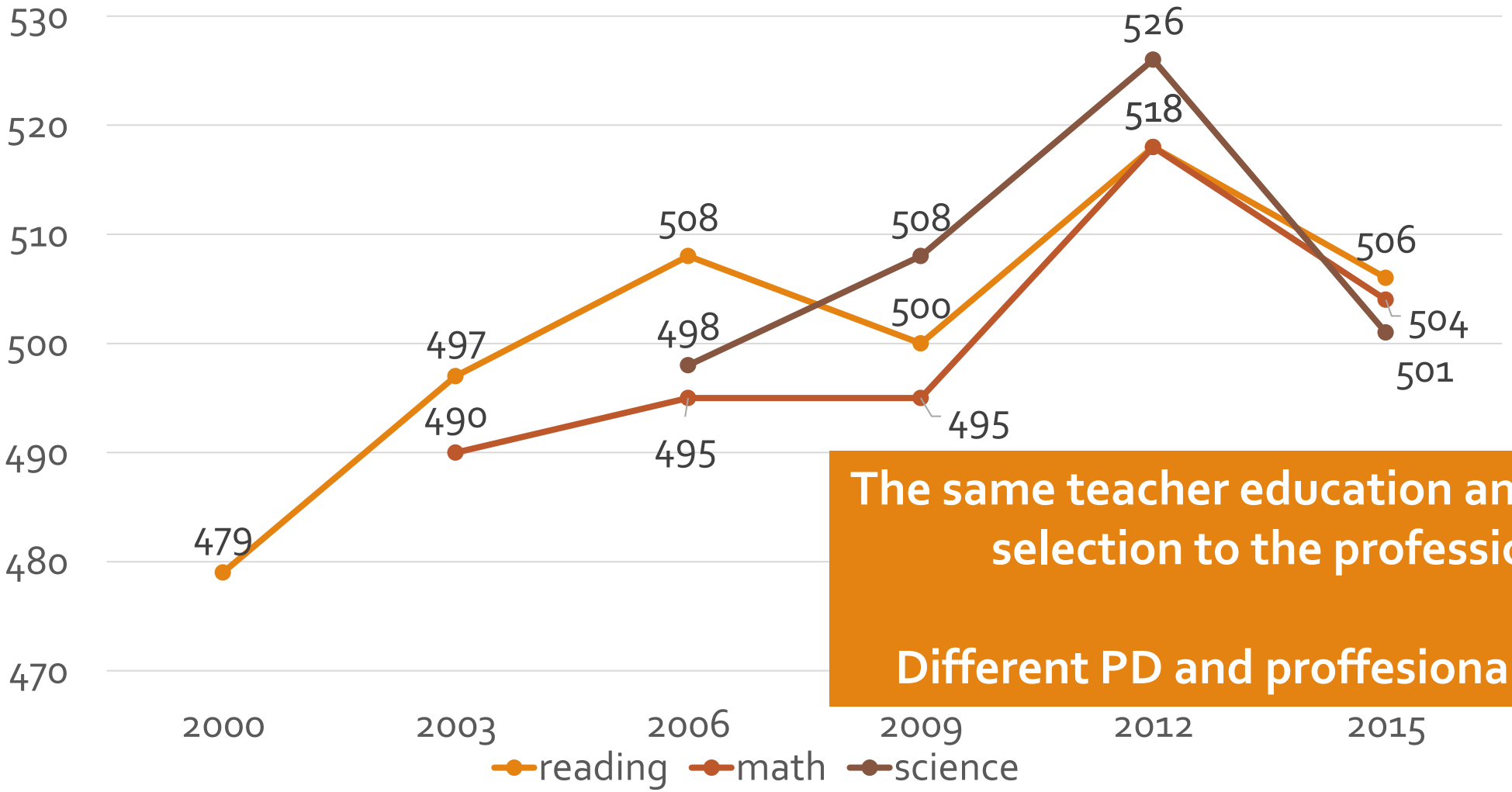
# Polish structural reforms provided learning opportunities and boosted student performance for students who would go to vocational education




Jakubowski, Patrinos, Porta, Wisniewski, 2016

EVIDENCE INSTITUTE

# Results of Polish students in PISA: increase from below OECD average to above-average level




**The same teacher education and similar selection to the profession**

**Different PD and proffesional levels**

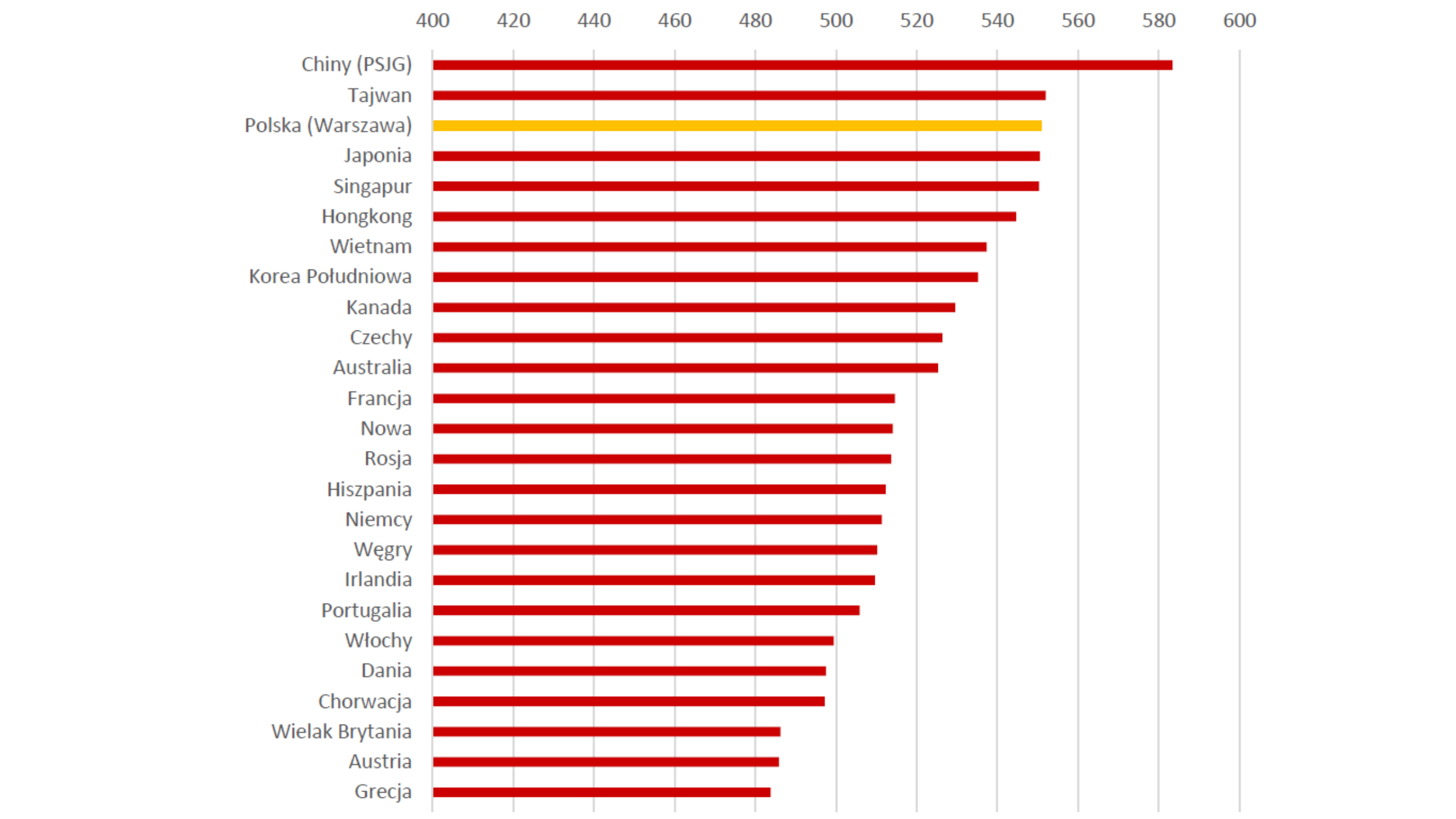
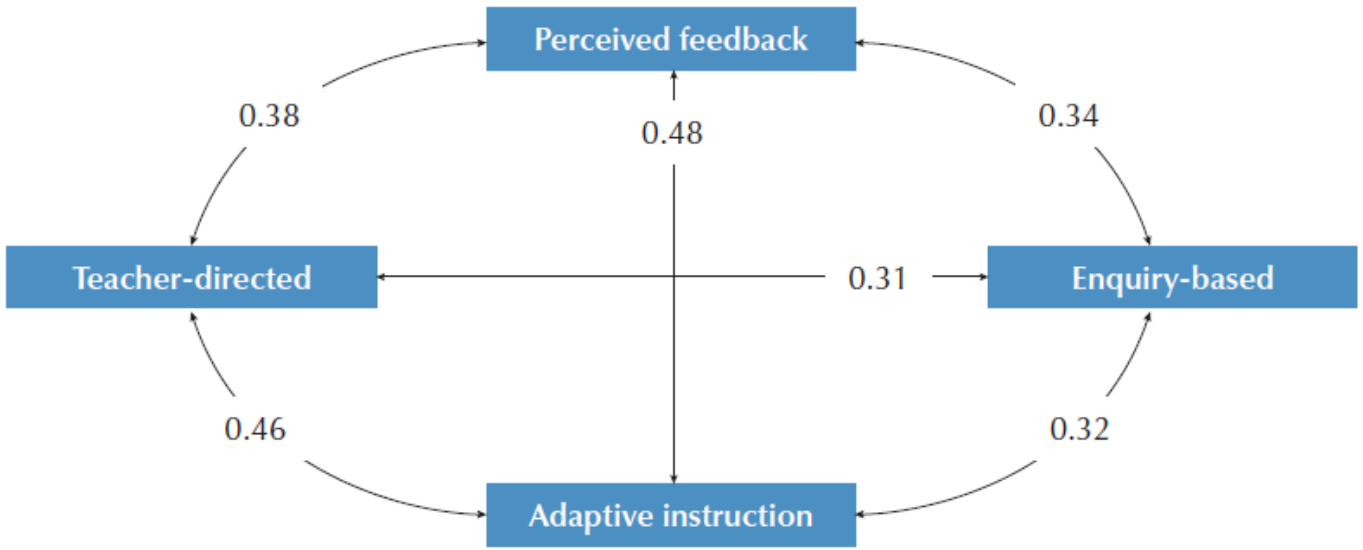

400
420
440
460
480
500
520
540
560
580
600
Chiny (PSJG)
Tajwan
Polska (Warszawa)
Japonia
Singapur
Hongkong
Wietnam
Korea Południowa
Kanada
Czechy
Australia
Francja
Nowa
Rosja
Hiszpania
Niemcy
Węgry
Irlandia
Portugalia
Włochy
Dania
Chorwacja
Wielak Brytania
Austria
Grecja

Figure II.2.12 ▪ **Relationships among instructional practices in science**

*Correlations at the student-level based on students' reports, OECD average*




**Source:** OECD, PISA 2015 Database, Table II.2.15.

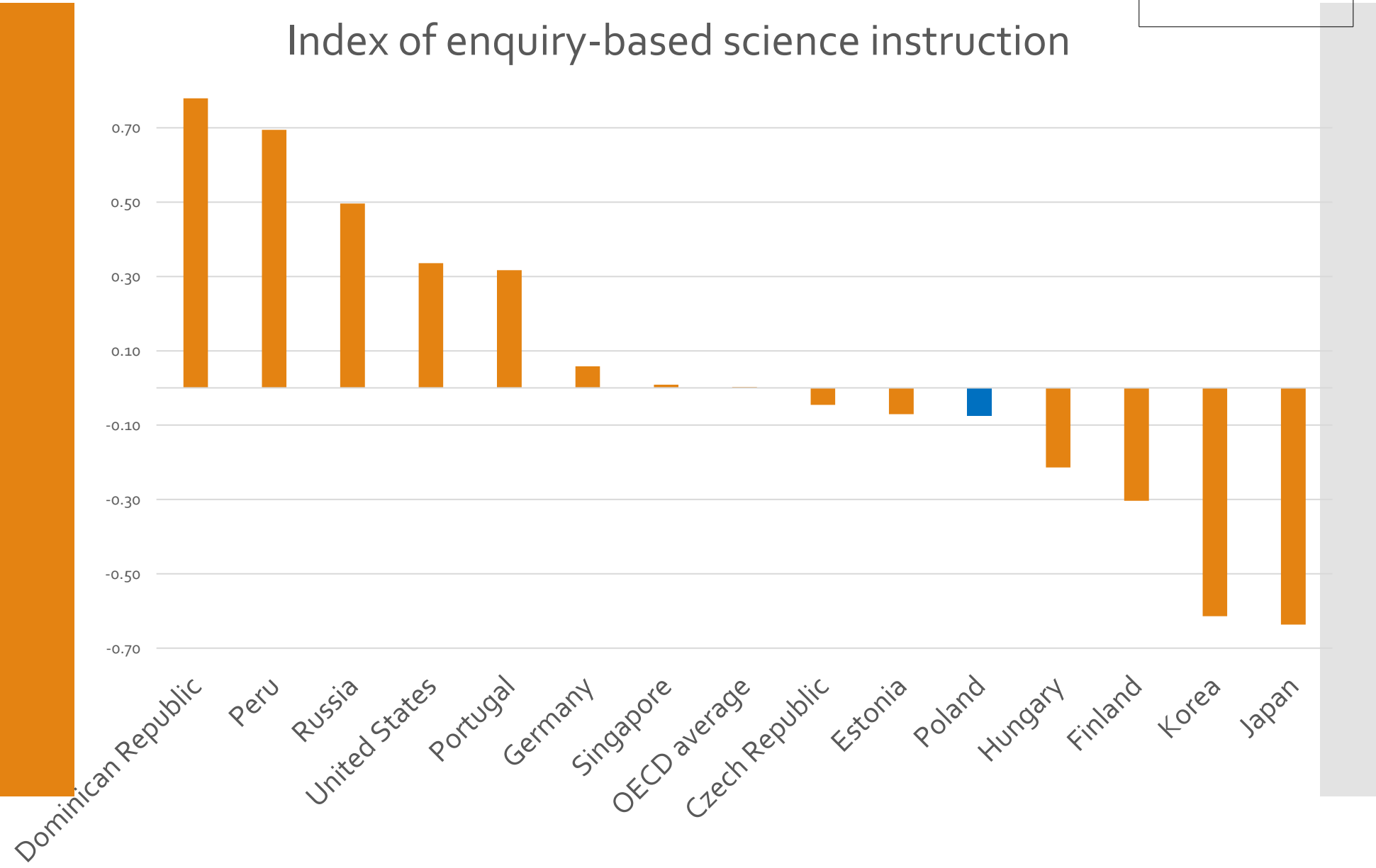
PISA 2015-school science questionnaire (inquiry-based teaching)

**ST098** **When learning <school science> topics at school, how often do the following activities occur?**

*(Please select one response in each row.)*

| | | *In all lessons* | *In most lessons* | *In some lessons* | *Never or hardly ever* |
|---|---|---|---|---|---|
| ST098Q01TA | Students are given opportunities to explain their ideas. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 |
| ST098Q02TA | Students spend time in the laboratory doing practical experiments. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 |
| ST098Q03NA | Students are required to argue about science questions. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 |
| ST098Q05TA | Students are asked to draw conclusions from an experiment they have conducted. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 |
| ST098Q06TA | The teacher explains how a <school science> idea can be applied to a number of different phenomena (e.g. the movement of objects, substances with similar properties). | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 |
| | Students are allowed to design | | | | |

# PISA 2015- school science questionnaire (inquiry-based teaching)



# PISA 2015-school science questionnaire (teacher-directed instruction)

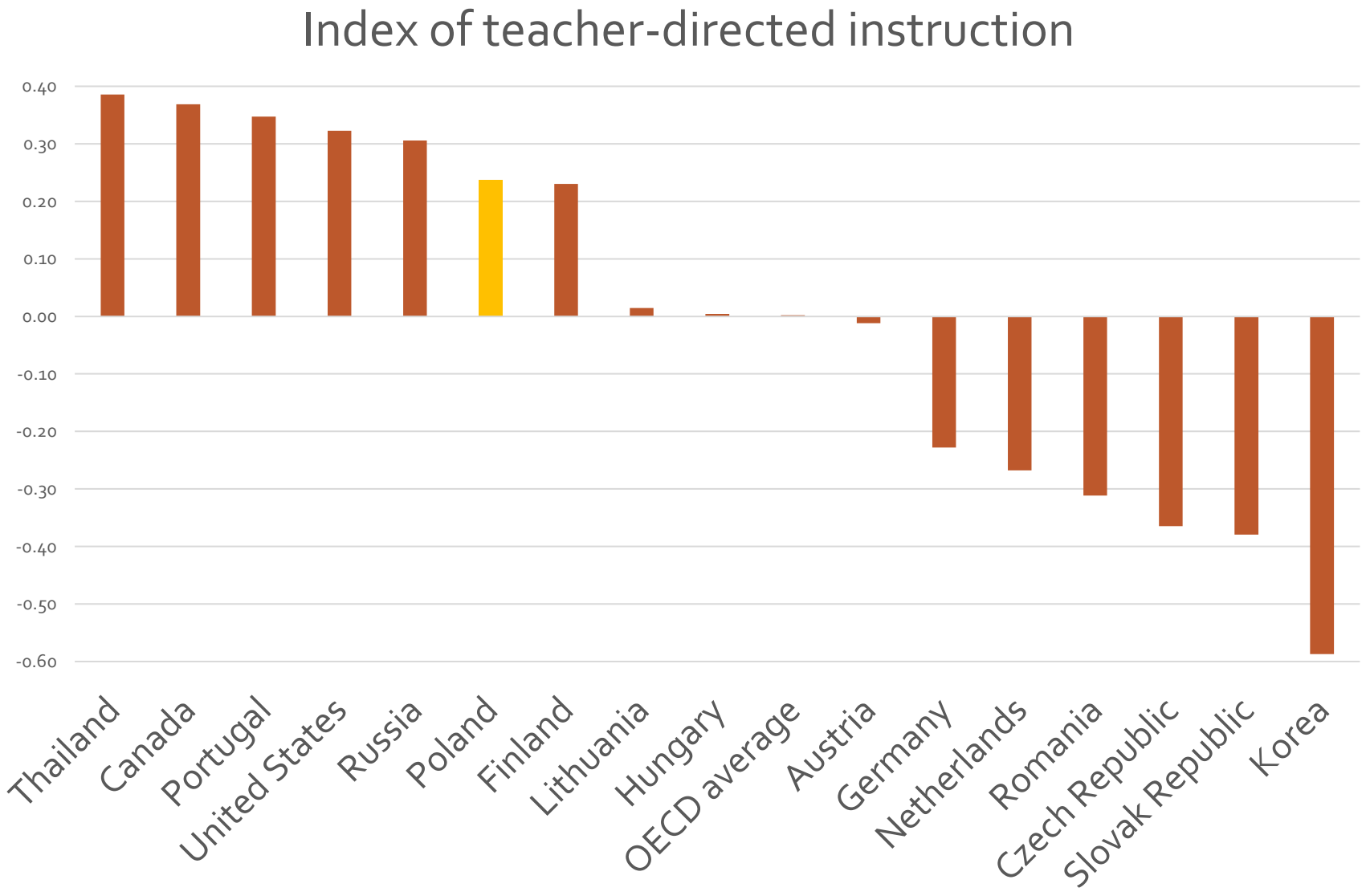
**ST103** **How often do these things happen in your lessons for this <school science> course?**

*(Remember to answer this question in reference to the <school science> course you indicated earlier.)*

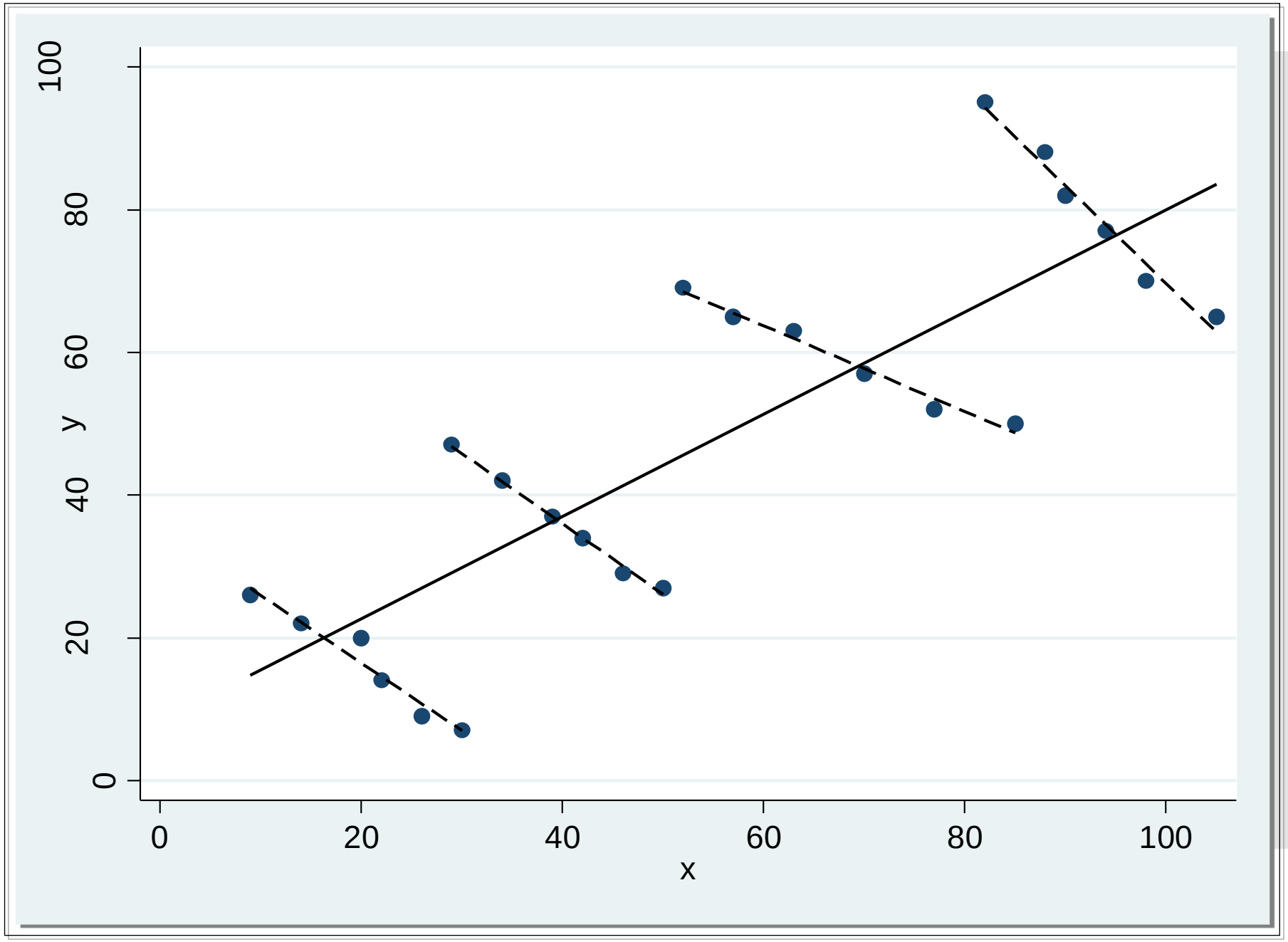
*(Please select one response in each row.)*

| | | *Never or almost never* | *Some lessons* | *Many lessons* | *Every lesson or almost every lesson* |
|---|---|---|---|---|---|
| ST103Q01NA | The teacher explains scientific ideas. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 |
| ST103Q03NA | A whole class discussion takes place with the teacher. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 |
| ST103Q08NA | The teacher discusses our questions. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 |
| ST103Q11NA | The teacher demonstrates an idea. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 |

# PISA 2015- school science questionnaire (teacher-directed instruction)



# Multilevel regression analysis



# Multilevel model results: individual effects

| | SCIENCE | EPISTEMO-LOGICAL | ENJOY-MENT | BROAD INTERESTS |
|---|---|---|---|---|
| Truancy | -8.61*** | -0.03 | -0.06* | -0.04 |
| Motivat | 15.33*** | 0.16*** | 0.22*** | 0.18*** |
| Teachsup | -4.22 | 0.05 | 0 | -0.01 |
| Disclisci | 7.80** | 0.03 | 0.06* | 0.04 |
| Instscie | -1.15 | 0.05 | 0.26*** | 0.17*** |
| x_ibteach | -1.06 | -0.11 | 0.11 | 0.06 |
| **c_ibteach** | **-10.48*** ** | **-0.07** | **0.04** | **0.05*** |
| x_tdteach | -22.04 | -0.12 | 0.27* | -0.01 |
| **c_tdteach** | **14.24*** ** | **0.14*** ** | **0.11*** ** | **0.02** |
| x_perfeed | -15.17 | -0.07 | 0.07 | 0.01 |
| **c_perfeed** | **-11.99*** ** | **-0.05** | **0.08** ** | **0.10*** ** |
| x_adinst | 14.96 | 0.13 | -0.13 | 0.11 |
| **c_adinst** | **7.92** ** | **0.04** | **0.05** | **0.02** |

# Multilevel model results: school-level effects

| | SCIENCE | EPISTEMO-LOGICAL | ENJOY-MENT | BROAD INTERESTS |
|---|---|---|---|---|
| Truancy | -8.61*** | -0.03 | -0.06* | -0.04 |
| Motivat | 15.33*** | 0.16*** | 0.22*** | 0.18*** |
| Teachsup | -4.22 | 0.05 | 0 | -0.01 |
| Disclisci | 7.80** | 0.03 | 0.06* | 0.04 |
| Instscie | -1.15 | 0.05 | 0.26*** | 0.17*** |
| **x_ibteach** | **-1.06** | **-0.11** | **0.11** | **0.06** |
| c_ibteach | -10.48*** | -0.07 | 0.04 | 0.05* |
| **x_tdteach** | **-22.04** | **-0.12** | **0.27*** | **-0.01** |
| c_tdteach | 14.24*** | 0.14*** | 0.11*** | 0.02 |
| **x_perfeed** | **-15.17** | **-0.07** | **0.07** | **0.01** |
| c_perfeed | -11.99*** | -0.05 | 0.08** | 0.10*** |
| **x_adinst** | **14.96** | **0.13** | **-0.13** | **0.11** |
| c_adinst | 7.92** | 0.04 | 0.05 | 0.02 |

# PISA 2015-school science questionnaire (teacher-directed instruction)

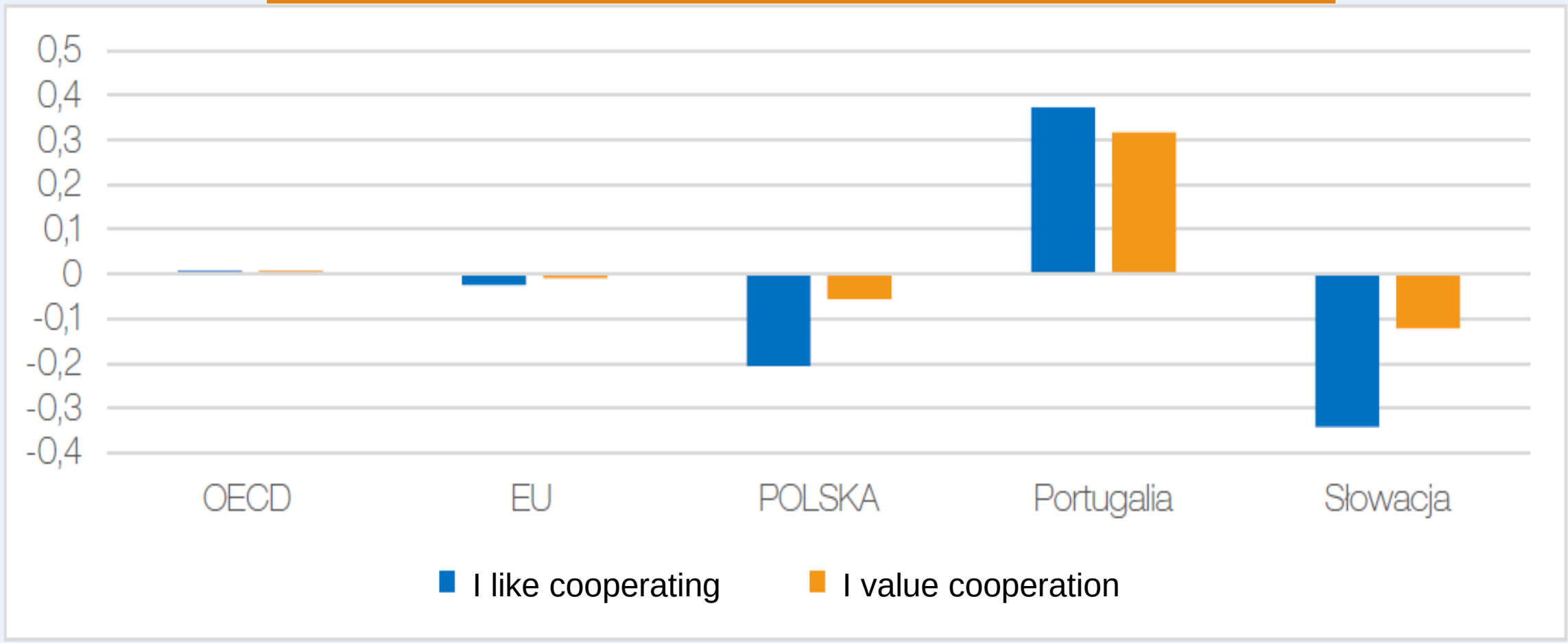
**ST103** **How often do these things happen in your lessons for this <school science> course?**

*(Remember to answer this question in reference to the <school science> course you indicated earlier.)*

*(Please select one response in each row.)*

| | | *Never or almost never* | *Some lessons* | *Many lessons* | *Every lesson or almost every lesson* |
|---|---|---|---|---|---|
| ST103Q01NA | The teacher explains scientific ideas. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 |
| ST103Q03NA | A whole class discussion takes place with the teacher. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 |
| ST103Q08NA | The teacher discusses our questions. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 |
| ST103Q11NA | The teacher demonstrates an idea. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 |

# PISA 2015: Polish students do not like to cooperate but what is their reference point?



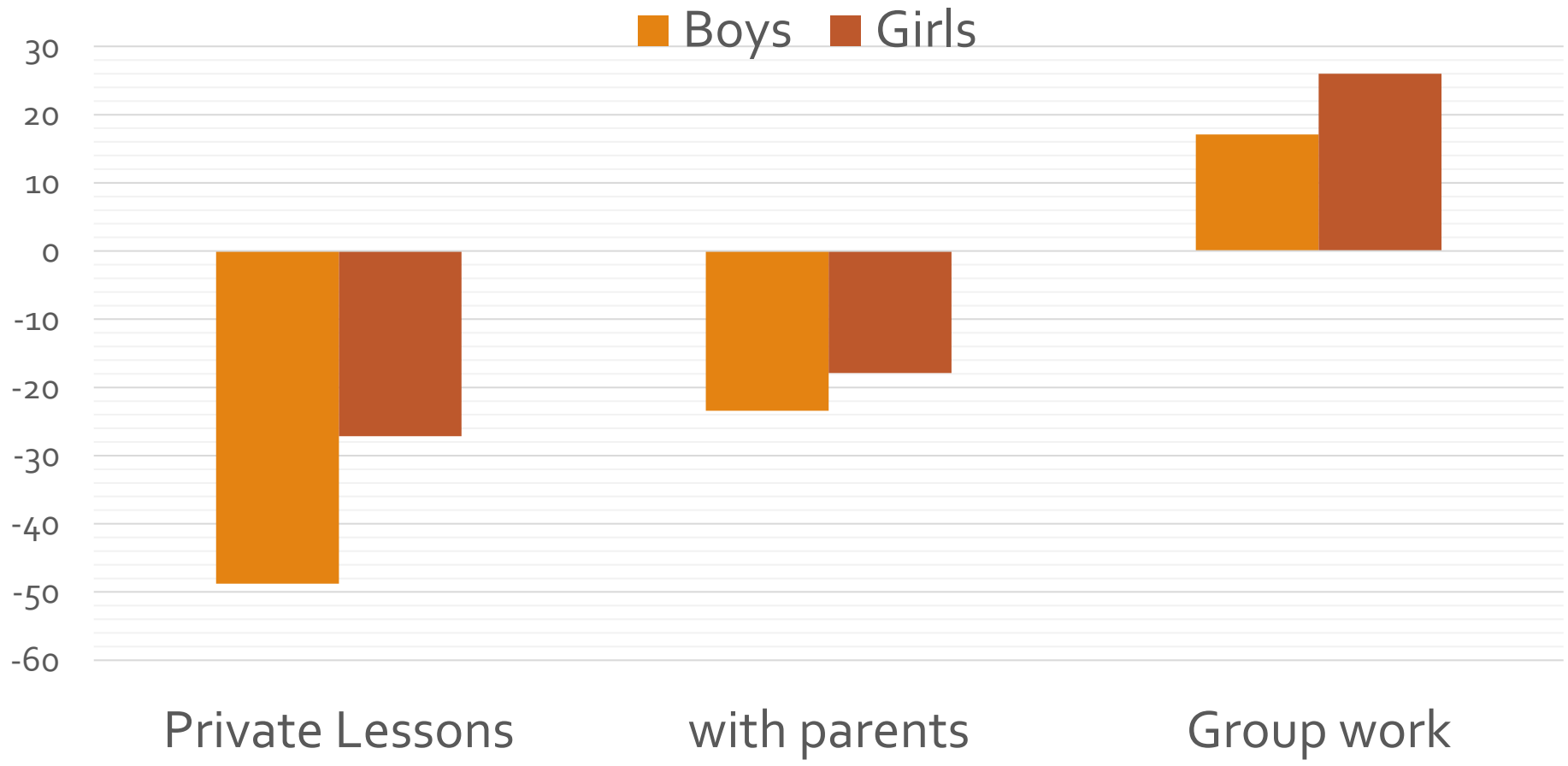

*Źródło: obliczenia własne na podstawie bazy danych PISA 2015*

## Instead of asking students what they value or what they like we asked them to say what would they prefer to do (DCE method)

**Suppose you have to do a project … which one do you prefer?**

| | Option 1 | Option 2 | Option 3 |
|---|---|---|---|
| **Subject** | Polish | Math | Geography |
| **Form of work on the task** | group work | with tutor | independently |
| **Work time** | 10 hours | 5 hours | 2 hours |
| **YOUR CHOICE** | ○ | ✖ | ○ |

# Comparing to individual work …



*European Research Studies Journal*
*Volume XXIII, Issue 4, 2020*

*pp. 1312-1333*

# Student Well-being Factors: A Multilevel Analysis of PISA 2015 International Data

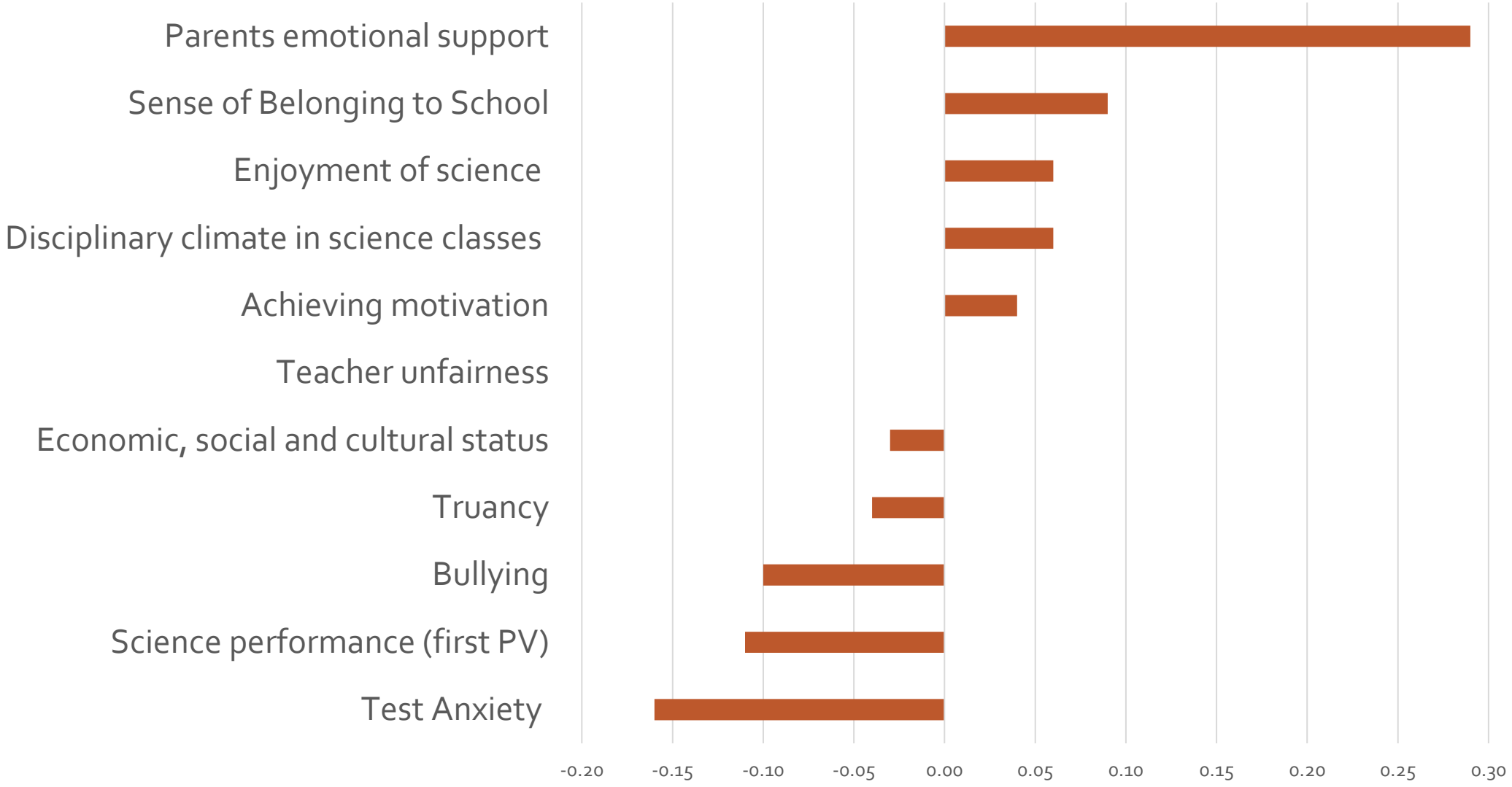
*Submitted 12/10/20, 1st revision 28/10/20, 2nd revision 15/11/20, accepted 30/11/20*

dr. Maciej Jakubowski[1], dr. Tomasz Gajderowicz[2]

## Multilevel regression explaining student life satisfaction with individual and school characteristics

| | | | |
|---|---|---|---|
| Country fixed effects | | | YES |
| School level variance | 0.074 | 0.046 | 0.013 |
| % of school variance explained | | 38% | 82% |
| Student level variance | 0.939 | 0.772 | 0.771 |
| % of student variance explained | | 18% | 18% |
| Intraclass correlation | 0.073 | 0.056 | 0.017 |
| Log pseudolikelihood | -6600.4 | -6124.3 | -6066.0 |
| N of schools | 10056 | 10056 | 10056 |
| N of students | 226916 | 226916 | 226916 |

# Standardized coefficients for school-level association with life satisfaction



# Summary

- ***Average achievement is remarkably stable***
- ***Descriptive data are crucial to understand what are the achievement and challenges for your education system***
- ***Non-cognitive measures are much less reliable and less useful***
- ***Plenty of data and a lot of interesting research but be careful…***

# Thank you

MJ@EVIDIN.PL

EVIDENCE INSTITUTE

UNIVERSITAS VARSOVIENSIS

UNIWERSYTET WARSZAWSKI
Wydział Nauk Ekonomicznych