Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

When the adjustment fails: population-denominator sensitivity in coverage-adjusted PISA trends

Jonas A. Mandalunes

jonas.mandalunes@perpetual.edu.ph

University of Perpetual Help System DALTA

Danica Jane S. Mandalunes

Saint Columban College

Research Article

Keywords: PISA, population coverage, Coverage Index 3, trend estimation, balanced repeated replication, denominator sensitivity

Posted Date: August 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10706435/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

When the adjustment fails: population-denominator sensitivity in coverage-adjusted PISA trends

Abstract

Programme for International Student Assessment (PISA) cycles sample different shares of each economy's 15-year-old population, and the OECD's coverage-adjusted top-quarter statistic depends on POP15, the reported population denominator. We analyse 2018–2022 Coverage Index 3 (CI3) changes for 73 economies and top-quarter trends for 69 under two named denominator scenarios: PISA POP15 and World Population Prospects (WPP). Estimates use 10 plausible values, 80 Fay balanced-repeated-replication replicates and a weighted-percentile convention calibrated to OECD's published percentiles. CI3 changed by a mean 1.82 percentage points (SD 8.64; range −45.19 to 27.05), with 24 absolute changes at least 5 points; an exact decomposition assigned the larger component to the survey channel in 50 economies. Mean absolute PISA–WPP trend contrasts were 3.4342, 3.4526 and 3.4290 points in mathematics, reading and science, exceeding the published link-error reference value for 38, 52 and 51 of 69 economies; link errors serve as descriptive references, not denominator-error standard errors. Aligning the external series to each economy's observed PISA age-eligibility window and removing systems with documented territorial, vintage or consistency problems leaves contrasts of 3.29, 3.55 and 3.44 points in 58 economies. Annex A7's documented synthetic-weight and zero-replacement rule was triggered in 6 of 159 economy–cycle cells: final-weight zeroing switched off the adjustment for Korea and Ireland; four further cases affected replicates only. Ratios to an explicitly invalid negative-weight continuation do not estimate bias, and neither population source is treated as ground truth. We recommend transparent POP15 provenance, named-scenario sensitivity reporting and non-reporting when synthetic weights are negative.

Keywords: *PISA; population coverage; Coverage Index 3; trend estimation; balanced repeated replication; denominator sensitivity*

Introduction

Cross-cycle comparisons in PISA are differences between independent cross-sections. The sampled students in each cycle represent a country- or economy-specific fraction of all 15-year-olds, and that fraction can change. OECD publications recognise the interpretive consequence: when newly represented students differ systematically from those already represented, a conventional trend can combine achievement change with a change in population coverage (OECD, 2017, 2023). OECD therefore reports a coverage-invariant statistic defined as the minimum score attained by the top quarter of the full age group. Its construction is documented in PISA 2018 Results Volume I, Annex A7 (OECD, 2019). OECD also states the limits of its own construction. Because the adjustment assumes that non-eligible 15-year-olds would not have scored in the top quarter, the reported estimates are "in reality, lower bounds on the true 75th percentiles", and those bounds move closer to the true value as Coverage Index 3 rises. In that context, "the reported changes and trends may over-estimate the true changes and trends" (OECD, 2019, Note 11). The concern addressed here is therefore not that the assumption is undisclosed. It is that the statistic's population input, which determines how much of the tested distribution the assumption is applied to, is neither published with its provenance nor stable across sources.

The statistic requires a count of all 15-year-olds. That count, POP15, is supplied to the OECD by national authorities and appears in Annex A2. It is the denominator of Coverage Index 3 (CI3),

$$\mathrm{CI3} = \frac{\text{weighted PISA participants}}{\mathrm{POP15}},$$

and it enters every percentile adjusted through this procedure. POP15 is therefore not merely descriptive metadata: changing it can change the percentile selected from the tested distribution and, consequently, the estimated trend.

This study asks three linked questions. First, how much did CI3 change from 2018 to 2022, and how does its exact accounting decomposition divide that change between enrolment and survey coverage? Second, how sensitive is the coverage-invariant trend to two named age-15 population series, PISA POP15 and United Nations World Population Prospects (WPP), with Eurostat used as a third series for a European subset? Third, what does the documented Annex A7 algorithm do when POP15 is less than the sum of final or replicate weights? The estimands are scenario contrasts. No denominator-error model is fitted, and none of the population series is designated as ground truth.

Data

We use the PISA 2018 and 2022 student questionnaire files, containing 612,004 and 613,744 students, respectively, across 80 economies per cycle. The files provide the final student weight, 80 replicate weights and 10 plausible values in mathematics, reading and science. Annex A2 of PISA 2022 Volume I supplies POP15, the grade-7-or-higher enrolled population, weighted participants and CI3 for seven cycles (OECD, 2023).

The first named alternative is the 2024 revision of WPP, from which we extract the mid-year age-15 population for each survey year (United Nations, Department of Economic and Social Affairs, Population Division, 2024). For 38 countries, Eurostat population on 1 January by single year of age supplies a third series (Eurostat, 2025). PISA POP15 follows the national target-population definitions reported in Annex A2. The series differ in reference date, production method, vintage and, for some places, population concept or territorial scope; the analysis therefore compares named scenarios and does not interpret pairwise differences as measurement error.

The three sources also differ in what they document. WPP publishes a stated reference date and a documented estimation method: single-age counts are produced by a reconciled cohort-component projection, and the published 1 July series is the arithmetic mean of the 1 January values (United Nations, Department of Economic and Social Affairs, Population Division, 2024). Eurostat publishes a stated reference date of 1 January and documents that the reported reference population is the usually resident population in some member states, the legal residence population in others and the registered residence population in others again, and that population from 2021 onwards incorporates the 2021–2022 census round with revised back series for several member states (Eurostat, 2025).

The PISA side is documented differently, and the distinction matters. The PISA 2022 Technical Report identifies the CI3 denominator by its collection field: Coverage Index 3 is defined as $P/\mathrm{ST7a_1}$, where $\mathrm{ST7a_1}$ is the total national number of 15-year-olds submitted by the national centre on Sampling Task 7a (OECD, 2025). For that submission the international contractor required the data sources and the year of the data, and verified the

submitted value against those sources where an English-language page was available (OECD, 2025). Provenance is therefore collected and partly checked. It is not published. Annex A2 reports the same quantity as a single column headed "Total population of 15-year-olds", with no accompanying population concept, reference date, source or vintage for any economy. Its only related note records that POP15 was retrospectively re-sourced for Albania, Brazil, Chile, Jordan, the Netherlands, Romania and Uruguay, so those values differ from earlier PISA reports (OECD, 2023). Section 4.4 uses this asymmetry to separate mismatches that can be reduced from mismatches that cannot.

The same technical report fixes the age concept. Age-eligible students are those born in a single specified calendar year, or in the appropriate 12-month age span agreed for the participating economy, and the assessment period may not exceed 56 days without agreement (OECD, 2025). The eligible population is thus a 12-month date-of-birth window rather than the population at one completed age on a fixed date, which is the mismatch quantified in Section 4.4.

School-closure checks use the Oxford COVID-19 Government Response Tracker (Hale et al., 2021) and the UNESCO Institute for Statistics closure-duration file (UNESCO Institute for Statistics, 2022). These checks speak only to the empirical association between CI3 change and the available duration measures. They are interpreted alongside cross-country closure studies that use PISA outcomes (Jakubowski et al., 2025; Kennedy & Striethолt, 2023).

The analytic samples are 82 economies with 2022 coverage data, 73 with coverage data in both 2018 and 2022, 72 matched in the PISA–WPP cohort-growth comparison, 69 with complete trend estimates under both primary scenarios, 38 with all three population series, and 159 economy–cycle cells in the Annex A7 implementation. Of the 73 two-cycle coverage cases, Baku lacks a comparable national WPP series, Cyprus is absent from the supplied student files, the 2018 national Ukraine and 2022 partial-region sample are not comparable, and Viet Nam has no usable 2018 plausible values. Four cross-cycle label changes (Hong Kong, Macao, Moldova and Türkiye) are harmonised explicitly before estimation.

Methods

Plausible values and Fay-BRR

Following the PISA 2018 and 2022 reporting framework (OECD, 2019, 2023), let $U = 10$ index plausible values and $R = 80$ index replicates. For plausible value u, let $\hat{\theta}_u$ be the estimate using the final weight and $\hat{\theta}_{ur}$ its estimate using replicate r. With Fay's $k = 0.5$, the sampling component is

$$V_{\text{samp}} = \frac{1}{U}\sum_{u=1}^{U}\left[0.05\sum_{r=1}^{R}\left(\hat{\theta}_{ur} - \hat{\theta}_u\right)^2\right]. \qquad (1)$$

Writing $\bar{\theta} = U^{-1}\sum_u \hat{\theta}_u$, the plausible-value component is

$$V_{\text{imp}} = \left(1 + \frac{1}{U}\right)\frac{\sum_{u=1}^{U}\left(\hat{\theta}_u - \bar{\theta}\right)^2}{U-1}, \qquad U = 10, \qquad (2)$$

and $V_{\text{total}} = V_{\text{samp}} + V_{\text{imp}}$. For a trend between the independent 2018 and 2022 cross-sections, the reported link-error-inclusive standard error in domain d is

$$\mathrm{SE}(\Delta_d) = \left(V_{\mathrm{total},d,2018} + V_{\mathrm{total},d,2022} + L_d^2\right)^{1/2}, \qquad (3)$$

where L_d is the published domain link error (2.24 for mathematics, 1.47 for reading and 1.61 for science) reported in PISA 2022 Results Volume I, Table I.A7.1 (OECD, 2023). Equation 3 is used for the trend SEs and significance checks reported below; it does not convert a named-scenario contrast into a sampling standard error. Each economy–cycle–domain estimate requires 10 final-weight calculations plus 10×80 replicate-weight calculations: 810 weighted computations, not 800.

Exact decomposition of CI3 change

For cycle t, let P_t be POP15, E_t the number of enrolled 15-year-olds in grade 7 or above, and W_t the weighted participant count. Define $A_t = E_t/P_t$ and $B_t = W_t/E_t$. Then

$$\mathrm{CI3}_t = \frac{W_t}{P_t} = \underbrace{\frac{E_t}{P_t}}_{A_t\text{: enrolment channel}} \underbrace{\frac{W_t}{E_t}}_{B_t\text{: survey channel}}. \qquad (4)$$

For cycles 0 and 1, the midpoint decomposition is exact:

$$\Delta\mathrm{CI3} = \underbrace{\frac{B_1 + B_0}{2}(A_1 - A_0)}_{\text{enrolment contribution}} + \underbrace{\frac{A_1 + A_0}{2}(B_1 - B_0)}_{\text{survey contribution}}. \qquad (5)$$

The residual is at floating-point precision. Cross-economy variance shares are descriptive normalisations: the two component variances and twice their covariance are divided by the variance of $\Delta\mathrm{CI3}$.

Coverage-invariant statistic

Let W be the final-weight sum, P the selected population denominator and s the top share of the full age group. Under the identifying assumption that every non-covered 15-year-old falls below the target threshold, and before the zero-replacement boundary, the minimum score attained by the top sP students is the tested-sample quantile at

$$q = 1 - \frac{sP}{W} = 1 - \frac{s}{\mathrm{CI3}}, \qquad s < \mathrm{CI3} \leq 1. \qquad (6)$$

The lower bound ensures that the required tested-sample probability is positive. When $\mathrm{CI3} > 1$, the documented zero-replacement rule instead sets the negative synthetic weight to zero and gives $q = 1 - s$, as detailed below. The main analysis uses $s = 0.25$; $s = 0.10$ and $s = 0.50$ are sensitivity cases. The identifying assumption is strong and not testable with the observed student files.

Documented Annex A7 calculation at final and replicate weights

Annex A7, in the section on adjusted trends accounting for changes in enrolment, specifies the procedure operationally (OECD, 2019). A single case per economy is added to the database to represent all 15-year-olds not covered by the PISA sample. Its final student weight is the difference between the total population of 15-year-olds, referenced to Table I.A2.2, and the sum of final student weights. Each replicate weight for that case is the difference between the total population of 15-year-olds and the sum of the corresponding replicate weights. Any negative weights resulting from this procedure are replaced by zero. A

value below every plausible value in the sample is entered as its performance, and the median and upper percentiles are then computed on the augmented sample; where the coverage rate is below 50%, the adjusted median is reported as missing. For target percentile $p = 1 - s$, the documented final-weight calculation is

$$S = \max(P - W, 0), \quad T = S + W, \quad q = \frac{pT - S}{W}. \qquad (7)$$

The procedure must be repeated for each replicate-weight sum W_r:

$$S_r = \max(P - W_r, 0), \quad T_r = S_r + W_r, \quad q_r = \frac{pT_r - S_r}{W_r}, \qquad r = 1, \ldots, 80. \qquad (8)$$

Thus both the final quantile probability and all replicate probabilities can vary with the weight sums. If any required probability is non-finite or outside the open interval $(0,1)$, that estimate is labelled infeasible; probabilities are never clipped. This feasibility rule is not an addition to the documented procedure but a restatement of it: Equation 6 requires $s < \mathrm{CI3}$, which for the median reproduces exactly the condition under which Annex A7 reports the adjusted median as missing.

We calculate three variants. Variant A is the documented rule in Equations 7–8. Variant B retains $P - W$ and $P - W_r$ even when negative. Variant B is an explicitly invalid smooth continuation, because negative synthetic weights have no population interpretation. It is used only to describe boundary sensitivity. Variant C substitutes the WPP denominator. Comparisons of A with B cannot identify the direction or magnitude of standard-error bias because the denominator of that ratio, Variant B, is invalid.

Weighted-percentile convention

Equations 6–8 determine *which* probability is read from the tested distribution. They do not determine how a weighted percentile is evaluated at that probability, and different conventions give slightly different answers. We searched the three relevant OECD documents for a stated rule. Annex A7 says only that "the median and upper percentiles of the distribution are computed on the augmented sample" (OECD, 2019). The PISA 2022 Technical Report contains no weighted-percentile definition (OECD, 2025). The PISA Data Analysis Manual supplies a percentile macro, MCR_SE_PERCENTILES_PV, which weights the file by the final weight, calls the host package's frequency procedure to return the requested percentiles for each plausible value, repeats that call once per replicate weight, and combines the results with the usual sampling and imputation components of Equations 1 and 2 (OECD, 2009). The macro therefore fixes the combination rule (one percentile per plausible value, then averaged across plausible values with compute STAT=mean) but delegates the percentile itself to the statistical package and states no interpolation rule. None of the three documents contains the terms weighted percentile, weighted quantile, cumulative weight or interpolation in this sense. The weighted-percentile rule used to produce the Annex A7 statistic is therefore not established by the documentation we examined, and it cannot be copied from it.

It can, however, be identified from OECD output. 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 every participating entity (OECD, 2023). We recomputed those three percentiles from the delivered 2022 microdata for all 80 entities under five standard weighted-percentile definitions and two plausible-value combination rules, giving 240 published-value comparisons per candidate. Writing $x_{(1)} \leq \cdots \leq x_{(n)}$ for the sorted scores with weights $w_{(k)}$,

cumulative weights C_k and total weight W, the candidate that reproduces the published values is the inverse empirical distribution function computed separately for each plausible value and then averaged:

$$\hat{Q}(q) = x_{(k^*)}, \qquad k^* = \min\{k: C_k \geq qW\}. \qquad (9)$$

Its median absolute deviation from the published percentiles is 0.0000 score points, its mean absolute deviation 0.0004, and it agrees to the published four-decimal precision for 77 of 80 entities on all three percentiles; the largest single deviation is 0.0434 points in one small sample. The interpolating alternatives are further away and match no entity to four decimals: mean absolute deviations of 0.0070 points for the midpoint (Hazen-type) rule used in our first submission, 0.0286 and 0.0297 for the two edge-anchored linear rules. Pooling plausible values before computing the percentile is clearly not OECD's rule (mean absolute deviation 0.063 points), which independently reproduces the per-plausible-value combination that the documented macro prescribes. Equation 9 is therefore used throughout. This is an empirical identification of OECD's convention against OECD's published output, not a claim to have recovered OECD's code. The implementation sensitivity that remains is reported in Section 4 and is an order of magnitude smaller than any substantive quantity here.

Validation

The analysis was checked against published means, weight sums, coverage values and closure durations before substantive comparisons. Table 1 reports the checks. The CI3 arithmetic identity holds in 422 of 424 entity–cycle observations. The two exceptions are transparent: Jordan 2022 has a published value of 0.940218 versus a reconstructed ratio of 0.966865 after a retrospective denominator revision, and the Ukrainian-regions row is published as 0.639000 versus a reconstructed ratio of 0.639415.

Table 1 *Validation against published OECD and closure-source values*

Check	Reproduced	Published/source value
Philippines 2022 mathematics mean	354.7197	354.72
Philippines 2022 reading mean	346.5474	346.55
Philippines 2022 science mean	356.1673	356.17
Philippines $\sum W_{\mathrm{FSTUWT}}$, 2022	1,782,895.5818	1,782,895.59
Philippines $\sum W_{\mathrm{FSTUWT}}$, 2018	1,400,584.0599	1,400,584.05
CI3 $= \sum W_{\mathrm{FSTUWT}}/\mathrm{POP15}$	422/424	422/424
Annex A2.1 versus A2.2 CI3, 2022	82/82 equal	82/82 equal
Philippines OxCGRT closure days	621 level 3; 716 level ≥ 2	621; 716
Philippines UNESCO closure days	556 full; 6 partial	556; 6

Note. The two CI3 exceptions are Jordan 2022 (0.940218 published, 0.966865 reconstructed) and Ukrainian regions (0.639000 published, 0.639415 reconstructed). They are retained and documented rather than forced to equality.

Results

Magnitude and decomposition of CI3 change

Across 73 economies, the published 2018–2022 CI3 change had mean $+1.82$ percentage points, SD 8.64 and range -45.19 to $+27.05$ points. Twenty-four changes were at least 5 points in absolute value. The range includes cases affected by war, population-series revisions, survey participation and changes in the represented territory; it should not be read as a ranking of survey quality.

Equation 5 assigns a normalised variance share of 0.826 to the survey channel and 0.329 to the enrolment channel, with a covariance contribution of -0.155; the three terms sum to one by construction. The survey contribution is larger in absolute value in 50 of 73 economies. Fig. 1 shows the economy-level split for the 24 systems whose absolute change is at least 5 points. For the Philippines, a 15.4-point reconstructed increase consists of approximately -1.4 points from the enrolment channel and $+16.8$ from the survey channel. These are algebraic components, not causal effects.

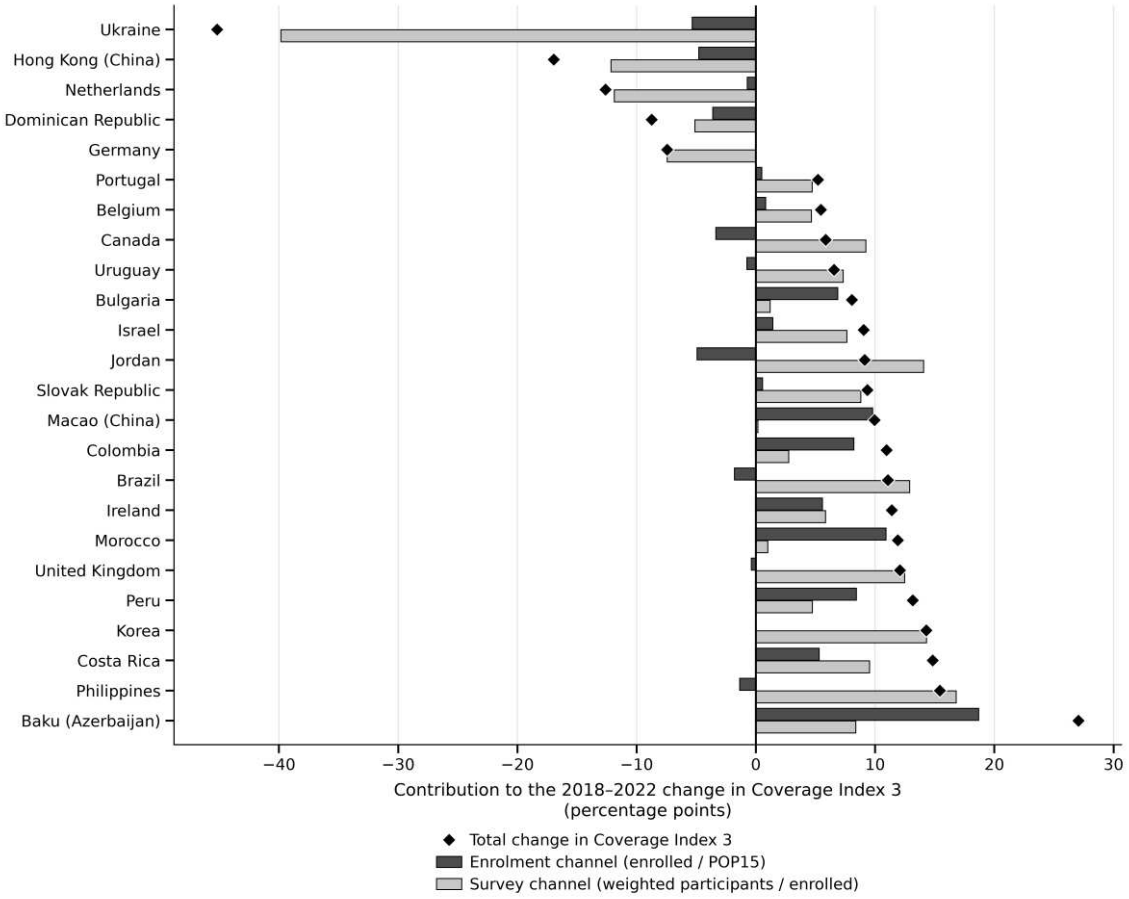


Fig. 1 *Exact accounting decomposition of the 2018–2022 change in Coverage Index 3 for economies with an absolute change of at least 5 percentage points. The midpoint contributions sum to the reconstructed change. Labels "enrolment" and "survey" describe the factors in Equation 4; the bars do not establish causal mechanisms*

A rising CI3 therefore need not mean that more 15-year-olds entered education. It can instead reflect a changed share of enrolled students represented by the assessment, a revised population count, or both. This distinction matters for any trimming argument that would

interpret a change in coverage as monotone entry or exit at the bottom of a latent achievement distribution.

Population-series contrasts

Two 2022 CI3 values exceed one: Korea (1.024) and Ireland (1.023). Six further enrolled-to-population ratios exceed one. These ratios flag numerator–denominator combinations that require reconciliation under their stated interpretation, but do not by themselves identify whether definitions, revisions, weighting adjustments or another input should change. Under the WPP scenario, no economy has CI3 above one.

For 72 countries, PISA and WPP age-15 growth have Pearson correlation 0.7933 and Spearman correlation 0.7333; Fig. 2 plots the two series against the identity line. An equal-error Deming fit 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 points; 25 of 72 absolute discrepancies exceed 5 points and 7 exceed 10 points. The two series imply opposite growth signs for 17 of 72 countries. These are named-scenario contrasts, and the interval reflects resampling of country pairs under the specified fit rather than a denominator-error model.

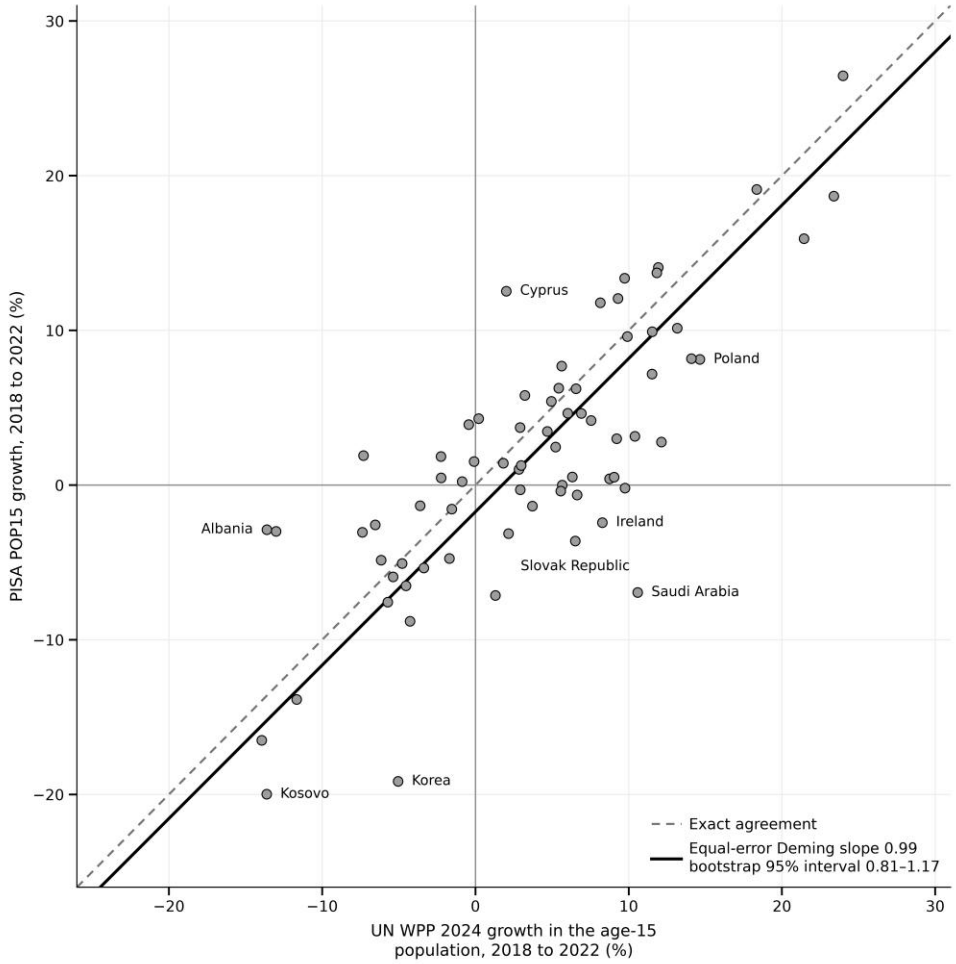


Fig. 2 *PISA POP15 and WPP age-15 population growth, 2018–2022, for 72 matched countries. The fitted line is the equal-error Deming description reported in the text; the identity line is a reference. Neither series is treated as ground truth, and the plot does not adjudicate country-specific differences*

For 38 countries, Eurostat provides a third series. The pairwise mean absolute growth differences are 4.74 percentage points for PISA versus WPP, 3.90 for PISA versus Eurostat and 3.31 for WPP versus Eurostat. All three CI3-change signs agree in 17 of 38 cases; PISA alone differs in 8 of 38. This three-way pattern does not select one universally preferable series.

Table 2 focuses on the four European economy–cycle cases in which zero-replacement later binds. At the binding cycle, both WPP and Eurostat yield lower CI3 than PISA in all four. Korea, the fifth binding case, is not in the table because Eurostat does not provide the required comparable series.

Table 2 *Coverage Index 3 under three named population-denominator scenarios in four European boundary cases*

		PISA POP15			**WPP**	**Eurostat**
Economy	2018	2022	2018	2022	2018	2022
Belgium	0.9365	0.9910	0.9559	0.9540	0.9620	0.9572
Germany	0.9934	0.9189	0.9650	0.9026	0.9667	0.9208
Ireland	0.9086	**1.0226**	0.9269	0.9402	0.9272	0.9525
Slovenia	0.9793	0.9957	0.9433	0.9094	0.9352	0.9361

Note. Bold marks a ratio above one. Values are scenario-specific weighted-participant counts divided by the named population count. The table does not designate any source as ground truth. Korea lacks the Eurostat comparison and is therefore omitted.

Sensitivity of the coverage-invariant trend

Replacing PISA POP15 with WPP changes both cycle-specific quantile levels and their difference. Fig. 3 displays the country-level propagation in mathematics and the domain distributions. Table 3 compares absolute scenario contrasts with the published link errors used as descriptive reference values.

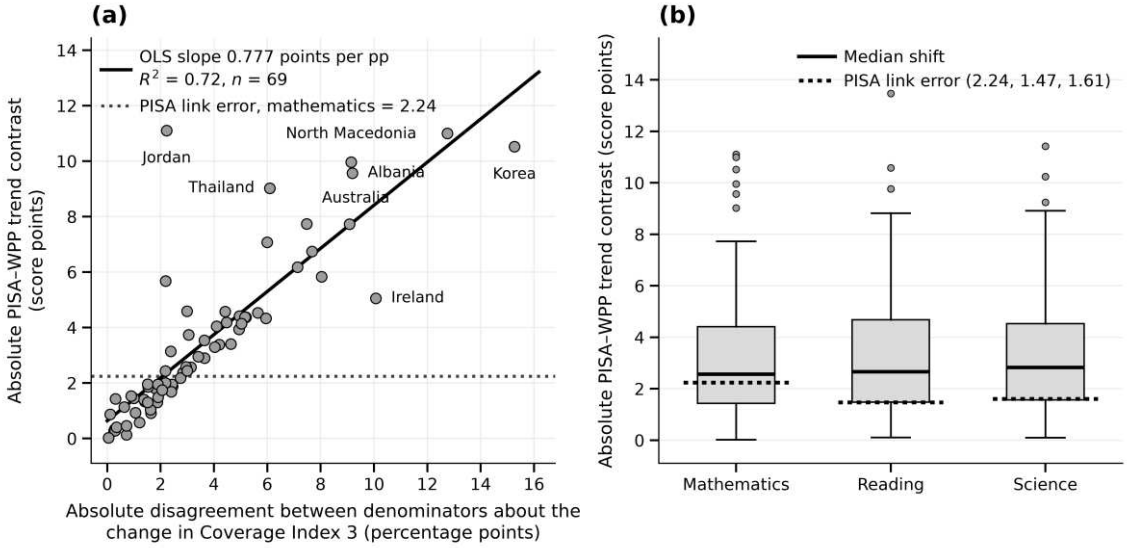


Fig. 3 *Named-scenario sensitivity of the 2018–2022 top-quarter trend. Panel (a) relates the absolute PISA–WPP disagreement in CI3 change to the absolute mathematics trend contrast; the fitted OLS line is descriptive and has no causal interpretation. Panel (b) shows domain-*

specific contrast distributions relative to published link-error reference values. Neither denominator scenario is treated as correct

Table 3 *Absolute PISA–WPP scenario contrast in the coverage-invariant trend*

Domain	n	Mean	Median	P90	Maximum	$>$ link	$>2\times$link
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

Note. Entries are absolute score-point contrasts between 2018–2022 changes under the PISA POP15 and WPP scenarios. Published link-error reference values are 2.24 for mathematics, 1.47 for reading and 1.61 for science. Counts above link error are 38/69, 52/69 and 51/69, respectively.

For mathematics, a descriptive OLS regression of the absolute trend contrast on the absolute CI3-change contrast has intercept 0.6352, slope 0.7773 score points per percentage point (SE 0.059), $R^2 = 0.7233$ and $n = 69$. This relationship summarises propagation through the quantile formula; it is neither a causal coefficient nor an estimate from a population-denominator error model.

Table 4 reports selected large contrasts and includes Ireland separately as a final-weight boundary case. The signed contrast is WPP minus PISA. Several estimates move materially without changing sign; sign changes should be interpreted against their standard errors rather than as categorical evidence that one scenario is preferable.

Table 4 *Selected large and boundary-case mathematics shifts under the two denominator scenarios*

Economy	Tested mean	PISA-adjusted	WPP-adjusted	WPP−PISA	SE
Jordan	−38.534	−48.087	−36.982	+11.105	5.567
Saudi Arabia	+15.543	+3.900	−7.097	−10.997	5.168
Korea	+1.369	+14.270	+3.749	−10.521	5.768
North Macedonia	−5.871	−17.985	−8.027	+9.959	3.513
Albania	−69.000	−68.527	−58.963	+9.564	5.038
Thailand	−24.610	−32.835	−23.816	+9.019	5.688
Australia	−4.276	−0.107	−7.838	−7.731	4.290
Slovak Republic	−22.170	−12.167	−19.894	−7.726	5.323
Ireland (boundary case)	−7.985	−0.921	−5.972	−5.051	4.084

Note. Score-point changes from 2018 to 2022. “Adjusted” denotes the top-quarter statistic under the named denominator scenario. SE is for the PISA-denominator change and includes the mathematics link error. Ireland is displayed separately because final-weight zeroing occurs in 2022.

Sensitivity increases as the target top share expands. In mathematics, mean absolute PISA–WPP contrasts are 2.311, 3.434 and 6.728 points for top shares 0.10, 0.25 and 0.50; the

corresponding maxima are 7.822, 11.105 and 43.512. This gradient is consistent with the larger denominator leverage in Equation 6 for larger s.

Reduced-mismatch denominator comparison

Table 3 compares POP15 with the WPP count at completed age 15. That comparison carries an age-definition mismatch that can be measured and removed. PISA's eligible population is a 12-month date-of-birth window, not the population at completed age 15 on a fixed date (OECD, 2025). The delivered microdata confirm the window: across entity–cycles the minimum recorded age is 15.08–15.42 and the maximum 16.17–16.42, and the observed birth window spans exactly 12 months in 145 of 154 entity–cycles, with 13 or 14 months in the remaining nine. Applying each student's birth year and month, the eligible cohort is on average 59.6% at completed age 15 and 40.1% at completed age 16 on 1 July of the cycle year, and 91.4% at completed age 15 and 8.6% at completed age 14 on 1 January. Reading a single completed-age-15 count therefore does not target the same cohort that PISA sampled.

We construct an aligned alternative by blending the external single-age counts with the observed composition of the eligible cohort at the external series' own reference date. Relative to the completed-age-15 specification, the aligned WPP denominator has a mean ratio of 0.9999 but ranges from 0.961 to 1.110, with a mean absolute deviation of 1.00%; the aligned Eurostat denominator has a mean absolute deviation of 0.08%. The alignment assumes that the age composition of the whole eligible cohort matches the weighted composition of the tested sample. That assumption cannot be evaluated from the PISA microdata and aggregate population series used here; population-level administrative or register data covering the full eligible cohort would be required to test it directly. The alignment also cannot be built where birth month or year is withheld from the released files: Iceland, Montenegro, Norway and Sweden release neither in at least one cycle.

Four further mismatch classes are documented rather than removed. PISA POP15 has no reported population concept, reference date, source or vintage, so no processing of the external series can align those dimensions. Eurostat's own reference population differs across member states. Eurostat values from 2021 onwards embed the 2021–2022 census round, whereas PISA's 2018 POP15 was supplied contemporaneously. WPP is a reconciled model-based estimate rather than an administrative count.

Table 5 recomputes the contrast on this basis. A reduced-mismatch set of 59 systems retains those that are whole national or economy units with a one-to-one WPP match, have an observable age window in both cycles, carry no OECD-documented POP15 re-sourcing, and reproduce the published CI3 identity; 58 of them have complete two-cycle trend estimates. A strict set additionally requires POP15 to be internally consistent with its own numerator, that is $\mathrm{CI3} \leq 1$ and enrolled 15-year-olds no greater than POP15, leaving 51. The mean absolute contrast is 3.4342, 3.4526 and 3.4290 points in the original specification and 3.2933, 3.5462 and 3.4395 points in the reduced-mismatch set under alignment. It falls by about 0.14 points in mathematics and rises by about 0.09 and 0.01 points in reading and science. Counts above the link-error reference value remain 33, 45 and 43 of 58. Sign disagreements remain 4, 6 and 2 of 58, of which 0, 1 and 0 have a statistically significant PISA-denominator change. The reduction in avoidable mismatch does not remove the sensitivity.

Table 5 *Absolute PISA–alternative contrast after reducing avoidable mismatch*

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

Note. Score points. The first block is mathematics. "WPP aligned" blends WPP single-age counts using each economy's observed PISA age-eligibility composition at 1 July of the cycle year. The reduced-mismatch and strict sets are defined in the text and enumerated system by system in Online Resource 1. Link-error reference values are 2.24, 1.47 and 1.61. Alignment cannot be built for Iceland, Montenegro, Norway and Sweden, which is why the aligned all-systems row has $n = 65$.

The convention in Equation 9 matters far less than the denominator. Recomputing every estimate under the midpoint interpolating rule instead changes adjusted top-quarter levels by a mean 0.004 and at most 0.036 score points, standard errors by a mean 0.003 and at most 0.031 points, and the economy-level PISA–WPP contrast by a mean 0.010 and at most 0.032 points, with a correlation of 1.000 between the two sets of contrasts. Domain means move from 3.4342, 3.4526 and 3.4290 to 3.4347, 3.4507 and 3.4297. Counts above link error, counts above twice link error, sign disagreements and significance calls are identical under both conventions.

The zero-replacement boundary

The documented Annex A7 rule is triggered in 6 of 159 economy–cycle cells. For Korea and Ireland in 2022, the final synthetic weights are negative (approximately $-9{,}984$ and $-1{,}446$), so Variant A sets them to zero. Equation 7 then gives $T = W$ and $q = p$: the coverage adjustment switches off and the reported point estimate becomes the ordinary tested-sample percentile. This occurs in all three domains for both economies.

Four additional cases have positive final synthetic weights but some negative replicate synthetic weights: Germany 2018 in 30 of 80 replicates, Belgium 2022 in 9 of 80, Slovenia 2022 in 2 of 80 and Hong Kong (China) 2018 in 1 of 80. Their point estimates do not change under final-weight zeroing, but their replicate probabilities follow the piecewise rule in Equation 8.

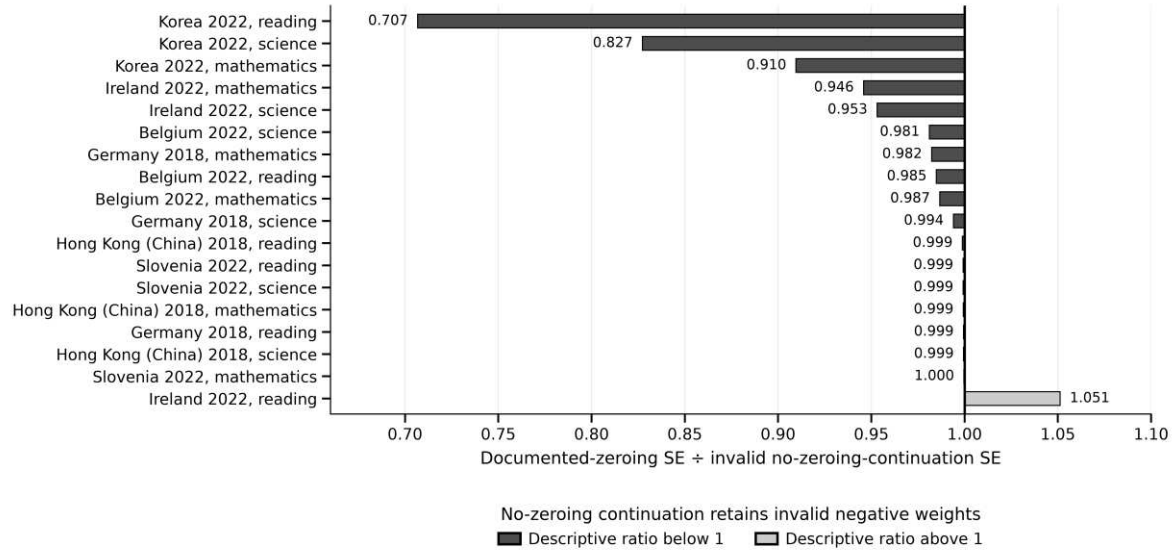


Fig. 4 *Descriptive standard-error ratios for the 18 affected economy–cycle–domain cells. Each ratio is Variant A (documented zero-replacement) divided by Variant B (negative synthetic weights retained). The denominator of every ratio is an explicitly invalid continuation, so values below or above one do not estimate standard-error bias. The figure describes boundary sensitivity only*

Table 6 *Boundary cases under the documented zero-replacement rule*

Economy–cycle	CI3	Replicates zeroed	Mathematics	Reading	Science
Korea 2022	1.024	64/80	0.910	0.707	0.827
Ireland 2022	1.023	80/80	0.946	1.051	0.953
Germany 2018	0.993	30/80	0.982	0.999	0.994
Hong Kong (China) 2018	0.984	1/80	0.999	0.999	0.999
Belgium 2022	0.991	9/80	0.987	0.985	0.981
Slovenia 2022	0.996	2/80	1.000	0.999	0.999

Note. Domain entries are standard-error ratios, Variant A divided by Variant B. Variant B, the denominator of each ratio, is invalid because it retains negative synthetic weights. Accordingly, the 17 of 18 ratios below one cannot be used to infer bias or miscalibration. Korea and Ireland also have final-weight zeroing; the other four cases are replicate-only.

The boundary is not fully visible from the published final CI3. Germany, Hong Kong (China), Belgium and Slovenia all have CI3 below one, yet at least one replicate invokes zero-replacement. The result is a reproducible discontinuity in the documented algorithm. The descriptive A/B ratios in Fig. 4 and Table 6 characterise that discontinuity but do not provide a valid counterfactual standard error.

Complete-sample summaries, sign sensitivity and closure checks

Unweighted complete-sample means across the same 69 economies are reported here to avoid implying membership weighting or an institutional aggregate. For mathematics, the tested-mean, PISA-adjusted and WPP-adjusted changes are -11.442, -11.771 and -12.960. For reading they are -9.612, -6.965 and -8.457; for science they are -1.870, $+1.458$ and $+0.100$. The mathematics decline is similar across these summaries, whereas the reading and science summaries are more scenario-sensitive.

The sign of the coverage-invariant change differs between PISA and WPP in 4 of 69 mathematics cases, 5 of 69 reading cases and 4 of 69 science cases. Among those sign changes, the number whose PISA-denominator estimate is statistically significant at the conventional two-sided 5% threshold is 0, 0 and 1, respectively. The main sensitivity is therefore often in magnitude rather than in a well-supported reversal.

Across the five closure-duration measures, the Pearson correlations with ΔCI3 range approximately from -0.14 to $+0.22$ in samples of 67–70. A nominal Pearson result for one specification is not reproduced by its rank correlation, and pre-pandemic placebo associations are at least as large. We therefore find no robust association in these checks. This result does not establish that closure-duration analyses are free of confounding; it only fails to support CI3 change as a stable measured correlate in the available country-level specifications.

We do not report a Lee-type common-coverage trimming bound. The exact decomposition and denominator contrasts show that coverage change cannot generally be interpreted as monotone entry or exit of students at one end of the achievement distribution, which is the counterfactual needed for that bound.

The Philippines

The Philippines has the largest survey-channel increase among sovereign participants in the decomposition: approximately $+16.8$ points of a $+15.4$-point CI3 increase, offset by about -1.4 points in the enrolment channel. Table 7 contrasts the tested means with the PISA-denominator top-quarter estimates. The final column is the absolute difference between the PISA- and WPP-scenario changes.

***Table** 7 Philippines: tested-sample means and PISA-denominator coverage-invariant statistics*

	Tested-sample mean				**PISA-adjusted top-quarter statistic**			
Domain	2018	2022	Change	SE	2018	2022	Change	Abs. contrast
Mathematics	352.57	354.72	+2.15	4.87	375.56	384.29	+8.73	1.43
Reading	339.69	346.55	+6.86	4.95	355.51	386.74	+31.23	0.30
Science	356.93	356.17	−0.77	4.73	371.99	388.69	+16.70	0.11

Note. Score points. Tested-mean SEs are for the 2018–2022 change and include link error. The final column is the absolute PISA–WPP contrast in the adjusted change; it is not a confidence interval.

The tested-distribution quantile used by the adjustment rises from about 0.63 in 2018 to 0.70 in 2022 as PISA CI3 rises from approximately 0.679 to 0.833. Conditional on the identifying assumption in Equation 6, the adjusted changes are larger than the tested-mean changes. Two cautions are central. First, the 2018 calculation assumes that none of the roughly 32% not covered would reach the top quarter of the age group. Second, the 2022 weighted participant count (1,782,896) exceeds the reported post-school-exclusion target-population count (1,709,495). The available documentation here does not establish whether this discrepancy reflects differing definitions, revisions, calibration or other weighting adjustments.

Discussion

The results identify two distinct sensitivity problems. The first is continuous: selecting a different named population series changes the quantile applied to each tested distribution and can alter the estimated trend by several score points. The second is a boundary discontinuity: when a final or replicate weight sum exceeds POP15, zero-replacement changes the mapping from population percentile to tested-sample percentile. These problems share an input but should not be conflated. The PISA–WPP contrast describes dependence on denominator choice; the A/B comparison describes dependence on a documented computational patch whose comparison variant is not substantively valid.

The denominator sensitivity is consequential relative to link error in a majority of complete cases in all three domains, but that comparison is not a probability statement. Link error and a named-scenario contrast arise from different constructions. The latter should not simply be added in quadrature to sampling and link components because no distribution for denominator error has been specified.

The sensitivity is not an artefact of avoidable harmonisation failures. Aligning the external series to each economy's observed PISA age-eligibility window and restricting to systems without documented territorial, vintage or internal-consistency problems leaves mean absolute contrasts of 3.29, 3.55 and 3.44 points, against 3.43, 3.45 and 3.43 in the original specification. What remains after that reduction is not measurement error either. It is the residual disagreement between a national submission whose population concept, reference date and vintage are unpublished and an external series whose concept, date and method are published. This study contains no common cross-national criterion that establishes which count is closer to PISA's eligible cohort, so the quantity identified here is dependence on denominator choice rather than denominator error. That limit is specific to a 69-economy design. Country-specific population registers can adjudicate individual systems: Borger et al. (2024) compare the Swedish PISA sample with Statistics Sweden register data covering the entire relevant birth cohort. Evidence of that kind is not available on comparable terms across the present sample, which is why the estimand here is a scenario contrast rather than an error.

The three-series comparison sharpens but does not settle attribution. WPP and Eurostat are closer to each other than either is to PISA on average in the 38-country subset, and both place the four European boundary cases below the PISA CI3 at the binding cycle. Yet the three series agree on the sign of CI3 change in fewer than half of those countries, and PISA is not uniquely the dissenting series in every disagreement. National register documentation and harmonised vintage information would be required for case-specific adjudication.

Substantively, the unweighted mathematics result is stable across the tested mean and the two adjustment scenarios. Reading loses part of its measured decline under either coverage-adjusted scenario, while science moves from a small negative tested-mean change to a value close to zero or modestly positive. These are complete-sample descriptive summaries, not institutional averages and not causal estimates.

Limitations

First, no population series is treated as ground truth here. WPP is model-based, Eurostat harmonises national submissions, and PISA POP15 can reflect national concepts or revisions. Scenario contrasts cannot determine which series is closest to the eligible population in a given country, and this design supplies no common cross-national criterion that could.

Country-specific registers can support such a judgement in individual systems (Borger et al., 2024), but not on comparable terms across 69 economies.

Second, the coverage-invariant statistic depends on the untestable assumption that every non-covered student lies below the target threshold. It describes the upper part of the age-group distribution and cannot recover the lower three quarters. Floor effects may remain relevant even though the selected tested-sample quantiles are above the median in many cases (Rutkowski et al., 2019).

Third, the weighted-percentile convention is identified from OECD's published percentiles rather than from a documented rule. Equation 9 reproduces the published mathematics percentiles to a mean absolute deviation of 0.0004 score points, and matches the per-plausible-value combination that the documented macro prescribes. Agreement on published output is nonetheless strong empirical evidence for the operative convention rather than proof of identity with the exact production implementation used for the Annex A7 statistic. The identification also rests on one published table for one cycle and one domain. The residual implementation sensitivity is bounded by the comparison in Section 4.4, at most 0.036 score points in any adjusted level. Fourth, mode changes and the 2022 hybrid multistage mathematics design are not modelled; these can affect trends and variance components (König et al., 2026; Robitzsch et al., 2020). Scaling-model choices are likewise outside the analysis (Robitzsch, 2022).

Fifth, the 69-economy complete sample supports the descriptive regressions reported here but not elaborate country-level causal models. Sixth, Ukraine's extreme CI3 change reflects wartime disruption and territorial coverage; it is retained transparently rather than interpreted as ordinary survey operations. Seventh, closure-duration correlations are ecological and imprecise and cannot rule out confounding in published outcome models. Finally, Variant B is deliberately invalid and provides only a smooth numerical continuation at the boundary; its standard errors are not a benchmark for correctness.

Conclusion and recommendations

PISA's coverage-adjusted trend statistic is mechanically sensitive to the population denominator that defines coverage. Across the two named primary scenarios, that sensitivity is commonly comparable to or larger than published link-error reference values. At the algorithmic boundary, the documented zero-replacement rule can switch the point adjustment off or change only a subset of replicate calculations. Neither finding implies that WPP, Eurostat or PISA POP15 is universally preferable.

Four cautious changes would make the statistic easier to interpret.

1. *Publish the POP15 provenance that is already collected.* The technical report shows that the data source and the year of the data are required on Sampling Task 7a and are verified against those sources where possible (OECD, 2025); releasing those two fields per economy in Annex A2, together with the population concept, the reference date and a flag for retrospective re-sourcing, would require no new collection. Without them, an external check can align calendar year, age window and territory but cannot align concept or vintage, so part of any comparison is irreducibly a comparison of an undocumented with a documented construct.
2. *Report a named-scenario sensitivity range.* Every coverage-adjusted trend should be accompanied by a clearly labelled named-scenario denominator-sensitivity range

using at least one external population series. This range should not be described as a confidence interval.

3. *Withhold the statistic when the synthetic weight is negative.* If $P - W < 0$ or $P - W_r < 0$ for any required replicate, the adjusted statistic should not be silently reverted or partially patched. It should be flagged and withheld until the population and weight inputs are reconciled.
4. *State the weighted-percentile definition in the technical documentation.* The published macro delegates it to the host statistical package (OECD, 2009), and no OECD document we located states the rule, so it had to be recovered here by matching candidate definitions to published percentile values. That is avoidable and should not be necessary for an audit of a published statistic.

The analysis does not overturn the broad mathematics decline in the complete sample. It does show that country-level coverage-adjusted magnitudes, and the procedure used to produce them at the boundary, require more transparent inputs and reporting.

Online resources

Online Resource 1 (ESM_1.pdf). *Supplementary methods, provenance and supporting results*. Contains extended methodological details, source provenance, validation material and supporting tables.

Online Resource 2 (ESM_2.zip). *Derived data and estimator*. Contains the machine-readable derived datasets and result tables behind every reported figure and table, the documented Annex A7 estimator module, the figures in vector and raster form, and the provenance manifest recording the source of each input. PISA microdata are not included.

References

Alvarenga Fournier, X. (2026). Educational resilience in Latin America: Divergent post-pandemic trajectories. *Ibero-American Journal of Education & Society Research*, *6*(2), e26009. https://doi.org/10.56183/iberoeds.v6i2.719

Anders, J., Has, S., Jerrim, J., Shure, N., & Zieger, L. (2021). Is Canada really an education superpower? The impact of non-participation on results from PISA 2015. *Educational Assessment, Evaluation and Accountability*, *33*, 229–249. https://doi.org/10.1007/s11092-020-09329-5

Andersson, C., & Sandgren Massih, S. (2023). PISA 2018: Did Sweden exclude students according to the rules? *Assessment in Education: Principles, Policy & Practice*, *30*(1), 33–52. https://doi.org/10.1080/0969594X.2023.2189566

Borger, L., Johansson, S., & Strietholt, R. (2024). How representative is the Swedish PISA sample? A comparison of PISA and register data. *Educational Assessment, Evaluation and Accountability*, *36*, 365–383. https://doi.org/10.1007/s11092-024-09438-5

Boussim, O. (2026). Correcting sample selection bias in PISA rankings. *Economics Letters*, *267*, 113063. https://doi.org/10.1016/j.econlet.2026.113063

Braun, H., Zhang, J., & Vezzu, S. (2008). Evaluating the effectiveness of a full-population estimation method. *ETS Research Report Series*, *2008*(1). https://doi.org/10.1002/j.2333-8504.2008.tb02104.x

Braun, H., Zhang, J., & Vezzu, S. (2010). An investigation of bias in reports of the National Assessment of Educational Progress. *Educational Evaluation and Policy Analysis*, *32*(1), 24–43. https://doi.org/10.3102/0162373709351137

Cosgrove, J., & Cartwright, F. (2014). Changes in achievement on PISA: The case of Ireland and implications for international assessment practice. *Large-Scale Assessments in Education*, *2*, 2. https://doi.org/10.1186/2196-0739-2-2

Eurostat. (2025). *Population on 1 january by age and sex (demo_pjan)*. https://ec.europa.eu/eurostat/databrowser/view/demo_pjan

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., & Tatlow, H. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, *5*, 529–538. https://doi.org/10.1038/s41562-021-01079-8

Hanushek, E. A., & Woessmann, L. (2010). *Sample selectivity and the validity of international student achievement tests in economic research* (NBER Working Paper No. 15867). National Bureau of Economic Research. https://doi.org/10.3386/w15867

Hedges, L. V., & Bandeira de Mello, V. (2013). *A validity study of the NAEP full population estimates*. American Institutes for Research. https://www.air.org/resource/report/validity-study-naep-full-population-estimates

Jakubowski, M., Gajderowicz, T., & Patrinos, H. A. (2025). COVID-19, school closures, and student learning outcomes: New global evidence from PISA. *Npj Science of Learning*, *10*, 5. https://doi.org/10.1038/s41539-025-00297-3

Jerrim, J. (2021). PISA 2018 in England, Northern Ireland, Scotland and Wales: Is the data really representative of all four corners of the UK? *Review of Education*, *9*(3), e3270. https://doi.org/10.1002/rev3.3270

Kennedy, A. I., & Strietholt, R. (2023). School closure policies and student reading achievement: Evidence across countries. *Educational Assessment, Evaluation and Accountability*, *35*, 475–501. https://doi.org/10.1007/s11092-023-09415-4

König, C., Fink, A., & Frey, A. (2026). Effects of the "highly adaptive testing design for PISA" on plausible value-based population estimates. *Large-Scale Assessments in Education*, *14*(1), 36. https://doi.org/10.1186/s40536-026-00309-7

Ludewig, U., Strietholt, R., & McElvany, N. (2025). Reading literacy decline in Europe: Disentangling school closures and out-of-school learning conditions during the COVID-19 pandemic. *Learning and Instruction*, *98*, 102150. https://doi.org/10.1016/j.learninstruc.2025.102150

OECD. (2009). *PISA data analysis manual: SPSS, second edition*. OECD Publishing. https://doi.org/10.1787/9789264056275-en

OECD. (2017). *Does the quality of learning outcomes fall when education expands to include more disadvantaged students?* (PISA in Focus No. 75). OECD Publishing. https://doi.org/10.1787/06c8a756-en

OECD. (2019). *PISA 2018 results (volume i): What students know and can do*. OECD Publishing. https://doi.org/10.1787/5f07c754-en

OECD. (2023). *PISA 2022 results (volume i): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

OECD. (2025). *PISA 2022 technical report*. OECD Publishing. https://doi.org/10.1787/01820d6d-en

Robitzsch, A. (2022). Exploring the multiverse of analytical decisions in scaling educational large-scale assessment data: A specification curve analysis for PISA 2018 mathematics data. *European Journal of Investigation in Health, Psychology and Education*, *12*(7), 731–753. https://doi.org/10.3390/ejihpe12070054

Robitzsch, A., & Lüdtke, O. (2019). Linking errors in international large-scale assessments: Calculation of standard errors for trend estimation. *Assessment in Education: Principles, Policy & Practice*, *26*(4), 444–465. https://doi.org/10.1080/0969594X.2018.1433633

Robitzsch, A., & Lüdtke, O. (2024). An examination of the linking error currently used in PISA. *Measurement: Interdisciplinary Research and Perspectives*, *22*(1), 61–77. https://doi.org/10.1080/15366367.2023.2198915

Robitzsch, A., Lüdtke, O., Goldhammer, F., Kroehne, U., & Köller, O. (2020). Reanalysis of the German PISA data: A comparison of different approaches for trend estimation with a particular emphasis on mode effects. *Frontiers in Psychology*, *11*, 884. https://doi.org/10.3389/fpsyg.2020.00884

Rutkowski, D., Rutkowski, L., & Liaw, Y.-L. (2019). The existence and impact of floor effects for low-performing PISA participants. *Assessment in Education: Principles, Policy & Practice*, *26*(6), 643–664. https://doi.org/10.1080/0969594X.2019.1577219

Sachse, K. A., Mahler, N., & Pohl, S. (2019). When nonresponse mechanisms change: Effects on trends and group comparisons in international large-scale assessments. *Educational and Psychological Measurement*, *79*(4), 699–726. https://doi.org/10.1177/0013164419829196

Spaull, N. (2019). Who makes it into PISA? Understanding the impact of PISA sample eligibility using Turkey as a case study (PISA 2003–PISA 2012). *Assessment in Education: Principles, Policy & Practice*, *26*(4), 397–421. https://doi.org/10.1080/0969594X.2018.1504742

UNESCO Institute for Statistics. (2022). *Duration of school closures by country*. https://covid19.uis.unesco.org/

United Nations, Department of Economic and Social Affairs, Population Division. (2024). *World population prospects 2024*. https://population.un.org/wpp/

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.pdf
- ESM2.zip