Online Resource 1

Contents

Online Resource 1

Supplementary methods, provenance and supporting results

Companion to *"When the adjustment fails: population-denominator sensitivity in coverage-adjusted PISA trends"*.

This Markdown file is the source for `ESM_1.pdf`. The complete machine-readable derived datasets, the documented Annex A7 estimator module and the provenance manifest are supplied in Online Resource 2 (`ESM_2.zip`). PISA student microdata are not redistributed.

S1. Reproducibility scope and provenance

The analysis uses sixteen externally sourced inputs: the PISA 2018 and 2022 student files; four OECD Annex or results workbooks; four OECD documentation sources retained for verification (PISA 2018 Results Volume I, the PISA 2022 Technical Report, the PISA Data Analysis Manual SPSS second edition, and the Annex A7

StatLink tables); two WPP 2024 single-age files (1 July and 1 January); two Eurostat `demo_pjan` extracts (age 15, and ages 14 to 16); OxCGRT; and the UNESCO closure-duration workbook. URLs, local paths, byte counts where a local source file is retained, and SHA-256 digests are recorded in `registry/provenance.json`.

The four documentation sources are checked programmatically rather than quoted from memory. Each cited passage was located in the source PDF by regular expression, and the matched page number and surrounding text were written to `results/oecd_primary_source_verification.json`, which is supplied in Online Resource 2 so that every quoted passage can be traced to its page. Every element of the Annex A7 procedure reported in the manuscript was located on one page of PISA 2018 Results Volume I: the single added case, the final-weight rule, the replicate-weight rule, the replacement of negative weights by zero, the below-range performance value, the lower-tail assumption, the reference to Table I.A2.2, the computation of the median and upper percentiles on the augmented sample, and the suppression of the adjusted median where coverage falls below 50%.

Input	Source	Status
PISA 2018 student file	OECD PISA 2018 archive	Not redistributed
PISA 2022 student file	OECD PISA 2022 archive	Not redistributed
PISA 2022 Annex A2 workbook	OECD StatLink `hpg9nd`	Recorded
PISA 2022 link-error workbook	OECD StatLink `48f0zo`	Recorded
PISA 2022 mean workbook	OECD StatLink `xluqor`	Recorded
PISA 2022 change workbook	OECD StatLink `muorhc`	Recorded
PISA 2018 Results Volume I (Annex A7 source text)	OECD Publishing	Recorded
PISA 2022 Technical Report	OECD Publishing	Recorded
PISA Data Analysis Manual, SPSS 2e	OECD Publishing	Recorded
PISA 2018 Volume I Annex A7 StatLink tables	OECD StatLink `888934028957`	Recorded
WPP 2024 single-age population, 1 July	UN DESA Population Division	Recorded
WPP 2024 single-age population, 1 January	UN DESA Population Division	Recorded
Eurostat age-15 population	Eurostat dissemination API	Extract retained
Eurostat age 14–16 population	Eurostat dissemination API	Extract retained
OxCGRT national file	Oxford policy tracker	Recorded
School-closure workbook	UNESCO Institute for Statistics	Recorded

The four OECD documentation sources are recorded with digests but are not redistributed, because they are copyrighted OECD publications available from OECD Publishing at the DOIs recorded in the manifest. Their digests are `d943dd7fc65b2a78…` (PISA 2018 Results Volume I), `6f6ccf26beb01372…` (PISA 2022 Technical Report), `c6b6ddeebd99797f…` (PISA Data Analysis Manual, SPSS second edition) and `122fabc8c19fdeb2…` (Annex A7 StatLink tables); full digests are in `registry/provenance.json`.

The two PISA archive SHA-256 digests are `df39c2b4c27649fd86866c0f7d23f9544ffc071634387d4ee52d1b433de70ea6` (2018 container) and `24ee020cd6315d8aca52e07fd9aaef21cbf4f9e5b21601734b4ee6d0928a0905` (2022 container). The WPP 1 July file digest is `f84c75789ccbd385122ad19bac70026d6306878442afbaa24e8e50a23b4bee2f` and the WPP 1 January file digest is `b9a2c95d1608e7ae2c506c1516f1dab60535513c143c49b3e55087515b7ebe91`. The Eurostat age 14–16 extract digest is `6e3c976e4f816a192b752d1f7ef26a9a6645a97abb34ef990c93651b84eba8ea`.

S2. Computing environment

The recorded environment is macOS 26.5.2 on Apple silicon (arm64), Python 3.14.3, eight logical CPUs and 16 GB RAM. The direct third-party dependencies of the analysis, with the exact versions used, are recorded here:

Package	Version	Use
numpy	2.4.4	array arithmetic and weighted estimators
scipy	1.17.1	descriptive correlations and regressions

Package	Version	Use
pandas	3.0.2	tabular assembly and joins
pyreadstat	1.3.6	selected-column SPSS input
openpyxl	3.1.5	OECD workbook input
matplotlib	3.11.1	figures

No network access is required after source acquisition. The core estimator, including the documented Annex A7 final-weight and replicate-weight rules and the weighted-percentile convention identified in S3, is implemented in `src/pcov/estimators.py`, which is supplied in Online Resource 2. The percentile-convention calibration, the denominator comparability audit and the recomputations reported in S3.3 and S7b all call that module; their outputs are the machine-readable tables in `results/` described in S13.

S3. Estimation detail

S3.1 Plausible values and Fay-BRR

The analysis uses the final student weight `W_FSTUWT`, 80 replicate weights `W_FSTURWT1`–`W_FSTURWT80`, and 10 plausible values per domain. With Fay's factor 0.5, the replicate multiplier is 0.05:

`V_samp = mean_u[0.05 × sum_r(theta_ur − theta_u)^2]`.

The plausible-value component is

`V_imp = (1 + 1/10) × sum_u(theta_u − theta_bar)^2 / 9`,

and `V_total = V_samp + V_imp`. Each economy-cycle-domain estimate requires 10 final-weight calculations plus 10 × 80 replicate calculations: 810 weighted computations.

S3.2 Documented Annex A7 estimator

For population denominator `P`, final-weight sum `W`, target percentile `p=1−s`, and top share `s`, the documented synthetic weight is

`S = max(P − W, 0)`, `T = S + W`, and `q = (pT − S)/W`.

The calculation is repeated independently for every replicate:

`S_r = max(P − W_r, 0)`, `T_r = S_r + W_r`, and `q_r = (pT_r − S_r)/W_r`.

Before the zero-replacement boundary, `q = 1 − sP/W = 1 − s/CI3` for `s < CI3 <= 1`. If `CI3 > 1`, the documented rule sets the negative synthetic weight to zero and gives `q = 1 − s`. Non-finite probabilities or probabilities outside `(0,1)` are labelled infeasible; probabilities are not clipped.

The same estimator is imported by production script 09 and implementation-check script 15. At stored output precision, the two generated outputs agree exactly for 8,163 jointly non-missing feasible scalar cells: PISA- and WPP-scenario estimates, standard errors and tested-sample quantile levels for `p` equal to 0.75, 0.90 and 0.50. They also agree for 15 retained out-of-range raw required-level diagnostics at `p = 0.50` (9 PISA and 6 WPP), giving 8,178 stored scalar comparisons in total; the corresponding estimates and standard errors remain missing, as required by the feasibility rule. The unsupported cells are Baku (Azerbaijan) 2018 under PISA POP15, Guatemala 2022 under both denominators and Cambodia 2022 under both denominators, each in all three domains.

S3.3 Weighted-percentile convention, identified against OECD output

OECD does not publish the weighted-percentile algorithm behind its reported percentiles. Three primary documents were checked directly for a stated rule, with the page-level search record in `results/oecd_primary_source_verification.json`:

1. *PISA 2018 Results Volume I, Annex A7.* Says only that "the median and upper percentiles of the distribution are computed on the augmented sample". No algorithm is given.
2. *PISA 2022 Technical Report.* Contains no weighted-percentile definition anywhere in the report.
3. *PISA Data Analysis Manual (SPSS, second edition), Box 17.3.* Supplies the macro `MCR_SE_PERCENTILES_PV`. It weights the file by the final weight, calls the package frequency procedure for the requested percentiles once per plausible value, repeats that call once per replicate weight, and combines the results as `SE = sqrt(V_samp + (1 + 1/M) B_m)`. The combination rule is fixed; the percentile itself is delegated to the statistical package and no interpolation rule is stated.

None of the three documents contains the strings "weighted percentile", "weighted quantile", "cumulative weight/frequency" or "interpolation" in a statistical sense. The weighted-percentile rule behind the Annex A7 statistic is therefore not established by the documentation examined here, and the convention was identified empirically against OECD's own published percentile values instead of being copied from a stated rule. This is a claim about what these three documents specify; it is not a claim about the full set of code that OECD releases, which was not audited.

The documented macro does, however, independently confirm two conventions used here. First, the combination rule: the percentile is obtained once per plausible value and the final statistic is their mean (`compute STAT=mean(!k)`), which is the rule the calibration selected empirically over pooling. Second, the variance structure of Equations 1 and 2: the macro's `CONS` default of 0.05 is the Fay BRR multiplier, and its `1.2` and `0.25` factors are `1 + 1/M` and `1/(M - 1)` for the five plausible values of that era, corresponding to 1.1 and 1/9 for the ten used here.

Criterion: PISA 2022 Results Volume I, Figure I.2.4 (source Table I.B1.2.1) publishes the mathematics score at the 10th, 50th and 90th percentile for each entity. All three percentiles were recomputed from the delivered 2022 microdata for all 80 entities under five weighted-percentile definitions and two plausible-value combination rules, giving 240 comparisons per candidate. Writing sorted scores `x_(k)` with weights `w_(k)`, cumulative weights `C_k` and total weight `W`:

Definition	Rule at probability `q`	Combination	Mean abs. dev.	Max abs. dev.	n
inverse CDF	`x_(k*)`, `k* = min{k : C_k >= qW}`	per plausible value	0.0004	0.0434	
midpoint (Hazen)	interpolate at knots `(C_k - w_k/2)/W`	per plausible value	0.0070	0.0505	
right-edge CDF	interpolate at knots `C_k/W`	per plausible value	0.0286	0.0958	
left-edge CDF	interpolate at knots `(C_k - w_k)/W`	per plausible value	0.0297	0.1195	
inverse CDF	as above	pooled plausible values	0.0628	0.3417	

Score points. The left- and right-continuous variants of the inverse CDF are numerically indistinguishable in these data because exact ties at `C_k = qW` do not occur with continuous weights.

The inverse-CDF definition computed per plausible value and then averaged agrees with the published values to the published four-decimal precision for 77 of 80 entities on all three percentiles. The three exceptions are Iceland (0.0378 at the 50th and 0.0434 at the 90th percentile, n = 3 360), Georgia (0.0088 at the 90th) and Australia (0.0009 at the 10th). No entity matches to four decimals under the midpoint interpolating rule. This definition is used for every percentile reported in the manuscript. It is an identification of OECD's convention from OECD's output; it is not a claim to have recovered OECD's code.

Residual implementation sensitivity is reported rather than asserted to be small. Recomputing the whole analysis under the midpoint interpolating rule changes:

Quantity	Mean absolute change	Maximum absolute change
Adjusted top-quarter level, PISA denominator	0.0044	0.0256
Adjusted top-quarter level, WPP denominator	0.0045	0.0358
Standard error of an adjusted level	0.0026	0.0310
Economy-level PISA–WPP trend contrast, mathematics	0.0096	0.0270

Quantity	Mean absolute change	Maximum absolute change
Economy-level PISA–WPP trend contrast, reading	0.0113	0.0323
Economy-level PISA–WPP trend contrast, science	0.0098	0.0322

Score points; correlation between the two sets of economy-level contrasts is 1.000 in all three domains. Domain mean absolute contrasts move from 3.4342, 3.4526 and 3.4290 to 3.4347, 3.4507 and 3.4297. Counts above link error (38, 52, 51), counts above twice link error (17, 33, 30), sign disagreements (4, 5, 4) and significant sign disagreements (0, 0, 1) are identical under both conventions. The calibration record is in `results/quantile_definition_calibration.csv` with its deviations and verdict in `results/quantile_definition_deviations.csv` and `results/quantile_definition_verdict.json`; the recomputation under both conventions is in `results/quantile_convention_levels.csv` and `results/quantile_convention_contrast.csv`.

S3.4 Analytic-sample flow

The 82 economies with 2022 Annex A2 coverage data reduce to 73 with coverage records in both 2018 and 2022 because nine have no 2018 row in the extracted Annex A2 panel. Baku (Azerbaijan) is then absent from the 72-country PISA--WPP growth comparison because it is a subnational city entity without a comparable national WPP join. The complete two-cycle trend sample contains 69 economies after explicit harmonisation of four cross-cycle student-file labels: Hong Kong/Hong Kong (China), Macao/Macao (China), Republic of Moldova/Moldova and Turkey/Türkiye. Four of the 73 two-cycle coverage cases remain outside the complete trend sample:

Entity	Data-supported exclusion reason
Baku (Azerbaijan)	Subnational entity; no comparable national WPP join
Cyprus	No record in either supplied PISA student file
Ukraine	2018 national sample and 2022 partial-region sample are not comparable
Viet Nam	All required 2018 plausible values are missing in the supplied student file

The 38-country Eurostat sample is a separate branch requiring non-missing 2018 and 2022 age-15 values. The 159 Annex A7 economy--cycle cells are another branch from the student files (79 usable 2018 cells and 80 usable 2022 cells), not a reduction from the Eurostat sample.

S4. Validation record

Check	Reproduced	Published/source	Result
Philippines 2022 mathematics mean	354.7197	354.72	Pass
Philippines 2022 reading mean	346.5474	346.55	Pass
Philippines 2022 science mean	356.1673	356.17	Pass
Philippines 2022 sum of final weights	1,782,895.5818	1,782,895.59	Pass
Philippines 2018 sum of final weights	1,400,584.0599	1,400,584.05	Pass
CI3 arithmetic identity	422/424 rows	422/424 rows	Pass (2 exceptions)
Annex A2.1 versus A2.2 CI3, 2022	82/82 equal	82/82 equal	Pass
Philippines OxCGRT days, level 3 / level >=2	621 / 716	621 / 716	Pass
Philippines UNESCO full / partial closure days	556 / 6	556 / 6	Pass

The two CI3 exceptions are retained rather than forced to equality. Jordan 2022 is published as 0.940218 while the tabulated weighted-participant-to-POP15 ratio is 0.966865. The Ukrainian-regions row is published as 0.639000 while the reconstructed ratio is 0.639415.

S5. Coverage change and exact decomposition

For 73 economies with both cycles, the published 2018–2022 CI3 change has mean +1.82 percentage points, SD 8.64 and range −45.19 to +27.05. Twenty-four absolute changes are at least 5 points.

Writing CI3 as the product of the enrolled-to-population ratio and the weighted-participant-to-enrolled ratio gives an exact midpoint decomposition. The survey component is larger in absolute value in 50 of 73 economies. Descriptive normalised variance components are 0.826 for the survey channel, 0.329 for the enrolment channel and −0.155 for twice their covariance; these sum to one by construction and are not causal shares.

Two final CI3 values exceed one (Korea and Ireland, 2022), and six further enrolled-to-population ratios exceed one. These values flag numerator--denominator combinations requiring reconciliation under the ratios' stated interpretation; they do not identify whether definitions, revisions, weighting adjustments or another input should change.

S6. Agreement between PISA POP15 and WPP

For 72 matched countries, PISA and WPP age-15 cohort growth have Pearson correlation 0.7933 and Spearman correlation 0.7333. An equal-error Deming description of PISA growth on WPP growth has intercept −1.7284 percentage points and slope 0.99125. A paired bootstrap with 100,000 resamples and seed 20260814 gives a 95% interval of 0.80982–1.17256 for the slope.

The paired growth discrepancy has SD 5.4041 percentage points. Absolute discrepancies exceed 5 points in 25 of 72 countries and 10 points in 7 of 72; the two growth series have opposite signs in 17 of 72. Separately, the two denominators imply opposite signs for the resulting CI3 change in 25 of 72. These are distinct diagnostics and must not be conflated.

For the 38-country Eurostat subset, pairwise mean absolute growth differences are 4.74 points for PISA–WPP, 3.90 for PISA–Eurostat and 3.31 for WPP–Eurostat. All three implied CI3-change signs agree in 17 of 38 cases; PISA alone differs in 8. No series is designated as ground truth.

S7. Denominator sensitivity of the adjusted trend

Domain	n	Mean absolute contrast	Median	P90	Maximum	Above link error	Above twice link error
Mathematics	69	3.4342	2.5668	7.7272	11.1050	38	17
Reading	69	3.4526	2.6612	7.6754	13.4691	52	33
Science	69	3.4290	2.8297	7.6848	11.4213	51	30

The contrast is the absolute difference between the 2018–2022 change under PISA POP15 and under WPP. Link errors are descriptive reference values, not denominator-error standard errors.

For mathematics, the descriptive OLS regression of absolute trend contrast on absolute CI3-change contrast has intercept 0.6352, slope 0.7773 (SE 0.0587), R^2 0.7233 and n=69. It describes propagation through the quantile mapping; it is not an agreement test or a causal model.

Target-top-share sensitivity in mathematics is:

Top share	n	Mean	Median	P90	Maximum
0.10	69	2.3107	1.8304	5.4294	7.8219
0.25	69	3.4342	2.5668	7.7272	11.1050
0.50	69	6.7275	4.6473	14.4089	43.5116

The PISA- and WPP-denominator trends have different signs in 4 of 69 mathematics cases, 5 of 69 reading cases and 4 of 69 science cases. Five significance calls produced by the former fixed-probability replicate

implementation disappear under the literal replicate-specific rule: United Arab Emirates and United States in mathematics, and Belgium, Indonesia and Thailand in reading.

Unweighted complete-sample means across the same 69 economies are:

Domain	Tested mean	PISA-adjusted	WPP-adjusted	Correlation of adjusted changes
Mathematics	−11.4415	−11.7710	−12.9596	0.9631
Reading	−9.6118	−6.9652	−8.4574	0.9734
Science	−1.8703	+1.4583	+0.1000	0.9659

These are unweighted complete-sample summaries, not OECD averages.

S7b. Denominator comparability audit and reduced-mismatch analysis

S7b.1 What each source documents

PISA POP15 (Annex A2; collected as Sampling Task 7a field 1). Reference date: not published. Age definition: the PISA age-eligibility window, a single specified birth calendar year or an agreed 12-month age span, with an assessment period of at most 56 days. Population concept: not published. Production method: national submission; the PISA 2022 Technical Report shows that the data source and the year of the data were required from the national centre on Sampling Task 7a and verified against those sources where an English-language page was available, so provenance is collected and partly checked but not released. Territory: the PISA entity as published. Vintage: not published; the Table I.A2.2 note records retrospective re-sourcing for seven economies. The technical report defines Coverage Index 3 as `P/ST7a_1`, which identifies the denominator with that submitted field.

WPP 2024. Reference date: 1 July of the year, with the published 1 July series equal to the arithmetic mean of the 1 January values. Age definition: completed single years of age. Population concept: de facto population. Production method: reconciled cohort-component estimate. Territory: WPP country or area, footnoted per country. Vintage: the 2024 revision.

Eurostat `demo_pjan`. Reference date: 1 January of the year. Age definition: completed single years of age. Population concept: usually resident, legal residence or registered residence, varying by country. Production method: national register or census-based count. Territory: national territory, with documented restrictions for Cyprus, Georgia, Moldova and Ukraine. Vintage: current; 2021 onwards embeds the 2021–2022 census round, with revised back series for several member states.

Sources: the delivered Annex A2 workbook (StatLink `hpg9nd`), the WPP 2024 methodology report, and the Eurostat ESMS metadata for population statistics.

S7b.2 Harmonisable versus structurally non-harmonisable

Harmonisable with official data, and harmonised here:

1. calendar year: the external series are read for the PISA cycle year;
2. reference date between the two external series: WPP publishes both 1 July and 1 January single-age series, so both external sources can be placed on a common 1 January basis;
3. age definition: the delivered microdata contain standardised birth year and month, so each economy's eligible cohort and its completed-age composition at any external reference date are observable and can be applied to external single-age counts;
4. territory: PISA entities that are not whole national or economy units are identified and excluded rather than joined to a national series;
5. documented denominator vintage breaks: the seven economies named in the Table I.A2.2 note are identified and excluded.

Structurally non-harmonisable, and documented rather than removed:

a. PISA POP15's population concept, reference date, source and vintage are not published for any economy, so those dimensions cannot be aligned from the external side. The source and the year of the data exist in OECD's Sampling Task 7a records, so this is a disclosure gap rather than an absence of information, but it is not closable by an external analyst; b. Eurostat's reported reference population differs across member states, which is internal to Eurostat; c. Eurostat values from 2021 onwards embed a later census round than PISA's contemporaneous 2018 submission, and a past vintage cannot be restored; d. WPP is a reconciled model-based estimate rather than an administrative count, which is a production-method difference by construction.

S7b.3 Observed PISA age-eligibility window

Across 154 entity–cycles with usable birth data the minimum recorded age is 15.08–15.42 and the maximum 16.17–16.42, consistent with the documented eligibility rule. The observed birth window spans exactly 12 months in 145 entity–cycles and 13 or 14 months in nine (Iceland, Israel, Japan, Serbia, the United Kingdom and the United States in 2018; Japan, Korea and the United Kingdom in 2022). Weighted completed-age composition of the eligible cohort:

Reference date	Completed age 14	Completed age 15	Completed age 16
1 July of the cycle year	0.003 (0.000–0.165)	0.596 (0.476–1.000)	0.401 (0.000–0.524)
1 January of the cycle year	0.086 (0.000–0.658)	0.914 (0.342–1.000)	0.000 (0.000–0.010)

Mean share, with the range across entity–cycles in parentheses. Reading a single completed-age-15 count therefore does not target the cohort PISA sampled. Relative to the completed-age-15 specification, the age-aligned WPP denominator has mean ratio 0.9999, range 0.961–1.110 and mean absolute deviation 1.00%; the aligned Eurostat denominator has mean absolute deviation 0.08%. Changing WPP's reference date from 1 July to 1 January at completed age 15 moves the denominator by a mean absolute 1.34% and up to 11.2%.

The alignment assumes that the age composition of the whole eligible cohort matches the weighted composition of the tested sample. That assumption is stated rather than tested: it cannot be evaluated from the PISA microdata and aggregate population series used here, and population-level administrative or register data covering the full eligible cohort would be required to test it directly. Alignment also cannot be built where birth month or year is withheld from the released files: Iceland, Montenegro, Norway and Sweden in at least one cycle, and Baku (Azerbaijan) in both.

S7b.4 Comparability status of the 73 two-cycle systems

Status	n
Reduced mismatch on all checkable dimensions	52
Internal-consistency mismatch only (CI3 > 1 or enrolled > POP15)	7
Denominator vintage mismatch (OECD-documented re-sourcing)	6
Insufficient microdata for age alignment	4
No comparable student sample in both cycles	2
Multiple mismatches	2

The reduced-mismatch set is the 59 systems that are whole national or economy units with a one-to-one WPP match, have an observable age window in both cycles, carry no documented POP15 re-sourcing and reproduce the published CI3 identity. Of those, 58 have complete two-cycle trend estimates; Viet Nam does not, because every 2018 plausible value is missing in the delivered file. The strict set additionally requires CI3 to be at most one and enrolled 15-year-olds to be at most POP15, leaving 51.

Excluded from the reduced-mismatch set, with the recorded reason: Albania, Brazil, Chile, the Netherlands, Romania and Uruguay (documented POP15 re-sourcing); Jordan (re-sourcing, CI3 identity failure

and enrolled above POP15); Baku (Azerbaijan) (not a whole national unit, and no observable age window); Cyprus and Ukraine (no comparable student sample in both cycles); Iceland, Montenegro, Norway and Sweden (birth month or year withheld). Retained in the reduced set but not in the strict set: Brunei Darussalam, Canada, Finland, Malta and Slovenia (enrolled above POP15) and Ireland and Korea (CI3 above one). The system-by-system table is `results/denominator_comparability_audit.csv`.

S7b.5 Contrast after reducing avoidable mismatch

Sample	Denominator	Domain	n	Mean	Median	P90	Maximum	Above link error
All systems	WPP age 15	Mathematics	69	3.4342	2.5668	7.7272	11.1050	38
All systems	WPP aligned	Mathematics	65	3.4306	2.4337	7.1859	11.4820	37
Reduced mismatch	WPP age 15	Mathematics	58	3.4574	2.7359	7.2711	10.9969	34
Reduced mismatch	WPP aligned	Mathematics	58	3.2933	2.4361	6.7714	10.7272	33
Strict	WPP aligned	Mathematics	51	3.3238	2.4385	6.6295	10.7272	29
All systems	WPP age 15	Reading	69	3.4526	2.6612	7.6754	13.4691	52
All systems	WPP aligned	Reading	65	3.5113	2.6393	7.4690	12.9901	51
Reduced mismatch	WPP age 15	Reading	58	3.6399	3.3238	7.6997	13.4691	44
Reduced mismatch	WPP aligned	Reading	58	3.5462	2.8483	7.3260	12.9901	45
Strict	WPP aligned	Reading	51	3.5854	3.0572	7.6598	12.9901	39
All systems	WPP age 15	Science	69	3.4290	2.8297	7.6848	11.4213	51
All systems	WPP aligned	Science	65	3.5025	2.8469	7.3145	11.2025	49
Reduced mismatch	WPP age 15	Science	58	3.5224	3.0621	7.6950	11.4213	43
Reduced mismatch	WPP aligned	Science	58	3.4395	2.8251	7.1617	11.2025	43
Strict	WPP aligned	Science	51	3.4514	2.9622	7.1721	11.2025	37

Score points. Sign disagreements under the reduced-mismatch aligned specification are 4, 6 and 2 of 58, of which 0, 1 and 0 have a statistically significant PISA-denominator change. Unweighted complete-sample means under the aligned specification (n = 65) are tested −10.4007, PISA-adjusted −10.7697 and aligned-WPP-adjusted −11.3900 in mathematics; −8.7290, −6.1004 and −6.9940 in reading; −1.0903, +2.1411 and +1.3273 in science. The mathematics top-share gradient under the aligned specification is 2.2956, 3.4306 and 6.9073 for top shares 0.10, 0.25 and 0.50.

For the Eurostat branch, the mean absolute PISA–Eurostat contrast is 2.9791, 3.1217 and 3.0811 points at completed age 15 (n = 36) and 3.3538, 3.4970 and 3.4594 points under alignment (n = 32).

The reduction of avoidable mismatch therefore leaves the magnitude, direction and interpretation of the reported sensitivity intact. It falls by about 0.14 points in mathematics and rises by about 0.09 and 0.01 points in reading and science relative to the original specification.

S8. Annex A7 zero-replacement boundary

The documented rule affects six economy-cycle cells, or 18 economy-cycle-domain cells at `p=0.75`.

Cycle	Economy	Final synthetic weight	Replicates zeroed	Mathematics A/B	Reading A/B	Science A/B
2018	Germany	+4,876.72	30/80	0.982	0.999	0.994
2018	Hong Kong	+834.03	1/80	0.999	0.999	0.999
2022	Belgium	+1,172.02	9/80	0.987	0.985	0.981
2022	Ireland	−1,445.56	80/80	0.946	1.051	0.953
2022	Korea	−9,983.64	64/80	0.910	0.707	0.827
2022	Slovenia	+81.76	2/80	1.000	0.999	0.999

“Hong Kong” in this compact table denotes the published entity Hong Kong (China).

A/B is the standard error under Variant A divided by that under Variant B. Variant A is the documented zero-replacement rule. Variant B retains negative synthetic weights and is an explicitly invalid smooth continuation used only to describe boundary sensitivity. The minimum unrounded ratio is 0.706773; 17 ratios are below one and one is above one. Because Variant B is invalid, these ratios do not estimate bias, undercoverage or the amount by which a valid standard error is over- or understated.

Final-weight zeroing switches off the point adjustment for Korea and Ireland in 2022. Germany 2018, Hong Kong (China) 2018, Belgium 2022 and Slovenia 2022 have positive final synthetic weights but zeroing in some replicates, so their point estimates are unchanged while a subset of replicate probabilities follows the piecewise rule.

S9. Selected corrected country results

Economy	Mathematics PISA-adjusted trend	WPP−PISA contrast	SE including link error
Korea	+14.2699	−10.5208	5.7675
Ireland	−0.9214	−5.0511	4.0845

For the Philippines, tested-mean changes are +2.15 in mathematics, +6.86 in reading and −0.77 in science. PISA-denominator top-quarter changes are +8.73, +31.23 and +16.70; absolute PISA–WPP contrasts are 1.43, 0.30 and 0.11. The case is descriptive and depends on the same strong lower-tail assumption as every other adjusted estimate.

S10. Closure-duration checks

Across five closure-duration measures, Pearson correlations with CI3 change range approximately from −0.14 to +0.22 in samples of 67–70. A nominal Pearson association in one specification is not reproduced by its rank correlation, and pre-pandemic placebo associations are at least as large. The result is therefore reported as no robust association in these checks, not as proof of no confounding in other models.

S11. Bounded literature search

A bounded search completed on 14 August 2026 used 24 targeted queries across Crossref, OpenAlex, Semantic Scholar, arXiv, Europe PMC and ERIC, citation sweeps over 23 anchor works, and screening of 604 records, including 38 high-threat candidates. No located study combined the exact denominator-propagation audit reported here. This is a dated search null, not a claim of universal novelty or guaranteed journal acceptance. The query list, the anchor works swept and the screening decisions are held by the authors and can be supplied to the editor on request.

S12. Limitations and interpretation

1. PISA POP15, WPP and Eurostat differ in concept, vintage and reference date; none is treated as truth. Calendar year, age window and territory can be aligned; concept and vintage cannot, because Annex A2 reports none of them.
2. The coverage-invariant statistic assumes every non-covered student lies below the target threshold.
3. Named-scenario contrasts are not confidence intervals and are not added to sampling or link-error variance.
4. Variant B is invalid and cannot benchmark the correctness of Variant A.
5. Mode, scaling and adaptive-design changes are outside the estimator.
6. Country-level closure checks are ecological and imprecise.
7. The weighted-percentile convention is identified from OECD's published percentile values rather than from a documented rule. Agreement on published output is strong empirical evidence for the operative convention but does not establish identity with the exact production implementation used for the Annex A7 statistic, and the identification rests on one published table for one cycle and one domain. The

documented macro corroborates the per-plausible-value combination rule but delegates the percentile to the host statistical package, so the interpolation rule remains unstated in the documentation examined.

8. The age-window alignment in S7b assumes that the age composition of the whole eligible cohort matches the weighted composition of the tested sample; testing that assumption would require population-level administrative or register data for the full eligible cohort. The alignment cannot be built where birth month or year is withheld from the released files.
9. This design supplies no common cross-national criterion for deciding which population series is closest to PISA's eligible cohort. Country-specific registers can adjudicate individual systems, as in the Swedish comparison of PISA with Statistics Sweden register data, but equivalent evidence is not available on comparable terms across the analytic sample. The reported quantity is therefore dependence on denominator choice, not denominator error.

S13. Contents of Online Resource 2

`ESM_2.zip` contains:

- `src/pcov/`, the documented Annex A7 estimator module together with the entity crosswalk and label helpers it uses;
- `data/interim/`, the harmonised inputs, including the coverage panel, the CI3 series under each denominator, the aligned denominators, the observed PISA age window and the external population and closure series;
- `results/`, the estimation outputs, including the by-entity estimates, the Annex A7 variants, the percentile-convention calibration and its deviations, the denominator comparability audit and the primary-source verification record;
- `tables/`, the machine-readable form of every table reported in the manuscript and in this file;
- `figures/`, the four figures in vector PDF and high-resolution PNG form;
- `registry/provenance.json`, recording for each of the sixteen inputs its URL, retrieval date, byte count and SHA-256 digest.

Every number in the manuscript and in this file can be traced to a file in `results/` or `tables/`, and every input can be traced through `registry/provenance.json` to its published source.

The archive carries derived data, the estimator module and provenance. It does not carry the driver scripts, the pinned dependency list or the manuscript sources: the manuscript is supplied separately as the manuscript upload, and the analysis environment is recorded in S2 above. Raw PISA microdata are not redistributed, and temporary build files, absolute local paths and operating-system metadata files are excluded.