Interpreting the PISA score declines

Evidence and cautions for ten proposed explanations

Dr. Milos Kankaras | 16 September 2026

This analytical brief examines ten proposed explanations for the decline in average PISA scores between 2018, 2022 and 2025. It draws on common-item, response-time, measurement-model and population comparisons, TIMSS pupil panels, and other international and national achievement assessments.

The comparisons offer reasons for caution about simple, single-factor accounts. They do not establish that any proposed factor is irrelevant, nor identify how much each contributed to the decline. A weak association, a sensitivity to measurement choices or a residual decline after adjustment is evidence about a particular comparison; none alone settles the underlying causal question.

Substantial reading and mathematics declines remain in several specifications, with differences across countries, subjects, tasks and populations. The purpose is to examine the evidence and limits of competing interpretations before drawing conclusions about causes or remedies.

Ten proposed explanations

1. Social media use
2. Screen time and school phone restrictions
3. AI use and the socioeconomic pattern
4. Curriculum reform
5. Pandemic disruption and school closures
6. Migration and changing pupil composition
7. Rapid responding and test effort
8. Changes in assessment methods
9. Attention and reading comprehension
10. Disengagement and pupil profiles

Selection and scope

The ten topics reflect prominence in the retained traditional and social media discussion, relevance to the decline, policy consequences and the availability of informative evidence. They are not a ranking of online frequency. Each section states an interpretation to examine, rather than attributing that exact wording to a commentator. Tentative suggestions remain distinguished from strong causal claims. Sources include a 52-item media collection and a 46-programme/component cross-assessment review.

1 Social media use

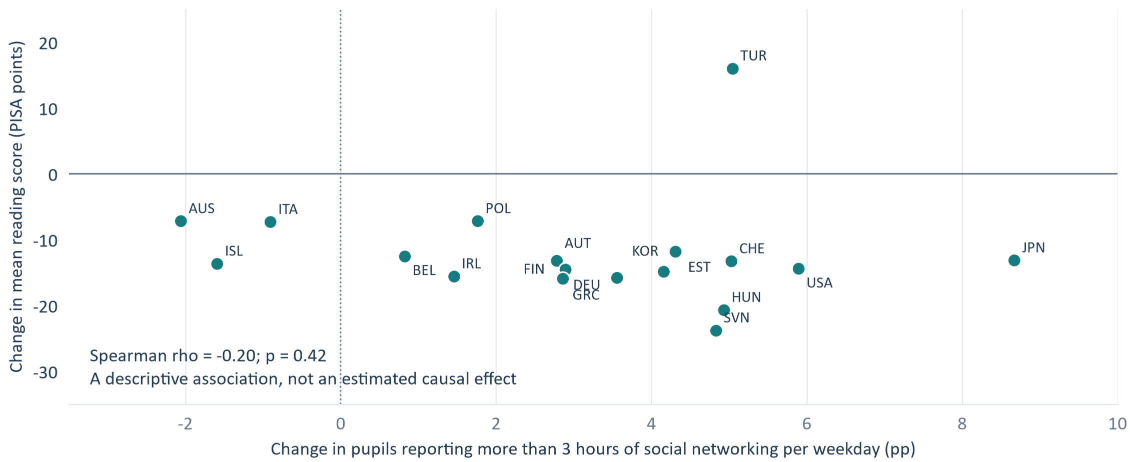
Interpretation to examine: Increasing social-media use is the principal cause of the recent reading decline.

Discourse: social-media and screen explanations recur in LinkedIn, blogs and the r/Professors discussion [M1, M4]. Some posts suggest a contribution; this section examines the stronger causal interpretation.

Caution: the retained country-level association is weak and cannot identify social media as the main cause, or establish that its causal effect is absent.

Changes in social-media use and reading scores

PISA 2022 to 2025 | all 18 countries with at least 50% questionnaire coverage in each round



Country averages are equally weighted; pupil weights are used within each country. Self-reported exposure.
A weak country correlation does not rule out individual harm, delayed effects or confounding. Source: Step 4.

Across all 18 countries meeting the retained questionnaire-coverage rule, growth in the share reporting more than three hours of social networking on a typical weekday has a rank correlation of -0.20 with reading-score change (p = .42). Reading falls also occur where that high-use share decreased. This is more informative than juxtaposing one global screen trend with one global score trend.

It is still a country-level comparison of self-reports, not an experiment. Individual associations can differ from country associations. School-phone interventions in the next figure address a narrower exposure, and cannot be relabelled as experiments on Instagram or TikTok. No analysis in the completed package identifies the fraction of the international decline caused by social media.

Evidence limit: Eighteen countries, exposure misclassification, questionnaire selection, delayed effects and confounding limit the test. A nonsignificant correlation is not evidence that the individual causal effect is zero.

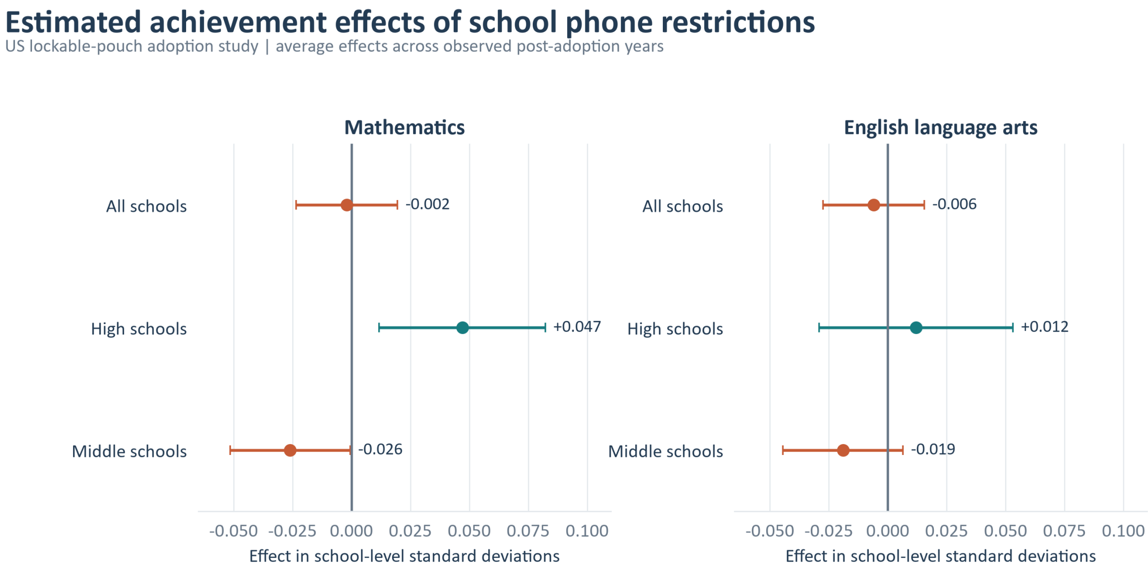
Sources: L1 Step 4, digital_changes.csv, digital_summary.csv and country_changes.csv; R2-R4 for narrower intervention evidence.

2 Screen time and school phone restrictions

Interpretation to examine: General screen exposure drives the decline, and restricting school phone access should restore achievement.

Discourse: broad device-use explanations and school-ban arguments appear in the collected social-media and blog accounts [M1-M4]. The interpretation connects a proposed cause to a policy response.

Caution: estimated policy effects differ by setting and school level. They do not identify a general screen-time effect or guarantee recovery in PISA scores.



95% intervals calculated as estimate +/- 1.96 x published school-clustered SE. These are not pupil SDs.
This policy reduces school phone access; it does not remove all screen time. Source: Allcott et al., Table 3.

The nationwide US study of lockable pouches found pooled effects close to zero, despite reduced phone use. In its August 2026 specification, mathematics effects were +0.047 school-level SD in high schools and -0.026 in middle schools; the pooled estimate was -0.002. English-language-arts estimates were small and their intervals crossed zero. These are school-level, not pupil-level, standard deviations.

The broader evidence is not uniformly null: studies in Florida and Rio de Janeiro report achievement gains from restrictions [R3-R4]. Policies, implementation and populations differ. This variation cautions against expecting a uniform effect or guaranteed recovery. It also gives no basis for assuming that phone restrictions cannot help.

Evidence limit: The plotted study is a quasi-experimental working paper, not random assignment. Restricting phones at school does not eliminate home use, computers, educational applications or other distractions; effects may change over time.

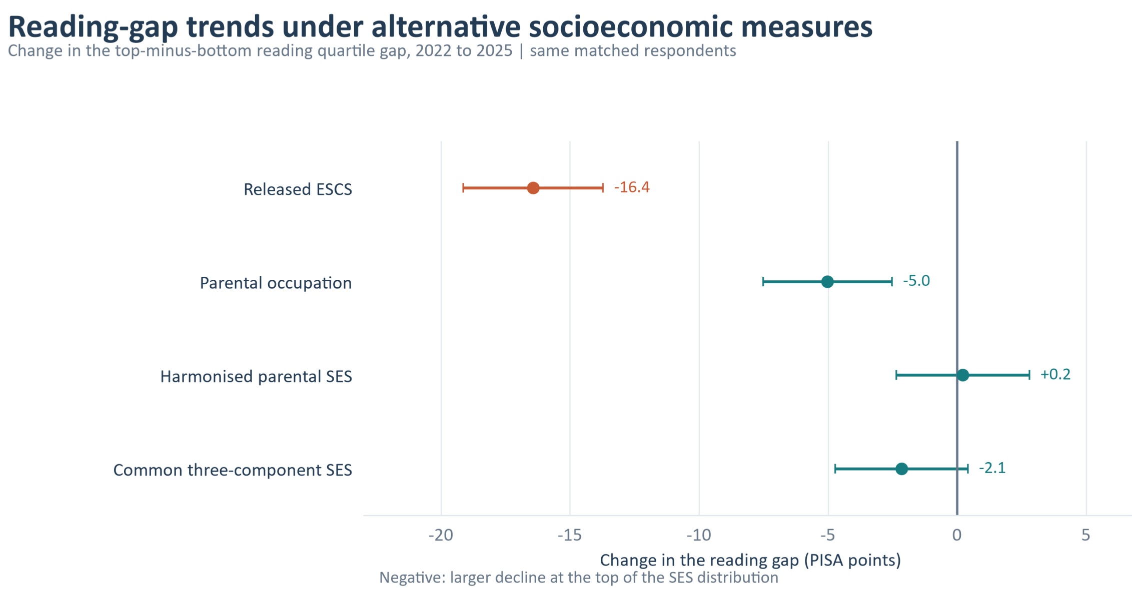
Sources: R2 Allcott et al., August 2026, Table 3, p. 27; L2 retained PhoneBans2026.pdf. R3 Figlio and Özek; R4 Lichand et al.

3 AI use and the socioeconomic pattern

Interpretation to examine: The larger decline among advantaged pupils points to generative AI as a leading cause.

Discourse: Gortazar/EsadeEcPol presents AI as a possible explanation of the socioeconomic pattern, not a proven result [M2]. Ilekytė and Eduentry also connect AI use with falling performance [M1, M3].

Caution: the socioeconomic pattern changes with the SES definition. It does not identify an AI effect, although harm is possible in particular learning settings.



Alternative measures change the construct and the quartile membership; none is a validated gold standard.
Conditional 95% intervals exclude linking error. This is a sensitivity analysis, not an AI effect estimate. Source: Step 4.

On the same matched respondents, the released-ESCS top-bottom reading gap narrows by 16.4 points from 2022 to 2025. With harmonised parental occupation and education, it changes by only +0.2 points, with an interval spanning zero. Occupation alone and a common three-component measure give intermediate results. The striking differential decline is therefore not invariant to the SES definition.

This does not show that AI cannot harm learning. Bastani et al.'s randomised mathematics study found a 17% lower unaided exam performance after unrestricted GPT access, while a tutor designed to safeguard learning mitigated the harm [R5]. That establishes a possible mechanism in a particular setting, not its prevalence or contribution to PISA. Earlier PISA losses also occurred before widespread generative-AI use.

Evidence limit: Alternative SES measures change both the construct and quartile membership. None is a gold standard. A 2025 AI-use association lacks prior proficiency and comparable earlier exposure histories, so reverse causation and confounding remain.

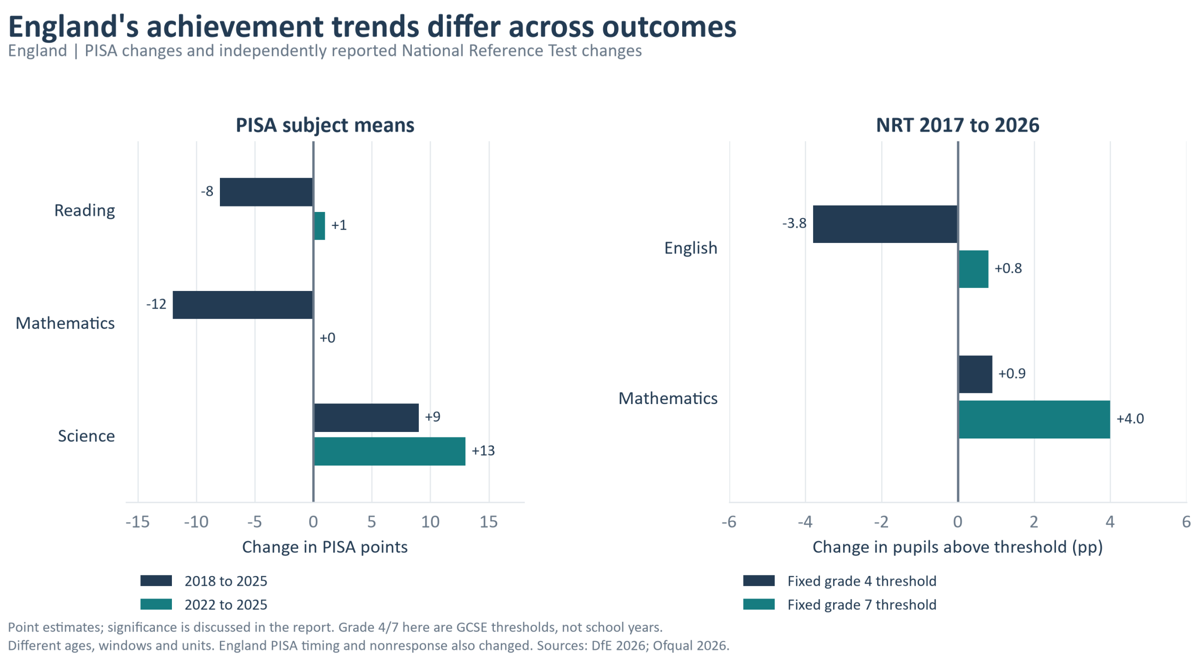
Sources: L1 Step 4, equity_changes.csv, matched sample; R5 Bastani et al. (2025); M2 EsadeEcPol.

4 Curriculum reform

Interpretation to examine: Differences in national trends reveal which curriculum approaches caused or prevented the decline.

Discourse: curriculum and foundational-knowledge explanations occur in Burn-Murdoch's X reply, the Scottish discussion and the German teachers' association commentary [M5-M7].

Caution: selected national trends differ across subjects and thresholds. They do not identify the effect of a curriculum reform.



England's PISA means give different accounts by subject and baseline. Mathematics is unchanged from 2022 but 12 points below 2018, a significant longer-period fall. Science is 13 points above 2022; reading is numerically 8 points below 2018 but that difference is not statistically significant in the national report. A single success-or-failure label obscures these distinctions.

England's independent National Reference Test is mixed too. Relative to 2017, its 2026 lower English threshold fell 3.8 percentage points, while the higher mathematics threshold rose 4.0 points; both changes were significant. The other two displayed changes were not. These results leave beneficial curriculum effects possible, but a selected trend cannot by itself demonstrate them.

Evidence limit: A causal test needs defined reform exposure, implementation timing, comparable untreated cohorts and earlier trends. England's PISA 2025 assessment ran in February-May with a changed year-group mix and nonresponse concerns. NRT has a different age, period and outcome scale.

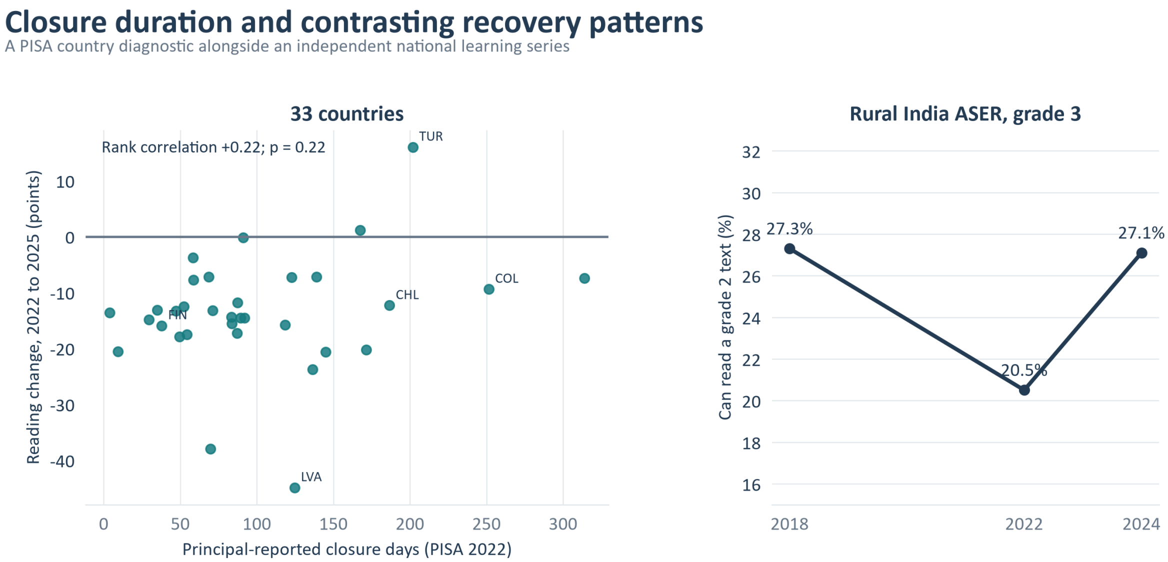
Sources: R6 England national report, Figures 7.5-7.7 and sampling sections; R7 Ofqual NRT 2026; L2 national_changes.csv.

5 Pandemic disruption and school closures

Interpretation to examine: The duration of school closures explains the country pattern and continued decline after 2022.

Discourse: pandemic disruption and absent recovery feature in the collected discussion [M4, M8]. This section considers how well closure duration alone describes the country pattern.

Caution: closure days track the country pattern weakly. This does not rule out pandemic learning losses or their persistence.



Country correlation is not a causal closure estimate; exposure quality and cohort timing are omitted.
ASER covers different pupils and ages from PISA. Recovery in one series does not rule out persistent losses elsewhere.

The 33-country correlation between principal-reported closure days and reading change is +0.08 for 2018-2022 and +0.22 for 2022-2025. Neither is statistically significant; neither has the simple predicted negative sign. This is a poor basis for assigning each country's decline to its closure count.

Independent evidence supports both learning disruption and heterogeneous recovery. A Dutch pupil-panel study estimated a closure-period learning shortfall of about 0.08 SD [R8]. In rural India's ASER, the share of grade-3 pupils able to read a grade-2 text fell from 27.3% in 2018 to 20.5% in 2022 and returned to 27.1% in 2024. Continuing decline is not a universal post-pandemic trajectory.

Evidence limit: Principal reports omit closure timing, remote-teaching quality and household disruption. The PISA cohorts were exposed at different developmental stages. ASER is a different population and age group, so its recovery cannot establish recovery among PISA pupils.

Sources: L1 Step 4, pandemic_closure_days.csv and pandemic_correlations.csv; R8 Engzell et al.; R9 ASER 2024; L2 national_series.csv.

6 Migration and changing pupil composition

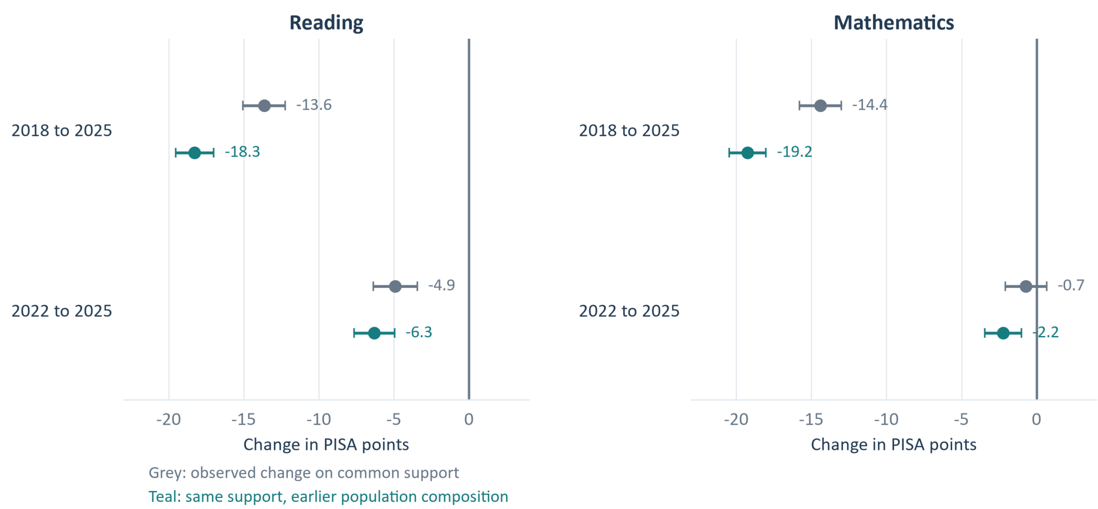
Interpretation to examine: A larger share of migrant or disadvantaged pupils accounts for the aggregate decline.

Discourse: migration and social-composition explanations were considered in the online discussion and in our earlier investigations; Gortazar explicitly argues against a dominant migration explanation for Spain [M2].

Caution: declines remain after holding measured composition constant in the supported sample. The comparison does not isolate an immigration effect or cover all pupils.

Reading declines after measured composition is held constant

29 matched countries | joint parental SES, age, grade and immigrant-status standardisation



Supported respondents differ from the full national samples. Complete cases and common cells only.
95% intervals include survey/PV uncertainty, not linking error. This is not an isolated migration effect. Source: Step 5.

The newly completed standardisation fixes joint parental SES, age, grade and immigrant-status composition. In 29 countries, the common-support reading fall is 13.6 points from 2018 to 2025; applying the earlier composition makes it 18.3 points. For 2022-2025, the corresponding falls are 4.9 and 6.3 points. Mathematics also remains lower after adjustment.

Measured composition thus partly offsets the fall in these selected populations. This does not mean the entire national trend has been corrected. In the same 29 countries, the all-student reading falls are 23.7 and 13.9 points. Their differences from the common-support estimates reflect respondent selection and must not be mislabelled as the migration contribution.

Evidence limit: The adjustment is joint, not a separate causal migration effect. Common cells and complete covariates exclude pupils; unmeasured composition and nonresponse remain. Country-specific migration effects can differ from the aggregate result.

Sources: L3 Step 5, aggregate_composition.csv, migration_added specification; composition_method.json and support audits.

7 Rapid responding and test effort

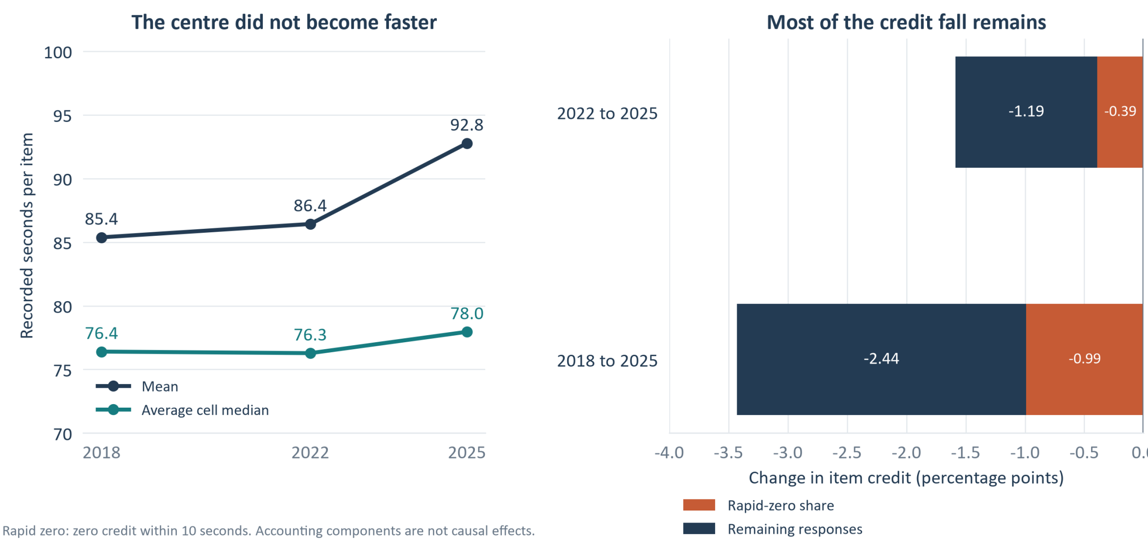
Interpretation to examine: Increased rushing through the test is the principal explanation for the reading decline.

Discourse: the hasty-reader and attention accounts in Ilekytė, Gortazar/EsadeEcPol and Eduentry [M1-M3]. Their wording and degrees of certainty differ.

Caution: most of the analysed common-item credit decline remains outside the rapid-zero-share component. This accounting does not identify the causal contribution of rushing.

Recorded response times and the common-item credit decline

19 identical non-adaptive reading items in 35 countries



Rapid zero: zero credit within 10 seconds. Accounting components are not causal effects.
Medians average country-item medians. Clock equivalence and missing timing remain limitations. Source: Step 3.

Across the same 19 non-adaptive reading items, recorded mean time rose from 85.4 seconds in 2018 to 92.8 in 2025; the average country-item median rose from 76.4 to 78.0. Rapid errors increased too. A larger fast-error tail can coexist with a slower centre.

For 2022-2025, the timed-credit fall was 1.59 percentage points. The fixed 10-second accounting attributes 0.39 points mechanically to the changing rapid-zero share and 1.19 to remaining responses. For 2018-2025, the corresponding components are 0.99 and 2.44 points. These are accounting terms, not causal effects or PISA-score corrections. Removing a selected class of responses does not recover a counterfactual level of proficiency.

Evidence limit: The 19-item panel is smaller than the whole assessment, and clock equivalence is unresolved. Slower time can itself reflect distraction or difficulty. The supporting package distinguishes official pupil-level and response-level hasty-reading indicators; neither is identical to the rapid-zero rule used here.

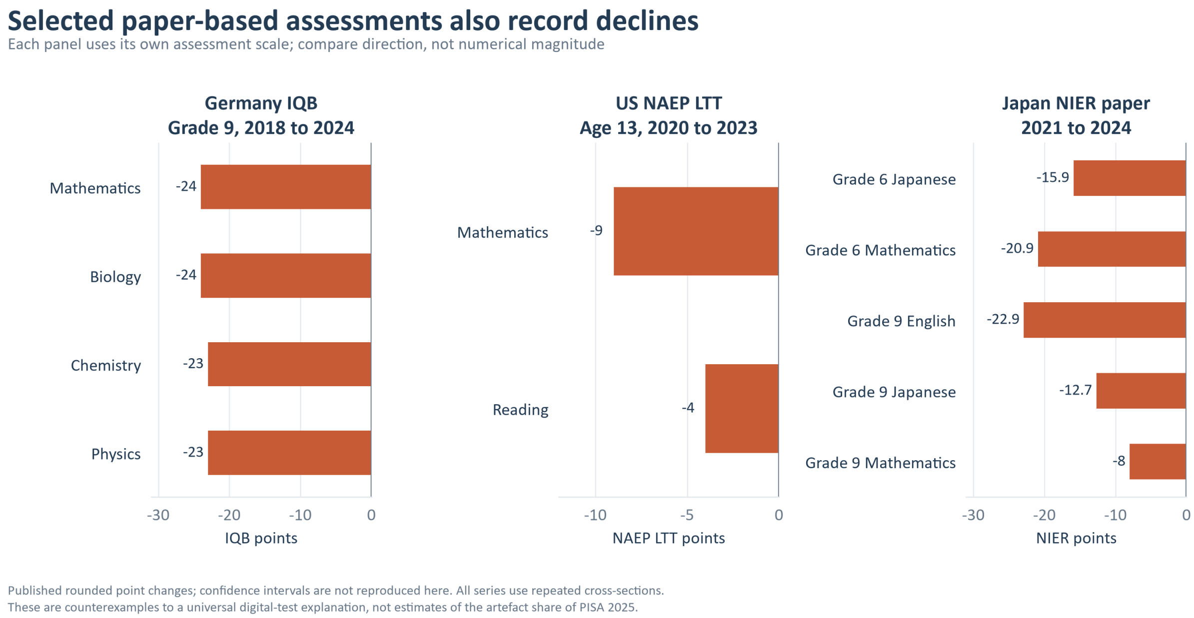
Sources: L1 Step 3, panel_levels.csv and credit_accounting.csv; R1 OECD Annex A1.

8 Changes in assessment methods

Interpretation to examine: Changes in PISA delivery, provider or scaling account for the observed decline.

Assessment comparability is a competing explanation considered in the preceding analyses. This section examines what evidence from other tests can establish about PISA-specific methodological changes.

Caution: declines in other assessments warrant attention, but cannot quantify or exclude methodological contributions to PISA trends.



Germany's paper-based IQB recorded 2018-2024 falls of approximately 23-24 points across mathematics and three science subjects. US paper-based NAEP long-term-trend scores fell 9 points in mathematics and 4 in reading at age 13 from 2020 to 2023. Japan's comparable paper trend sample also declined between 2021 and 2024. These observations concern independent assessment programmes, so a PISA platform change cannot account for them. Their causes and relation to PISA still require investigation.

Within PISA, common-item analyses and the completed anchor-sensitivity models also retain a reading and mathematics decline. Their robustness is limited to the tested models. The paper-test evidence establishes corroboration outside PISA, not the fraction of PISA's recent fall that is substantive rather than methodological.

Evidence limit: The external tests have different ages, periods and scales. PISA-specific changes in language, routing, scoring, coverage and platform remain possible contributors. A direct old/new platform bridge is still needed; stability under selected anchors is not proof of full invariance.

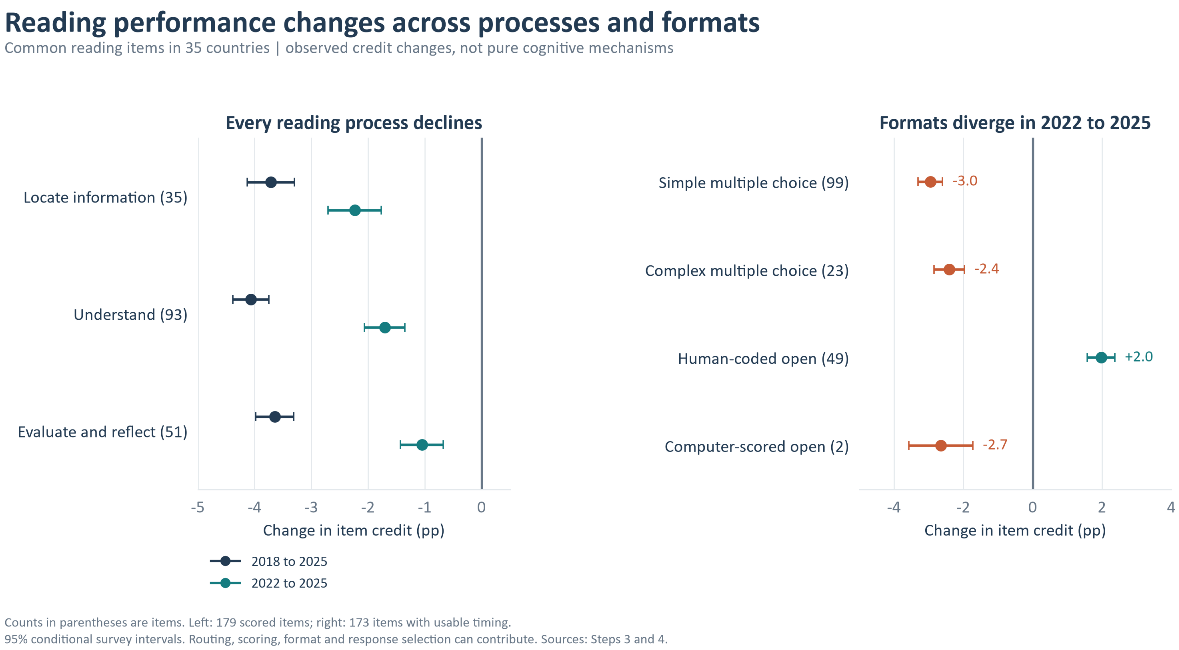
Sources: R10 IQB Bildungstrend 2024; R11 NCES NAEP LTT 2023; R12 NIER 2024; L1 Steps 1-2; L2 cross-assessment review.

9 Attention and reading comprehension

Interpretation to examine: The reading decline reflects reduced attention while comprehension capacity remains unchanged.

Discourse: the attention-versus-comprehension distinction is central to Gortazar's LinkedIn/X and EsadeEcPol accounts [M2]. This is more specific than saying attention may contribute.

Caution: comprehension-task performance also declines, but those tasks require attention. The results cannot separate changes in the two capacities.



Common-item accuracy falls in every reading-process group. The 93 understanding items lose 4.07 percentage points from 2018 to 2025 and 1.71 from 2022 to 2025; locating information and evaluating/reflecting also decline. Thus, observed performance on comprehension-classified tasks is also lower. This does not establish a change in comprehension capacity independent of attention.

At the same time, formats diverge: in the timed-item panel, simple multiple-choice credit falls 2.95 points from 2022 to 2025, while human-coded open responses gain 1.98. That is a reason to investigate scoring, task selection and interface demands as well as attention. A format is not a pure measure of a cognitive mechanism. These results neither isolate attention loss nor establish a general decline in intelligence.

Evidence limit: Comprehension tasks themselves require attention. Their lower scores do not prove reduced comprehension capacity independent of attention. The process and timed-format panels differ in item eligibility; the two sets of contrasts must not be subtracted as a causal decomposition.

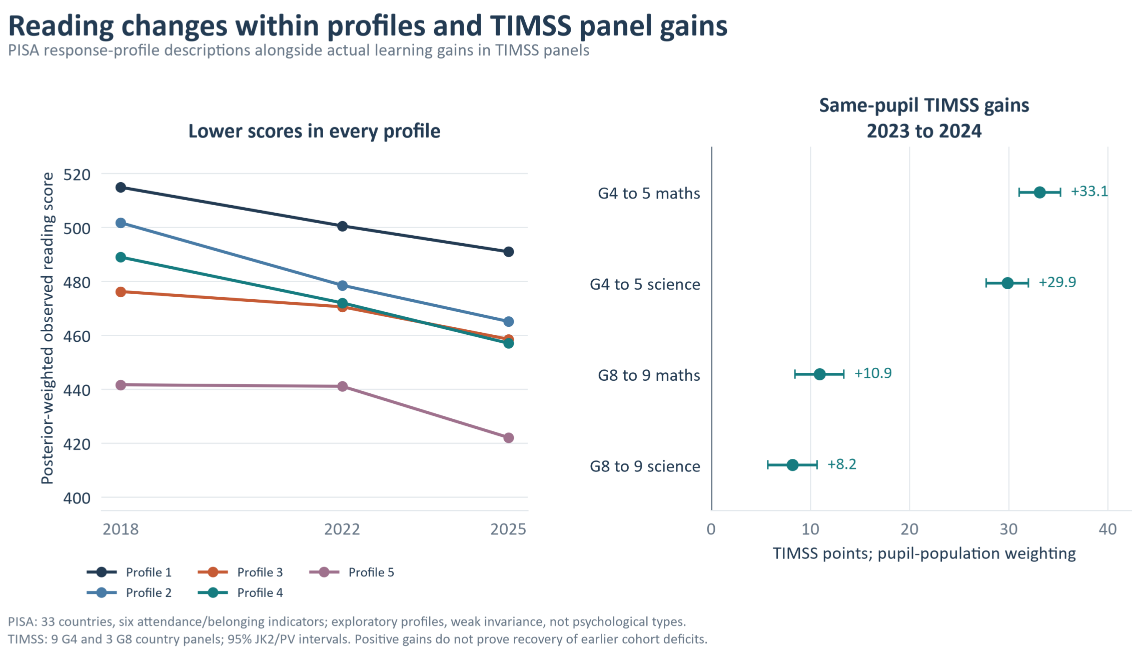
Sources: L1 Step 4, item_group_changes.csv; Step 3, format_changes.csv; M2 EsadeEcPol.

10 Disengagement and pupil profiles

Interpretation to examine: An expanding disengaged minority accounts for the aggregate fall while other pupils remain stable.

Discourse: accounts of withdrawal, weak effort, lost connection and a lost generation appear in the collected discussion [M4, M8]. The minority-composition mechanism is an explicit testable interpretation, not a quotation.

Caution: scores decline within all fitted profiles. The exploratory classes do not isolate the causal role of engagement.



Reading scores decline in all five attendance/belonging profiles, including pupils with regular attendance and strong school connection. In 2022-2025, the 13.2-point respondent-sample fall decomposes into -0.8 points from profile shares and -12.4 within profiles under the common-profile model. Allowing the profiles to vary by cycle changes these to +1.3 and -14.6. Neither specification supports a dominant minority-composition account.

TIMSS supplies a separate check on the claim that pupils have stopped learning. The same-pupil 2023-2024 panels show positive average gains in every country-subject cell. Student-population-weighted grade-4-to-5 gains are 33.1 mathematics and 29.9 science points. Positive learning can coexist with lower achievement than earlier cohorts reached at the same age.

Evidence limit: The five profiles are exploratory and not strongly invariant. Outcomes are posterior-weighted observed scores, not corrected latent-class means. TIMSS is a different age/country panel and does not establish that historical learning deficits have been recovered.

Sources: L1 Step 4, lca_profile_aggregate.csv and profile-accounting files; L4 TIMSS longitudinal national_estimates.csv. Profile 1: regular attendance/connected; 2: often late/connected; 3: skipping/connected; 4: disconnected/little skipping; 5: skipping/disconnected.

What the evidence currently supports

The evidence warrants caution about simple causal accounts. In the analysed samples, the reading decline remains outside the rapid-zero-share accounting term and after measured composition is held constant. Scores also fall within exploratory engagement profiles. Independent paper assessments record losses in other populations and periods. These comparisons constrain particular interpretations; they do not exclude a causal contribution from the proposed factors or allocate the PISA decline among them.

Several substantive mechanisms remain credible. School disruption has evidence from pupil panels with historical comparisons; particular forms of AI assistance can impair unaided performance in experiments; some phone restrictions improve achievement in quasi-experiments. Effects differ by exposure, teaching practice and context. Their existence is distinct from their contribution to the 2018-2025 or 2022-2025 international averages.

The completed TIMSS regressions consistently identify low mathematics confidence as a marker of weaker gains, with books and absence carrying more variable information. Broad contextual models add modest predictive value beyond baseline proficiency, age and sex: median held-out gains in R-squared are about 0.041/0.038 for grade-4 mathematics/science and 0.027/0.030 for grade 8. These are predictions of gains defined by released plausible values, not causal effects or shares of the historical decline explained. [L4-L5]

National circumstances matter. Latvia's 2022-2025 reading fall is about 19.4 points among Latvian-home-language pupils and 97.3 among Russian-home-language pupils, alongside the documented transition to Latvian-only testing and instruction. This is a major comparability issue and candidate mechanism, not an identified 78-point reform effect. Removing Latvia reduces the 35-country recent reading fall by only about 0.9 points. It cannot explain the international decline. [L1, R13]

Reading these findings cautiously

For policy and public commentary, specify the country, subject, period and population before explaining a trend. Separate a descriptive association from an identified effect, report measurement and sampling sensitivities, and consider interacting mechanisms. Evidence that an intervention helps in one setting is not an estimate of its contribution to an international score trend.

Scope and interpretation safeguards

The population and comparison period

PISA compares different samples of 15-year-olds, not the same pupils. The relevant earlier cycle is 2018, not 2019. Our fixed 35-country equal-country reading fall is 24.6 points from 2018 to 2025, versus 17.0 when released pupil-population weights are pooled and 16.1 with fixed 2018 population weights. Those are different estimands, neither a world-population estimate. They do not establish a weighting artefact that makes the decline disappear. Selected questionnaire and common-support samples are labelled separately throughout. [L1]

Using other assessments without treating them as interchangeable

The prior review contains 396 PISA-TIMSS country-pair rows, 270 pairs with other international assessments, 159 national/provincial observations and 46 programme/component records. TIMSS, PIRLS, ICILS, ICCS and PIAAC examine different ages and constructs. National evidence includes PCAP, NAPLAN, IQB, NAEP, NIER, NRT, INVALSI, ASER and others. The ten figures select comparisons that inform the ten interpretations; they are not a representative meta-analysis of programmes. Tests' raw points are never pooled across scales. [L2]

Concordance is partial

Across matched countries, PISA 2022-2025 and TIMSS 2019-2023 mathematics changes correlate 0.46 at grade 4 but only 0.17 at grade 8; science correlations are 0.34 and 0.02. The earlier PISA 2018-2022 mathematics comparison with grade-4 TIMSS gives -0.10. Period and cohort mismatches limit all these comparisons. Canada's PCAP mathematics and reading means are unchanged at displayed precision from 2019 to 2023, while other national tests show falls or partial recovery. Disagreement does not by itself invalidate PISA, and agreement does not identify a common cause. [L2, R14]

Uncertainty and causal interpretation

PISA microdata estimates use survey weights, 80 Fay-BRR replicates and ten plausible values where scores are involved. Displayed project intervals exclude systematic linking error; item/time intervals also exclude clock or classification uncertainty. TIMSS uses five paired plausible values and JK2 replication. No confidence interval is invented for a rounded national series. The phone-study intervals use its published school-clustered standard errors. A nonsignificant estimate is not proof of equivalence; descriptive decomposition is not mediation; cross-country correlation is not an individual treatment effect.

Public supporting evidence

The companion ZIP supplies the ten figures, editable SVGs, aggregate CSVs, selected fuller results and a methods guide. The guide resolves references L1-L5. It is an aggregate audit package, not a complete raw-data reproduction. https://doi.org/10.5281/zenodo.22801644

This is an independently prepared analytical brief, not a peer-reviewed article or an OECD publication. AI tools assisted code, literature collation and drafting; the author is responsible for the final report.

Source register

Media entries document interpretations in the public discussion. Empirical references support the comparisons. The README in this deposit explains the local analysis labels L1-L5.

L1 Supporting aggregate evidence and methods, this deposit, README L1; analyses/step3/results and analyses/step4/results. Includes common-item timing, digital exposure, SES, language and profile tables. Earlier measurement-model results are discussed but not reproduced in full in this aggregate package.

L2 Supporting package, README L2; analyses/cross_assessment/results. Includes evidence_registry.csv (46 programme/component records), selected national trend tables and PISA-TIMSS comparisons. The preceding review contained 68 source entries.

L3 Supporting package, README L3; analyses/step5/results. Includes aggregate_composition.csv, support_audit.csv and composition_method.json; estimates use independent-cycle replication.

L4 Supporting package, README L4-L5; analyses/timss_longitudinal. Includes national_estimates.csv and analysis_plan.json for the 2023-2024 pupil panels. Raw source archives are not redistributed.

L5 Supporting package, README L4-L5; analyses/timss_context_prediction/results. Includes prediction_summary.csv, scale_regressions.csv, country_slope_heterogeneity.csv and predictor_dictionary.csv.

R1 OECD (2026). PISA 2025 Results, Volume I, Annex A1, Tables I.A1.12 and I.A1.15. https://www.oecd.org/en/publications/pisa-2025-results-volume-i_73451bc5-en/full-report/trends-in-reading-scores-over-three-pisa-cycles-2018-2025_3fb6c775.html

R2 Allcott, H., Baron, J., Dee, T., Duckworth, A. L., Gentzkow, M., & Jacob, B. (2026). The Effects of School Phone Bans: National Evidence from Lockable Pouches. August 2026, Table 3. NBER Working Paper 35132. https://www.nber.org/papers/w35132

R3 Figlio, D., & Özek, U. (2025/2026). The Impact of Cellphone Bans in Schools on Student Outcomes: Evidence from Florida. NBER Working Paper 34388. https://www.nber.org/papers/w34388

R4 Lichand, G., Moreno-Louzada, L., da Costa, T., & Gentzkow, M. (2026). The Educational Impacts of School Phone Bans: Evidence from Brazil. NBER Working Paper 35233. https://www.nber.org/papers/w35233

R5 Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. PNAS, 122, e2422633122. https://doi.org/10.1073/pnas.2422633122

R6 Ingram, J., Hillier, J., Stiff, J., Richardson, M., Wilkinson, D., Maris, R., Barnes, I., & Swensson, C. (2026). PISA 2025: National Report for England. DfE. Figures 7.5-7.7 and sampling discussion. https://assets.publishing.service.gov.uk/media/6a9ea5415a0c25165ae469d9/pisa-2025-national-report-for-england.pdf

R7 Ofqual (2026). National Reference Test annual statement 2026. https://www.gov.uk/government/publications/the-national-reference-test-in-2026/national-reference-test-annual-statement-2026

R8 Engzell, P., Frey, A., & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. PNAS, 118, e2022376118. https://doi.org/10.1073/pnas.2022376118

R9 ASER Centre (2025). Annual Status of Education Report Rural 2024. National basic-learning results, retained as source S40 in L2. https://asercentre.org/aser-2024/

Source register continued

R10 IQB (2025). IQB-Bildungstrend 2024. National mathematics and science trends, with scaling/linking documentation. https://www.iqb.hu-berlin.de/media/documents/IQB_Bildungstrend2024_Berichtsband-6.pdf

R11 NCES (2023). NAEP Long-Term Trend Assessment Results, age 13 reading and mathematics. https://www.nationsreportcard.gov/highlights/ltt/2023/

R12 NIER/MEXT (2025). National achievement trend analysis 2024, summary, paper-based trend group. https://www.nier.go.jp/24chousakekkahoukoku/kannren_chousa/pdf/24keinen_summary_point.pdf

R13 OECD (2026). PISA 2025 Results, Volume I, Latvia country note. Test-language and instruction transition. https://www.oecd.org/en/publications/pisa-2025-results-volume-i-country-notes_2d4ff9ea-en/latvia_eea74d60-en.html

R14 CMEC (2025). PCAP 2023 Report on the Pan-Canadian Assessment. Trend tables B1.28, B2.6 and B3.6. https://cmec.ca/Publications/Lists/Publications/Attachments/447/PCAP2023_Public_report_FINAL_EN.pdf

M1 Greta Ilekytė. LinkedIn post on AI, cognitive skills and Baltic reading results; two author-retained screenshots. Exact publication date and permalink were not established. The post says nearly doubled since 2018, not exact doubling in the latest two rounds.

M2 Lucas Gortazar / EsadeEcPol (8 September 2026). PISA 2025 commentary on Spain, attention and candidate explanations; also disseminated on LinkedIn and X. https://www.esade.edu/ecpol/es/blog/pisa-2025-espana-cae-mas-que-su-entorno-la-caida-es-global-y-apunta-al-deterioro-de-la-atencion-del-alumnado/

M3 Eduentry (September 2026). PISA 2025 and work experience, student readiness. https://eduentry.com/blog/pisa-2025-work-experience-student-readiness

M4 r/Professors, selected participants (September 2026). Discussion of the global PISA decline. Individual comments differ in certainty and are not a representative sample of educators. https://www.reddit.com/r/Professors/comments/1wc78wb/pisa_2025_results_show_global_decline_in_math/

M5 John Burn-Murdoch (9 September 2026). X reply on curricula and the English system; retained media entry S05. https://x.com/jburnmurdoch/status/2097598993702297836

M6 r/Scotland, selected participants (September 2026). Discussion of educational reform and PISA decline; retained entry S26. https://www.reddit.com/r/Scotland/comments/1wbizjr/education_reform_under_way_as_scotland_sees/

M7 Cornelia Schwartz and Jochen Ring / Philologenverband Rheinland-Pfalz (September 2026). Commentary connecting PISA decline, foundational knowledge and earlier reforms. Author-retained media entry S45; not redistributed in this aggregate supplement.

M8 Steen William Reves (September 2026). LinkedIn discussion of Danish results, disconnection, COVID and AI; retained entry S09. https://dk.linkedin.com/posts/steenwreves_generationen-som-blev-tabt-pisa-2025-resultaterne-activity-7503368248541814784-7tkU