ILSAs
The OECD Programme for International Student Assessment (PISA)
PISA in brief

Every three years since 2000, over half a million students...
- representing 15-year-olds in now over 80 countries
- National sample sizes vary between 4000 and 30,000 students

... take an internationally agreed 2-hour test...
- Focus on mathematics, science and reading
- Problem-solving, collaborative problem-solving, creative thinking, financial literacy

... and respond to questions on...
- their personal background, their schools, their well-being and their motivation

Teachers, principals, parents and system leaders provide data on:
- school policies, practices, resources and institutional factors that help explain performance differences
Some design choices and trade-offs

ILSA’s are complements, not substitutes of other research methods
Balancing investment in measurement of outcomes vs. measurement of covariates that can help to explain outcomes – Improved questionnaires a priority

Larger samples vs. higher quality of measurement – Greater geographic granularity of results to compare ‘apples with apples’ – E.g. validity vs. efficiency or relevance vs. reliability

Being open to generating new insights and hypothesis on the nature of relationships (fishing) vs. constraining design to answer specific questions (hunting) – ILSA’s are complements, not substitutes of other research methods – In particular longitudinal components place constraints on design – Balancing type I and type II errors
Design choices and trade-offs

• **Measuring change vs changing the measures**
  – Every three years one of the frameworks is revised
    • Bridging studies for content and delivery
    – New measures are first explored through innovative assessment areas

• As **comparable** as possible but as country-specific as necessary
  – Adaptive assessment instruments
  – Greater investment in better and more modular context questionnaires
  – Integration and links with national assessments

• Frameworks **informed** but not constrained by national standards and curricula
  – Curriculum validation studies, extensive consensus building
Memorisation is less useful as problems become more difficult (OECD average)

Source: Figure 4.3
Elaboration strategies are more useful as problems become more difficult (OECD average)

Source: Figure 6.2
Interpretation and reporting

Keeping things as simple as possible but as complex as necessary; and as comparable as possible but as country-specific as necessary.
Things that can be addressed (and are)

• Comparable samples
  – Sampling errors are now well within measurement errors
  – Coverage issues addressed through supplementary studies (e.g. PISA-D)

• Meaningful units of comparisons
  – Most federal countries now collect state-level PISA data

• Better instruments to facilitate and support use of PISA data

• Investment in communication
  – Quality of media coverage has significantly evolved
The perfect can be the enemy of the good (and remember that without measurement policy-makers may just throw a coin)
A world without PISA

- One question is whether ILSA’s meet the methodological requirements of some gold standard of social research...

...another question is what the state of knowledge and policy development, both nationally and internationally, would have been without them.

*Some scientific revolutions have started with controlled experiments, many have started with better measurement*
A world with PISA

• Seeing what is possible in education
  – Helping policy-makers and educators to look outwards
• Placing national standards in a broader perspective
• Exposing ‘grade inflation’
• Contextualising curricular choices
• Lowering the political cost of action
• Raising the political cost of inaction
• Generating hypotheses
Science performance and equity in PISA (2015)

Some countries combine excellence with equity

Greater socio-economic equity
Brazil: School performance and schools’ socio-economic profile

Score points vs. PISA index of economic, social and cultural status

- Public schools
- Private schools
Viet Nam: School performance and schools’ socio-economic profile

PISA index of economic, social and cultural status

Score points

Viet Nam: School performance and schools’ socio-economic profile
Brazil: School performance and schools’ socio-economic profile

Score points vs. PISA index of economic, social and cultural status

Public schools
Private schools
Comparing apples with apples and oranges with oranges

PISA math performance by decile of social background

Mathematics performance of the 10% most privileged American 15-year-olds (~Japan)
Mathematics performance of the 10% most disadvantaged American 15-year-olds (~Mexico)
Poverty is not destiny – Learning outcomes by international deciles of the PISA index of economic, social and cultural status (ESCS)

Figure I.6.7

% of students in the bottom international deciles of ESCS
Generating hypotheses

Making policy alternatives visible
Category 1.
Analyse policy variation across countries

Country-level correlation and partial correlation
Three-level regression models
Spending per student from the age of 6 to 15 and science performance

Figure II.6.2
Student-teacher ratios and class size

Figure II.6.14

High student-teacher ratios and small class sizes

OECD average

Low student-teacher ratios and large class sizes

OECD average

R² = 0.25
Learning time and science performance

Figure II.6.23

Total learning time in and outside of school vs. PISA science score for various countries. The scatter plot shows a negative correlation, with countries like Finland, Japan, and Singapore having high PISA scores despite shorter learning times, while countries like Bulgaria and Brazil have lower PISA scores despite longer learning times. The OECD average is indicated by a line, and the R² value is 0.21.
Learning time and science performance

Figure II.6.23

Score points in science per hour of total learning time

Hours

Intended learning time at school (hours)  Study time after school (hours)  Score points in science per hour of total learning time
Category 2.
Analyse policy variation within countries

Regression models, by country
School fixed-effects models
Student fixed-effects models
Differences in educational resources between advantaged and disadvantaged schools

Disadvantaged schools have more resources than advantaged schools.

Disadvantaged schools have fewer resources than advantaged schools.
Category 5.
Analyse ILSA data that follow the same students over time.

Canadian YITS study
Canadian PISA/YITS study

Improvements in reading skills between the ages of 15 and 24, by individual and family-related factors associated with skills at age 15

Growth in reading skills

- Sense of mastery
- Family educational support
- Parental cultural communication
- Socio-economic background

Low    55  High  60
Low    55  High  62
Low    60  High  49
Disadvantaged 62
Advantaged 46

Improvements

- PISA-24
- PISA-15
In conclusion

- ILSA’s can help policy-development in many ways
  - Seeing what is possible in education
    - Helping policy-makers and educators to look outwards
  - Putting national standards in a broader perspective
  - Exposing ‘grade inflation’
  - Lowering the political cost of action
  - Raising the political cost of inaction
  - Generating hypotheses

- See ILSA’s as complements, not substitutes of other research methods
  - Don’t overload ILSA’s with unrealistic expectations and avoid time and energy traps

- Don’t underplay but also don’t overplay RCTs
  - Neuroscience, big data, predictive analytics

- Keep limitations in mind when reporting results
Thank you

Find out more about our work at www.oecd.org/pisa
- All publications
- The complete micro-level database

Email: Andreas.Schleicher@OECD.org
Twitter: SchleicherOECD
Wechat: AndreasSchleicher
Multi-dimensional constructs
Comparing countries and economies on the different science knowledge subscales

Figure I.2.30

Chinese Taipei

Overall science scale, 532
Content knowledge, 538
Procedural and epistemic knowledge, 528
Comparing countries and economies on the different science knowledge subscales

Figure I.2.30
Comparing countries and economies on the different science knowledge subscales

- **Singapore**: Overall science scale, 556; Content knowledge, 553; Procedural and epistemic knowledge, 558
- **Chinese Taipei**: Overall science scale, 532; Content knowledge, 538; Procedural and epistemic knowledge, 528
- **Austria**: Overall science scale, 495; Content knowledge, 501; Procedural and epistemic knowledge, 490

Score points: 440, 460, 480, 500, 520, 540, 560

*Figure I.2.30*
Gender

The difference is not how good they are at science but in their attitudes to science
It is harder for **boys**, on average, to perform well on these types of tasks...

---

### Science competencies

- Science
- Explaining phenomena scientifically
- Evaluating and designing scientific enquiry
- Interpreting data and evidence scientifically
- Content knowledge
- Procedural and epistemic knowledge
- Physical systems
- Living systems
- Earth and space

### Knowledge types

- Science
- Procedural knowledge
- Epistemic knowledge

### Content areas

- Physical systems
- Living systems
- Earth and space
Top-performing boys' and girls' strengths and weaknesses

...but the highest-achieving **boys** perform better than the highest-achieving girls on all types of tasks, including these...
... It is harder for girls to perform well on these types of tasks, even among low achievers.

Figure I.2.29

Bottom-performing boys' and girls' strengths and weaknesses

Score-point difference (boys - girls)

Science competencies

- Science
- Explaining phenomena scientifically
- Evaluating and designing scientific enquiry
- Interpreting data and evidence scientifically

Knowledge types

- Content knowledge
- Procedural and epistemic knowledge

Content areas

- Physical systems
- Living systems
- Earth and space