



EDUCATIONAL
RESEARCH INSTITUTE
National Research Institute

INS
TIT
UTE
ibe.edu.pl

EDUCATIONAL RESEARCH INSTITUTE

NATIONAL RESEARCH INSTITUTE





EDUCATIONAL
RESEARCH INSTITUTE
National Research Institute



Did Illiteracy Double in Poland? Response Processes and Interviewer Effects in PIAAC 2023

Michał Sitek · Katarzyna Chyl-Tanaś

Educational Research Institute - National Research Institute
Poland

Nordic-Baltic PIAAC Conference 2026 · Session C · Jyväskylä, 25–27 August 2026



The puzzle of the PIAAC results in Poland

Proficiency fell sharply in Poland

Literacy: **-30.9 points** | Numeracy: **-20.8 points**

Below Level 2 in literacy: 18.8% → 39%

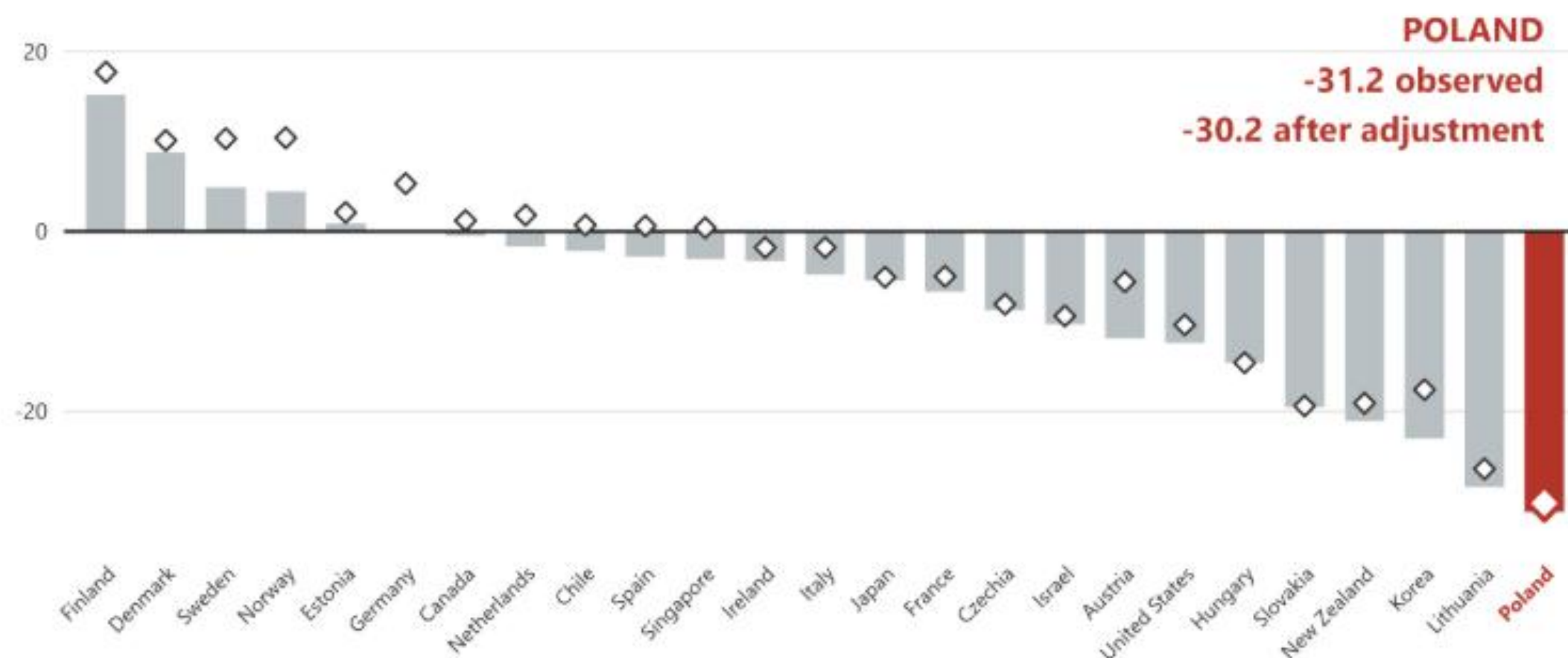
The decline was not unique to Poland.

Observed in 11 countries,
in 4 countries > 20 points.

Poland's literacy decline is not explained by demographic change

Bars: observed change in mean literacy score, Cycle 2 – Cycle 1.

Diamonds ◇: the same change after reweighting Cycle 2 to Cycle 1's age, gender and immigrant background.



OECD Table A3.1 (L). Published figures — the motivation for this talk, not a result of it.

The adjusted value is a reweighted scenario, not an observed change, and OECD's adjustment does not include education.

The puzzle of the PIAAC results in Poland

The puzzle is not explained by population change

Poland became substantially more educated

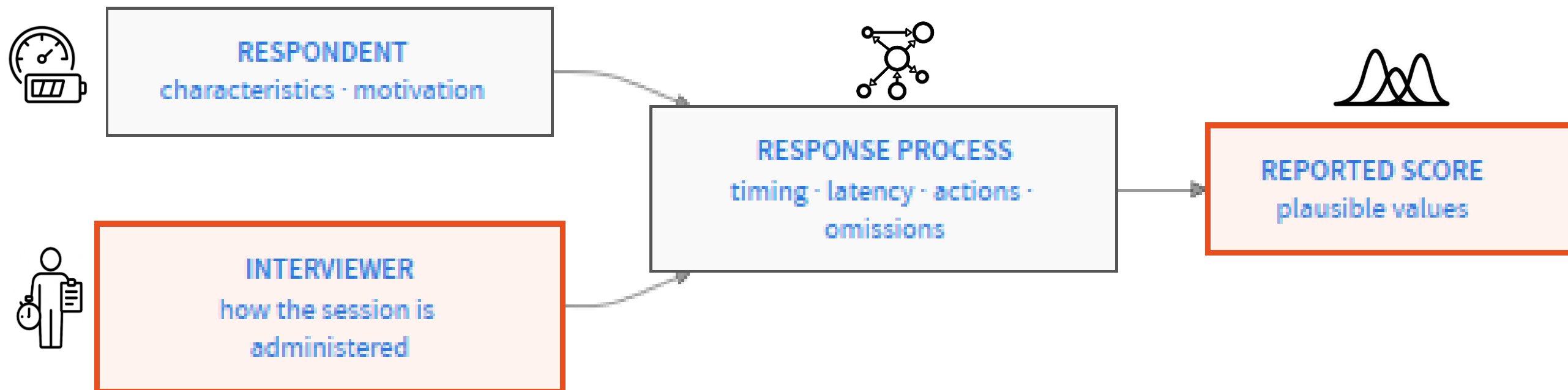
Yet proficiency declined, including among younger and more educated respondents

Could the measurement process be part of the story?

This is the question we investigate in this talk.

Measurement process is a general problem in surveys

In face-to-face surveys, research has linked interviewer presence, response styles, interview duration and interviewer observations to observed response patterns.



PIAAC and other assessments focus on performance on task and make the measurement process more important and more visible.



Identification: same locality, different interviewer

Same locality. Different interviewer.

This separates interviewer differences from locality differences.

The design creates the comparison; the model only estimates it.

5,014 full sample

114 interviewers · 1,014 localities

- ▶ *keep localities where **more than one** interviewer worked*

1,786 crossed localities

91 interviewers · 282 localities · 384 contrasts

- ▶ *and the smaller interviewer share in the locality is **at least two** respondents*

849 genuine overlap – the identification sample

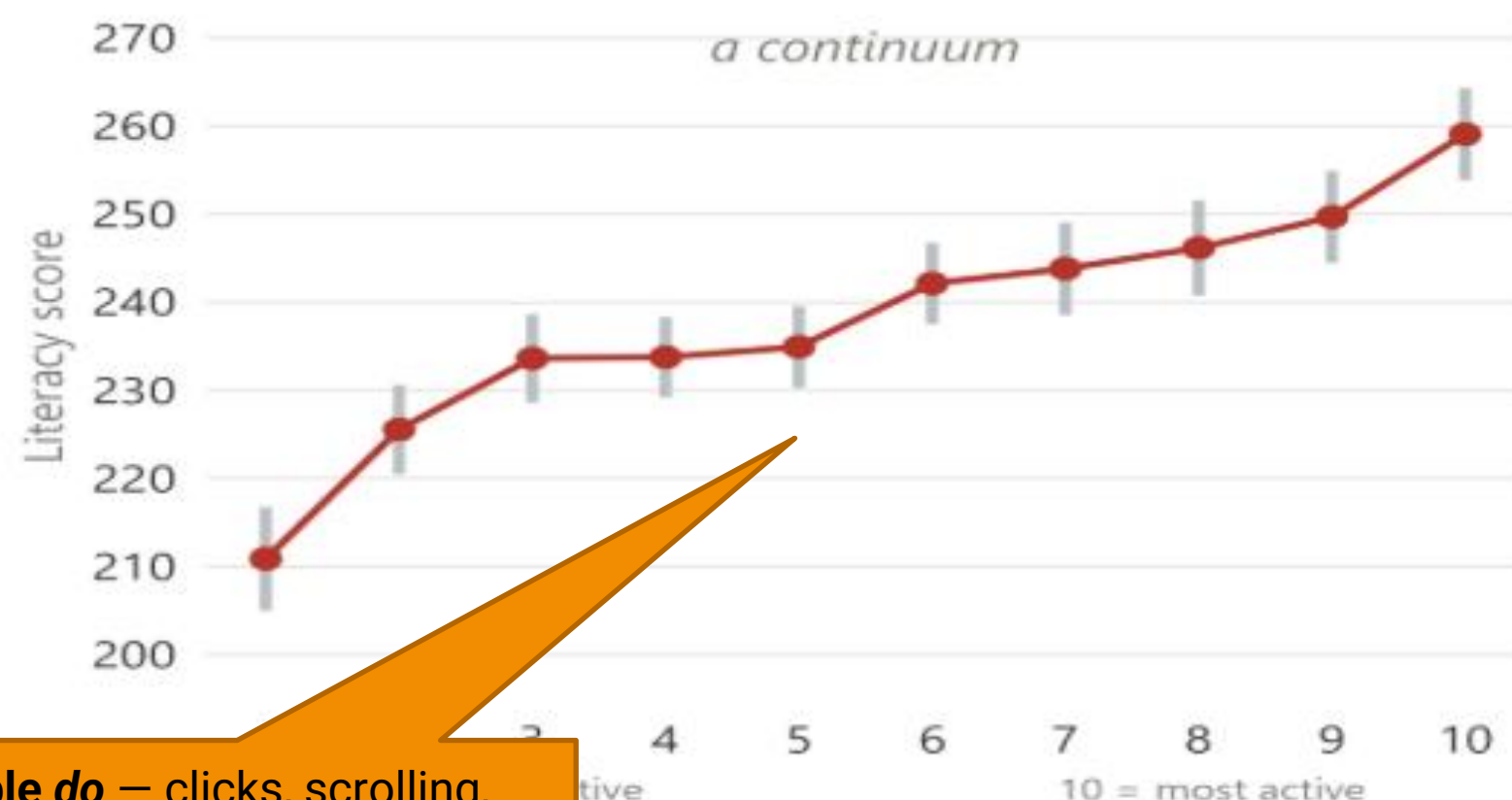
69 interviewers · 104 localities · 122 contrasts



A. Response behaviour is associated with proficiency

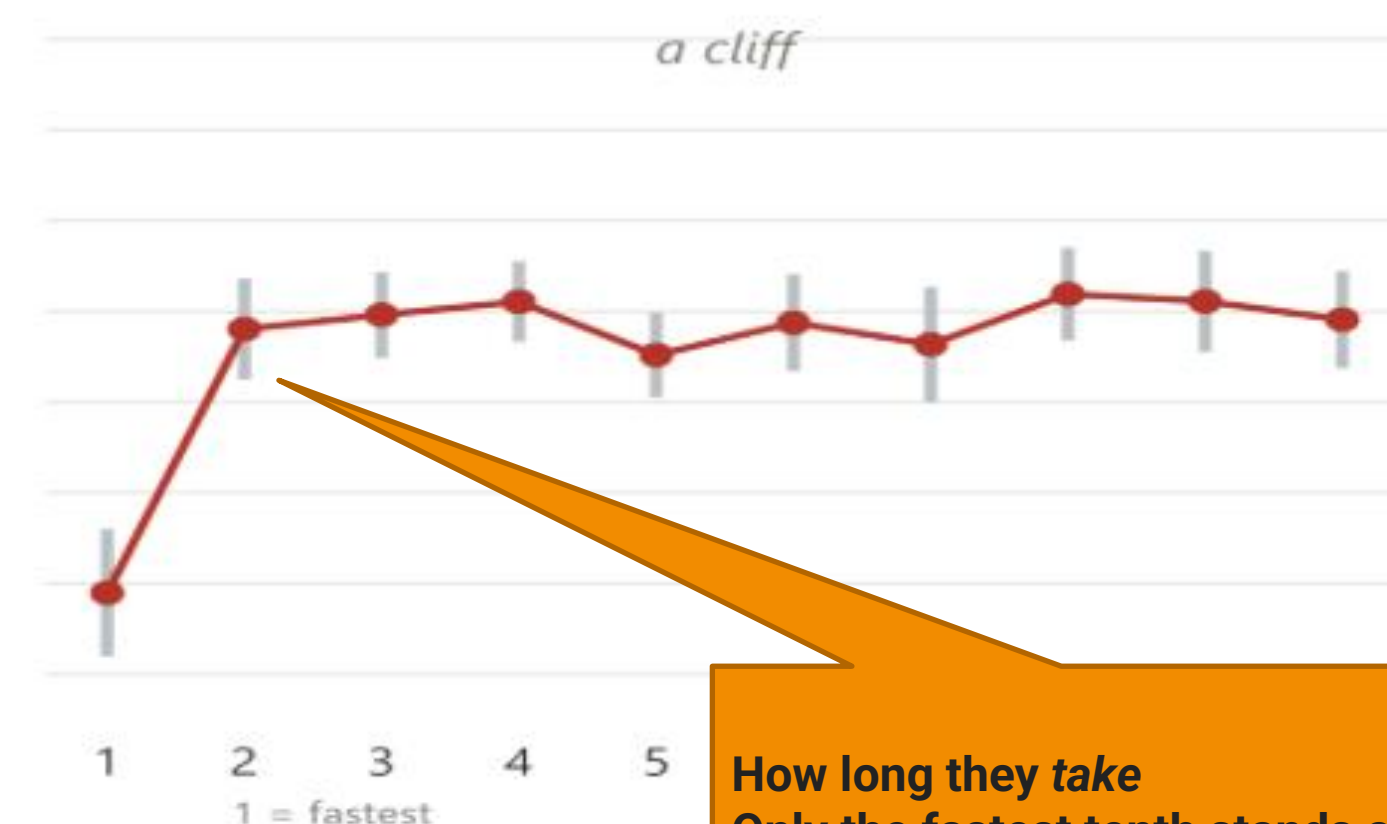
Same 16 tasks (PIAAC locator), before adaptive routing

How much they did



What people do — clicks, scrolling, exploring
11–14 score points for each 1-SD increase in engagement
About **50 score points** between the least and most active tenth

How long they took



How long they take
Only the fastest tenth stands apart,
by about 30 score points
Above that, the gradient is flat

The 16 locator tasks are scored. Engagement is measured from response times and actions; proficiency from correctness. This is an association, not an effect of engagement on proficiency.



B. Interviewers and the response process

Within-locality procedure, applied to session behaviour

Before accounting for locality:

interviewer-related association (ICCs) in response behaviour range: 32 to 55%.

Within the same locality:

26–41% of interviewer-related variance remains.

The pattern is consistent across the process measures: questionnaire and assessment duration, time and first-action latency, and number of actions.

The response process is structured by the interviewer.

C. Interviewers and proficiency

Within the same locality, interviewer-level differences account for 15–18% of proficiency variance.

This corresponds to an interviewer-level SD of 15–22 score points.

Sample	Literacy	Numeracy	APS
Full sample	30.2	31.0	25.3
Identification sample	27.9	31.1	25.1
Locality absorbed	17.2	21.6	14.6

Context: Poland's literacy SD is 51, and a proficiency level spans 50.

Observable respondent characteristics cannot reproduce this pattern: ICC = 0.01–0.03.

Careless interviewers or careless respondents?

We cannot fully separate the two.

*“...**Disengaged respondents** will always exist in surveys, and it is difficult to establish objective criteria to assess whether a reasonable level of effort was exerted. In some countries (Israel, Lithuania, New Zealand, Poland, the Slovak Republic, and Spain), it was found that **many of these respondents were clustered around a few interviewers.**” (OECD PIAAC Main Report Readers’ Companion)*

The practical intervention targeted interviewers.

Poland: **9 interviewers · 774 flagged cases · 559 cognitive cases deleted**

These were post-hoc quality adjustments after regular fieldwork controls. Some interviewers were dismissed and cases were re-assigned.

D. Does response behaviour enter the score?

Response behaviour changed sharply between the cycles in Poland:

- Very rapid responses increased
- Literacy: from 0.4% to 6.2%
- Numeracy: 0.3% → 2.8%

Omissions increased, but incorrect answers did not

- Literacy: not answered 5.6% → 14.2%
- incorrect 32.5% → 31.4%

Noise — or information about performance?

D. Scaling decisions matter

How response behaviour is treated changes the estimated decline.

Treatment	Literacy	Numeracy
Original change	-30.9	-20.8
Correction for rapid responding	-28.6	-17.1
Non-responses treated as missing	-21.0	-12.8

National report ch. 2, fig 2.3, p. 43. These are published sensitivity analyses from the national report, not our own analyses and not alternative estimates of the true decline. The rapid-response correction uses hybrid response-time modelling (Pokropek, 2016).

Same data. Different treatment of response behaviour. Different estimated change

Conclusions: Why this matters for interpreting PIAAC data

**The large decline in measured proficiency in Poland is a substantive finding.
Our analysis does not correct it.**

But interpreting a large change requires looking at the **whole measurement process**:

- **A. Respondents**
Response behaviour is related to proficiency.
- **B. Interviewers**
Interviewers structure response behaviour.
- **C. Proficiency**
Interviewer differences remain in proficiency outcomes.
- **D. Scaling**
How response behaviour is treated affects the estimated change.

What we cannot quantify:

How much each factor explains of the overall decline.

Recommendations

Data producers

Make process identifiers available.

Analysts

Use process data.

Decision-makers

Consider measurement error when interpreting large changes

Thank you!

Michał Sitek · Katarzyna Chyl-Tanaś

Educational Research Institute – National Research
Institute (IBE PIB), Warsaw

m.sitek@ibe.edu.pl

*Nordic-Baltic PIAAC Conference 2026 · Jyväskylä, 25–27
August 2026*



Identification: same locality, different interviewer

Why focus on the identification sample?

Defined on the interviewer-to-locality assignment structure, fixed before estimation, with **nothing about proficiency in either one**.

Removes *token* crossings, where an interviewer contributed a single respondent to a locality.

Limits

Only large localities need more than one interviewer. Urbanicity is constant within a locality by construction, so it **cannot confound the contrast** — what it bounds is external validity.

Why not adjust for respondents' characteristics? They affect proficiency through too many routes to treat as controls.

B. Interviewers and the response process

within-locality procedure, applied to session behaviour

Outcome	ICC before locality absorption	ICC after locality absorption	τ^2 surviving
Questionnaire duration	0.316	0.164	39%
Assessment duration	0.457	0.193	26%
Time on locator	0.552	0.268	26%
First-action latency on locator	0.487	0.340	41%
Number of actions on locator	0.343	0.172	37%

The last column is a variance-component (τ^2) ratio, not the ratio of the two ICCs beside it – those two divide to 0.52, and the difference is the changing denominator, not an error.

26–41% of interviewer variance in the response process survives locality absorption – the behaviour we just linked to proficiency is itself structured by the interviewer.

C. Interviewers and proficiency

Within the same locality, interviewers differ in proficiency outcomes: ICC = 0.15–0.18.

Step	Comparison / sample	n / intvrs / contrasts	Literacy ICC	Numeracy ICC	APS ICC
1. Full sample	Raw interviewer variation	5,014 / 114 / —	.361	.324	.352
2. Crossed localities	Localities with >1 interviewer	1,786 / 91 / 384	.308	.295	.343
3. Identification sample	+ smaller interviewer share ≥ 2	849 / 69 / 122	.291	.292	.325
4. Locality absorbed	The same 849, locality mean removed	849 / 69 / 122	.152	.184	.158

*unweighted one-way ANOVA (Lone way), single plausible value. Steps 1–3 sit inside the precision floor, so read the ladder as a **structure**, not a measured descent.*

Robustness check: Observable respondent characteristics yield ICCs of only **0.01–0.03**.