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.

Socioeconomic inequalities in linked family and classroom reports of AI evaluation

Wenxuan Li

lwx201621@henu.edu.cn

Henan University

Research Article

Keywords: AI literacy, information evaluation, parental self reports, classroom practice, socioeconomic inequality, PISA 2025

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11403418/v1

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

Additional Declarations: No competing interests reported.

Socioeconomic inequalities in linked family and classroom reports of AI evaluation

Abstract

Evaluating information generated by artificial intelligence (AI) is becoming an everyday educational task. Families and classrooms offer different settings for learning this practice, yet their socioeconomic distribution is usually examined separately. This exploratory study analysed linked parent and student reports from the Programme for International Student Assessment 2025 across sixteen education systems. Among 94,262 paired respondents, the equal system mean proportion combining a lower or unfamiliar parental AI evaluation rating with never or almost never practising AI evaluation in lessons was 15.45%. This combination occurred among 24.69% of students in the lowest socioeconomic resource quintile and 8.27% in the highest, a difference of 16.42 percentage points (95% confidence interval 15.45 to 17.39). The contrast was positive in all sixteen systems. In 93,687 records, lower parental ratings were associated with a 3.87 percentage point lower probability of reporting practice at least sometimes after adjustment for school fixed effects, student backgrounds and the full parental rating of ordinary online information (95% confidence interval 3.02 to 4.73 points lower). Estimates changed little with science performance adjustment or response weighting. Allowing arbitrary missing focal responses widened the resource contrast to an identification interval of 0.69–30.40 points, with sampling uncertainty at the lower endpoint including zero. The findings identify a socioeconomic concentration of joint low reports among respondents and position linked family and classroom experiences as a useful focus for equitable AI education.

Keywords: AI literacy; information evaluation; parental self reports; classroom practice; socioeconomic inequality; PISA 2025

Introduction

The educational significance of artificial intelligence (AI) increasingly depends on how young people judge the information it produces. A chatbot can provide a fluent explanation, a summary or a suggested argument, while the learner still has to decide whether its claims are accurate, relevant and adequately supported. AI literacy frameworks therefore place evaluation alongside understanding, application and ethical judgement (Long & Magerko, 2020; Ng et al., 2021). Discussions of large language models in education similarly identify critical thinking, fact checking and informed human oversight as central educational tasks (Kasneci et al., 2023). For adolescents, learning to evaluate AI information connects everyday digital activity with the knowledge practices expected in school.

That learning can take place in several settings. A parent may discuss an AI response during homework, a teacher may ask students to compare a generated explanation with other sources, or a young person may seek help independently. These settings provide different relationships and learning experiences. Treating them together raises an educational equity question: which students report little classroom evaluation practice while their parent also finds the task difficult or unfamiliar? Answering this question requires linked information about families and classrooms, because separate descriptions of either setting leave their overlap unresolved.

Recent AI education research provides a clear reason to examine this intersection. A four-country analysis of early childhood curricula found uneven recognition of AI literacy and argued that dependence on home environments could deepen educational disadvantage (Cohrssen et al., 2026). Although that study concerned curriculum documents for younger children, its question about the relationship between family environments and formal education extends to adolescence. International assessment data offer a complementary perspective. The Programme for International Student Assessment (PISA) 2025 includes parental self ratings and student reports about evaluating AI chatbot information, alongside socioeconomic backgrounds and school identifiers (OECD, 2025a; OECD, 2025b). OECD's initial reporting already identifies a socioeconomic gradient in students' reported classroom AI evaluation opportunities (OECD, 2026). The linked reports make it possible to extend that description to the joint distribution of parental ratings and classroom experiences across education systems.

AI evaluation as an educational practice

The evaluation of information has a substantial research history that predates generative AI. Research on civic online reasoning has developed tasks that assess how students examine claims, sources and evidence (McGrew et al., 2018). A national study of US high school students documented difficulties with evaluating online material and variation across student backgrounds (Breakstone et al., 2021). Studies of skilled evaluators identify practices such as leaving an unfamiliar website to investigate its source and compare external accounts (Wineburg & McGrew, 2019). These studies establish the educational value of making information evaluation an explicit activity, rather than assuming that frequent digital participation supplies the relevant judgement.

AI generated information introduces a related set of demands. Learners need to recognise the difference between a plausible response and an adequately supported account, consider the limits of the system, and decide when additional evidence is needed. Cognitive tools for navigating online information environments provide a useful foundation for these activities (Kozyreva et al., 2020). AI literacy extends that foundation by including knowledge about AI systems and their ethical and social implications (Long & Magerko, 2020; Ng et al., 2021). The focal practice in the present study is consequently one specific component of a wider educational agenda: assessing the quality of information generated by AI chatbots.

Research on AI education also distinguishes learning through an AI application from learning to understand and evaluate AI. Reviews describe a wide range of educational applications, tools and pedagogical approaches (Chiu et al., 2023; Su & Yang, 2022). A recent synthesis of early childhood AI literacy reviews argues for closer attention to developmental needs, social context and educational purposes (Wang et al., 2026). This perspective makes the circumstances in which children encounter AI an important part of the research question. Access to a tool, participation in an AI supported activity and deliberate practice in judging information represent different experiences.

Measurement studies reinforce the need to specify the construct under examination. An AI literacy scale developed for ordinary users distinguishes awareness, use, evaluation and ethics (Wang et al., 2023). The Meta AI Literacy Scale includes several perceived competencies and related psychological dimensions (Carolus et al., 2023). Research on ordinary online skills has also demonstrated differences between self assessed and demonstrated ability (Hargittai & Shafer, 2006). The parental item studied here records a judgement about one evaluation task, while the student item records the frequency of undertaking that task in school lessons. Their combination describes linked reports about an educational practice; it does not substitute for a comprehensive AI literacy assessment.

Family resources and the learning environment

Digital inequality research explains why a specific learning practice may be unevenly distributed even where devices and online services are widely available. The field distinguishes access, skills, patterns of use and the benefits obtained through digital engagement (Robinson et al., 2015; Scheerder et al., 2017). Research on Internet use shows that the amount of time spent online and the purposes of that use follow different social patterns (van Deursen & van Dijk, 2014). Work on the third level digital divide further focuses on the tangible benefits that people obtain from being online (van Deursen & Helsper, 2015). For AI education, these distinctions direct attention to what learners and families do with information, and to the support available for interpreting it.

Young people are particularly poorly described by a simple distinction between digital users and nonusers. Research on children's Internet participation identifies gradations in access, use and opportunities (Livingstone & Helsper, 2007). Evidence on young adults also challenges the assumption that growing up with digital media produces uniformly high levels of skill (Hargittai, 2010). A corresponding fields perspective links digital inclusion to economic, cultural, social and personal resources (Helsper, 2012). Applied to AI evaluation, this approach suggests examining

socioeconomic differences in the conditions surrounding a task, while keeping the measured task distinct from those broader resources.

Families contribute to those conditions through relationships as well as technical knowledge. Research on parental mediation distinguishes conversations, shared activity, restrictions and monitoring (Livingstone & Helsper, 2008). A survey across eight European countries linked enabling and restrictive mediation strategies to parent and child digital skills and to different balances of online opportunities and risks (Livingstone et al., 2017). More recent work separates parental digital skills, perceptions of risk and mediation practices (Helsper et al., 2025). These distinctions matter for the present question because a parent's rating of their own evaluation ability is one feature of the family setting, rather than a direct record of how that parent supports a child.

AI specific studies show how family learning can develop through joint activity. An in-home study designed learning activities for children and parents and identified several parental roles in their shared exploration of AI (Druga et al., 2022). A Hong Kong survey examined associations among parental AI literacy, educational level and parenting self efficacy (Guo et al., 2025). A survey of Chinese parents connected digital capital, AI related beliefs and relational support with intentions concerning AI enabled home learning (Song et al., 2026). These studies provide accounts of family processes and perceptions that a large survey cannot reproduce in the same detail. Linked assessment data can extend this work by examining the distribution of a specific parental rating alongside students' classroom experiences.

Socioeconomic resources may also shape how families engage with formal education. Qualitative research on parental digital involvement in online learning describes different ways in which competence, acceptance of technology and forms of capital are connected (Hou et al., 2024). Studies of technologies intended to connect home and school emphasise purposeful activities and support that responds to parents' circumstances (Lewin & Luckin, 2010). This literature makes joint family and classroom patterns an informative object of study. It also encourages an analysis that retains the distinct meaning of each report, because family involvement and classroom activity are not interchangeable measures.

Classroom practice within a wider learning ecology

A learning ecology perspective examines how young people pursue and connect activities inside and outside formal education (Barron, 2006). Neo-ecological theory extends developmental accounts to include the physical and virtual contexts in which adolescents learn and interact (Navarro & Tudge, 2023). For the present study, these perspectives motivate studying connected settings. A family rating and a classroom report may coincide because learning experiences are related across settings, because students share other circumstances, or because the reports reflect different experiences within a school. Linking the reports is a first step towards specifying those possibilities.

Schools can provide a structured setting for practising evaluation, but the relationship between school provision and family resources is an empirical question. Cross-national research using student assessment data and teacher survey reports found substantial inequalities in students' digital

readiness, alongside limited evidence that school readiness was systematically stratified by student composition (van de Werfhorst et al., 2022). Another analysis of international computer and information literacy data found both unequal access to school digital resources and potentially greater benefits of technical expertise for students from less advantaged backgrounds (Loh et al., 2025). Together, these studies make it important to examine the particular classroom activity and the level at which differences occur.

The instructional context also matters. A systematic review of AI for teachers identifies opportunities in planning, classroom implementation and assessment, together with challenges in applying AI appropriately (Celik et al., 2022). Work on teachers' professional knowledge argues that technological understanding needs to be connected with pedagogical and ethical judgement (Celik, 2023). A classroom study of GenAI supported multimodal writing illustrates how AI literacy and learning engagement can be examined within a concrete activity (Fang et al., 2025). Human centred guidance for generative AI likewise connects educational use with age appropriate design and equitable participation (Miao & Holmes, 2023). These contributions support a focus on the learning practices students report, while leaving open how those practices were organised in particular classrooms.

The present study

This study examines linked parental and student reports across sixteen PISA 2025 education systems. It first describes the full joint distribution of parental AI evaluation ratings and classroom evaluation frequency, including parental task unfamiliarity as a substantive response. It then asks how the combination of lower or unfamiliar parental ratings and never or almost never classroom practice varies across socioeconomic resource quintiles. Finally, it estimates the association between the parental rating and classroom practice among students attending the same school, accounting for measured backgrounds and the parent's rating of evaluating ordinary online information.

The study contributes a joint description, a within school comparison and an assessment of response selection. These components answer related questions about where low reports coincide, how much of their association remains within schools, and how the observed distribution relates to the wider eligible sample. Alternative parental categories and classroom thresholds, science performance adjustment and nonresponse analyses provide further detail.

Methods

Data and analytic samples

The PISA 2025 student public use file contained 755,721 records across 90 country or economy identifiers. Eighteen systems contained valid parental AI evaluation responses. Luxembourg and Macao (China) lacked usable public school identifiers, and Luxembourg also lacked valid age, leaving no eligible records for the common school fixed effects design. The sixteen systems retained were United Arab Emirates, Belgium, Brazil, Brunei Darussalam, Colombia, Germany,

Dominican Republic, Georgia, Hong Kong (China), Ireland, Italy, Korea, Mongolia, Portugal, Romania and Chinese Taipei.

This fixed set came from the preceding screening. Each system met the original requirements of at least 500 complete records and 30 schools, at least 50 records and 20 schools in each grouped parental category, and adequate regional module coverage. The extension retained the same set throughout. The exclusion of Luxembourg and Macao therefore reflects the common analytic design and public variable availability; descriptive questions that do not require these variables can have a different coverage.

The background complete parent option domain, D, required usable identifiers, the index of economic, social and cultural status (ESCS), the file's sex indicator, age, grade, immigrant background, and positive final and all eighty replicate student weights. D contained 142,508 students in 5,234 schools. Pair added both focal responses, giving 94,262 students in 4,917 schools. Core also required the parental ordinary information rating, giving 93,687 students in 4,915 schools. Context required both parental ratings within D without requiring the classroom response. Usage added all four AI use responses to Core. Table 1 reports system specific counts, and the Supplementary Information defines the analytic domains and missing response codes.

Measures

The parental AI item, PA205Q13DA, asks respondents to rate their ability to assess the quality of AI chatbot information (OECD, 2025a). Its five substantive categories distinguish inability, difficulty doing the task independently, ability with effort, ease and unfamiliarity with the task. The primary lower rating indicator, L, combined inability, difficulty and unfamiliarity, using effort or ease as the comparison. Task unfamiliarity, code 5, remained a valid nominal response. Full category models retained all five categories, with ease as the reference. We use the shorter term lower parental rating for the grouped indicator after defining its components here.

The ordinary parental evaluation item, PA205Q02DA, asks about assessing information found online and uses the same response categories. The main model included its full nominal categories. A binary adjustment offered a comparison on the same Core sample. Within Context, additional analyses compared paired AI and ordinary ratings at two thresholds: ability with effort or ease, and ease alone. These thresholds distinguish perceived ability to undertake the task from perceived ease.

The student item, ST436Q16DA, asks how often the student assessed AI chatbot information using digital resources in school lessons during the current school year (OECD, 2025b). Responses were never or almost never, sometimes, often and very often. The main outcome, Y, was practice at least sometimes. Additional outcomes were often or very often, very often, and each exact response category. The joint low report indicator, J, combined L with never or almost never classroom practice. The parental and student items refer to the same type of information evaluation, but use different respondents and response formats.

Student covariates were ESCS, MALE, age, grade relative to the modal grade, and immigrant background indicators. Resource quintiles were defined within each system from ESCS. The Usage analysis used four items about schoolwork with AI chatbots: summarising a text, preliminary

research, drafting text and help with learning. Weekly use meant at least one activity undertaken about once or twice a week or more.

Survey estimation

International assessment analyses require attention to survey weights, replication and plausible values (OECD, 2009; Rutkowski et al., 2010). Each education system was estimated separately using the final student weight and all eighty replicate weights. Weighted inverse cumulative distribution cut points defined ESCS quintiles, with equal ESCS values kept together. Cut points were recomputed for every replicate. Quintiles were defined separately within Pair for observed rates and within D for nonresponse bounds.

The main adjusted association was the coefficient of L in a weighted linear probability model with school fixed effects, full nominal ordinary parental ratings and student covariates. Weighted school means were removed separately for every weight set. Exactly dependent columns were removed in a fixed sequence and recorded. The coefficient is an adjusted probability difference among students within the same school. Full category models replaced L with the four nonreference parental AI categories. The Supplementary Information gives the model specification, computation and prediction diagnostics (Supplementary Table 2).

International summaries gave each of the sixteen systems equal weight. The sampling variance of a mean was the sum of the system variances divided by 256. System directions, ranges and leave one system summaries describe heterogeneity around this fixed set mean. Pointwise 95% confidence intervals used 1.96 standard errors. Balanced repeated replication variance used Fay rho 0.5, equivalent to 0.05 times the sum of eighty squared replicate deviations. The international codebook confirms the eighty BRR Fay weights; direct confirmation of the numeric PISA 2025 rho remains pending, so sampling intervals are conditional on this convention. Holm adjustment was applied to the two primary contrasts (Holm, 1979). Other comparisons used exploratory pointwise intervals.

Sensitivity and response selection

Science performance adjustment fitted each model separately for all ten science plausible values and all replicate weights. Estimates were averaged over plausible values. Total variance combined the mean replicate variance with 1.1 times the between plausible value sample variance. A logistic sensitivity with school fixed intercepts used schools containing both outcome categories, with a linear model fitted to the same sample for comparison. All final and replicate fits were attempted, and convergence failures were retained in the reporting.

Nonresponse was examined using bounds and response weighting. The bounding analysis follows the distinction between what observed data identify and what additional assumptions supply (Manski, 1989). Within D, the lower J indicator counted records observed to satisfy both conditions. The upper indicator additionally allowed unresolved responses to satisfy J unless an observed answer ruled it out. Resource contrast bounds subtracted the upper highest-quintile endpoint from the lower lowest-quintile endpoint, and vice versa. Endpoints and their sampling intervals were estimated separately (Supplementary Table 3). These identification ranges describe uncertainty about missing responses within D.

Inverse response weighting used D schools containing at least five Core respondents. Within each system, a weighted response model included student backgrounds, school mean ESCS, log school sample size and nominal region indicators. Predicted response probabilities had a prespecified floor of 0.05. Respondent weights were divided by those probabilities, with response and outcome models refitted for every replicate. The comparison used the original survey weights on the same supported schools. Interpretation requires conditional missing at random and sufficient support (Seaman & White, 2013). Response target and weight diagnostics are reported in Supplementary Table 4.

Results

Joint low reports were concentrated among students with fewer resources

Among paired respondents, the equal system mean lower parental rating proportion was 41.46%, and 66.67% reported classroom AI evaluation at least sometimes. Joint low reports accounted for 15.45% (95% confidence interval 15.14 to 15.77). Figure 1 presents the full five by four joint distribution, retaining parental task unfamiliarity separately. This distribution describes how parental task ratings align with each level of classroom frequency.

Joint low report prevalence declined across resource quintiles, from 24.69% in the lowest to 18.24%, 14.59%, 11.47% and 8.27% in the highest (Figure 2a). The lowest minus highest contrast was 16.42 percentage points (95% confidence interval 15.45 to 17.39). The contrast was positive in all sixteen systems, ranging from 4.04 to 25.69 points (Supplementary Table 1). Leaving out one system at a time gave means of 15.80–17.24 points. Thus, the average pattern was accompanied by a consistent direction across the fixed system set.

An association remained among students attending the same school

In Core, lower parental AI ratings were associated with 3.87 percentage points less reported classroom practice after adjustment for school fixed effects, backgrounds and nominal ordinary parental ratings (95% confidence interval 3.02 to 4.73 points less; Figure 3a). Fifteen system estimates were negative, while Portugal's was slightly positive. System estimates ranged from 7.61 points less to 0.10 points more. Leave one system means ranged from 4.14 to 3.63 points less. Both primary contrasts retained statistical support after Holm adjustment.

Full category models estimated 5.77 points less classroom practice for parental inability relative to ease (95% confidence interval 4.55 to 6.99 points less), 2.78 points less for difficulty independently (1.54 to 4.01), 1.02 points less for effort (0.05 to 1.99), and 4.71 points less for unfamiliarity (2.25 to 7.18; Supplementary Figure 1 and Supplementary Table 5). The grouped lower rating association was also negative for practice often or very often, at 2.64 points less, and very often, at 1.93 points less.

Binary ordinary rating adjustment gave 3.85 points less. Adding all ten science plausible values gave 3.90 points less (95% confidence interval 3.04 to 4.76 points less). Response weighting gave 3.77 points less, compared with 3.78 points less before response adjustment in the same supported schools (Figure 3b; Table 2). The full sixteen system logistic contrast was unavailable because

Korea's final and eighty replicate fits failed convergence. The matched sample linear model gave 3.93 points less.

Response selection widened the range of the resource contrast

Core retained 38.48–87.93% of D student weight across systems. Allowing arbitrary missing focal responses in D gave a mean joint low report prevalence bound of 10.94–24.88% (Figure 2b). The lowest minus highest resource contrast lay between 0.69 and 30.40 points. The lower endpoint had a 95% sampling interval of −0.16 to 1.54 points, and eight systems had negative lower endpoint estimates. The positive respondent contrast therefore had stronger support than a corresponding contrast generalised to the wider domain under unrestricted focal nonresponse.

Parental thresholds distinguished different resource patterns

In Context, ability with effort or ease was reported by 58.08% for AI information and 90.03% for ordinary online information. Ease alone accounted for 29.78% and 64.57%, respectively. The within school ESCS slope of the paired AI minus ordinary rating was +5.68 percentage points per ESCS unit for effort or ease (95% confidence interval 5.23 to 6.13), and −0.75 for ease alone (−1.27 to −0.24; Supplementary Figure 2). The relative resource pattern consequently changed with the parental threshold.

In Usage, 59.71% reported at least one of the four AI activities weekly. Weekly use combined with both focal low reports accounted for 6.78% of that sample and 11.55% of weekly users. These are equal means of within system proportions with their respective denominators. Frequent schoolwork use and joint low evaluation reports therefore occurred together for a subset of respondents.

Discussion

The distribution of AI evaluation reports was socially patterned across linked family and classroom settings. About one quarter of respondents in the lowest resource quintile, compared with about one twelfth in the highest, combined a lower parental rating with never or almost never practising evaluation in lessons. The contrast was positive in every system studied. A smaller association remained among students attending the same school after adjustment for measured backgrounds and ordinary information evaluation. The joint distribution adds an educationally relevant view of inequality that separate accounts of families or classrooms would leave incomplete.

This contribution complements research on family AI learning. Studies of joint activities describe how parents and children explore AI together, while surveys connect parental literacy with educational perceptions and relational resources (Druga et al., 2022; Guo et al., 2025; Song et al., 2026). The present findings place one parental task rating alongside the adolescent's reported classroom experience. This gives family learning research a large scale distribution among paired respondents to investigate further within the participating systems. It also connects the learning ecology perspective to a specific practice that can be recognised in both settings (Barron, 2006).

The findings speak particularly closely to the curriculum question raised by the four-country study in this journal (Cohrssen et al., 2026). That document analysis identified uneven formal recognition

of AI literacy and considered the implications of relying on home environments. Our adolescent survey analysis examines a different age group, source of evidence and educational level. It shows that lower family task ratings and little classroom practice can coincide in a socioeconomic pattern. Read together, the two approaches suggest that curriculum development and family learning research benefit from asking which learners encounter evaluation opportunities across settings.

The within school result adds a level of detail to this account. Sharing a school did not eliminate differences in students' reports associated with parental ratings. These differences can inform work on class allocation, subject choices, individual participation, recognition of classroom tasks and learning exchanges within families. Earlier international studies demonstrate that school resources can exhibit both compensatory and socially patterned relationships (Loh et al., 2025; van de Werfhorst et al., 2022). Our finding concerns the particular experience of practising AI evaluation within schools. Studying that experience more directly can help connect institutional provision with the activities learners actually undertake.

The category and threshold results also matter for educational design. A parent who can undertake evaluation with effort may need different resources from one who is unfamiliar with the task. The reversal between effort-or-ease and ease-only AI–ordinary gradients shows why these responses should remain visible. Measurement work already distinguishes several AI competencies and perceived capacities (Carolus et al., 2023; Wang et al., 2023). The present comparison gives a concrete example of how an apparently modest threshold choice can change a socioeconomic description. Family resources can be designed around specific tasks and levels of familiarity, with evaluation exercises that allow parents and adolescents to learn together.

For classroom practice, the results motivate accessible activities that make judgement visible. Students can compare an AI response with independent sources, identify unsupported claims, and explain why a response should be revised. Research on online evaluation supplies examples of source investigation and evidence checking, while classroom AI studies show how literacy can be addressed within subject activities (Fang et al., 2025; Wineburg & McGrew, 2019). Accessible family versions of these exercises could support participation across settings. Their design should attend to parents' needs and the practical demands of engagement (Lewin & Luckin, 2010), alongside age appropriate and equitable use of AI (Miao & Holmes, 2023).

Response selection determines the reach of the descriptive finding. The clear gradient among paired respondents and the broad bounds within the background complete domain answer different questions. The response weighted association was close to its matched comparison, while the bounds showed how much wider the resource contrast could be under arbitrary missing focal responses. Presenting both allows the respondent pattern to inform educational questions without treating it as a fully identified distribution for everyone eligible to participate. This distinction is useful when linked family responses are used to guide further assessment or the recruitment of families for educational research.

The study has three main limitations. Single item self reports capture perceived ability and recalled practice, without performance validation or a measurement equivalence audit. The sixteen systems and linked respondents form selected analytic domains; nonresponse bounds limit generalisation to

D. The cross-sectional exploratory design supports associations, while the incomplete logistic sensitivity and provisional Fay coefficient qualify model and precision checks. Full response, computation and threshold diagnostics are provided in the Supplementary Information.

Conclusion

Linked parental and student reports reveal a socioeconomic concentration of low AI evaluation reports across family and classroom contexts. The resource gradient and the association within schools identify a concrete intersection for research on equitable AI education. Direct assessments, classroom observation and evaluations of shared family learning can build on this account to determine how opportunities to practise evaluation develop across settings and which educational activities improve that practice.

Data availability

The PISA 2025 public use data and master questionnaires are available from the OECD PISA database. The Zenodo record at https://doi.org/10.5281/zenodo.23418309 provides source provenance, aggregate system estimates, figure source data and sample flow tables. Individual student records are not redistributed.

Code availability

The Zenodo record at https://doi.org/10.5281/zenodo.23418309 provides analysis code snapshots, the aggregate validation script, figure scripts, recorded dependencies and aggregate outputs. Its README documents inputs and historical workspace dependencies for raw-data re-estimation.

Statements and declarations

Ethics and consent: This study analyzes publicly released, deidentified secondary survey data. No new participants were recruited and no additional personal information was collected. The analysis follows the access conditions of the public data source.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests: The author declares no relevant financial or non-financial competing interests.

Author contributions: The sole author was responsible for conceptualization, methodology, software, formal analysis, data curation, visualization, writing the original draft, and reviewing and editing the manuscript.

Acknowledgments: The author acknowledges the OECD and the participating education systems for making the PISA public-use data available.

References

Barron, B. (2006). Interest and self-sustained learning as catalysts of development: A learning ecology perspective. *Human Development, 49*(4), 193–224. https://doi.org/10.1159/000094368

Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., & Saavedra, A. (2021). Students' civic online reasoning: A national portrait. *Educational Researcher, 50*(8), 505–515. https://doi.org/10.3102/0013189X211017495

Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS - Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies. *Computers in Human Behavior: Artificial Humans, 1*(2), Article 100014. https://doi.org/10.1016/j.chbah.2023.100014

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior, 138*, Article 107468. https://doi.org/10.1016/j.chb.2022.107468

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends, 66*(4), 616–630. https://doi.org/10.1007/s11528-022-00715-y

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 4*, Article 100118. https://doi.org/10.1016/j.caeai.2022.100118

Cohrssen, C., Madanipour, P., Guarrella, C., Chen, M. Y., & Park, S. (2026). AI literacy for AI natives? A four-country comparison. *AI, Brain and Child, 2*(1), Article 21. https://doi.org/10.1007/s44436-026-00045-3

Druga, S., Christoph, F. L., & Ko, A. J. (2022). Family as a third space for AI literacies: How do children and parents learn about AI together? In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–17). Association for Computing Machinery. https://doi.org/10.1145/3491102.3502031

Fang, X., Ng, D. T. K., Li, D. X., Tao, Y. Y., Wu, H. K., & Chu, S. K. W. (2025). GenAI-supported digital multimodal composing: Facilitating EFL students' GenAI literacy, writing engagement, motivation, and abilities. *AI, Brain and Child, 1*(1), Article 23. https://doi.org/10.1007/s44436-025-00020-4

Guo, J., Law, S. T. T., Qiao, S. M., & Yeung, S. S. (2025). AI literacy, educational level, and parenting self-efficacy of children's education among parents of primary school students. *Computers and Education Open, 9*, Article 100318. https://doi.org/10.1016/j.caeo.2025.100318

Hargittai, E. (2010). Digital na(t)ives? Variation in Internet skills and uses among members of the "net generation". *Sociological Inquiry, 80*(1), 92–113. https://doi.org/10.1111/j.1475-682X.2009.00317.x

Hargittai, E., & Shafer, S. (2006). Differences in actual and perceived online skills: The role of gender. *Social Science Quarterly, 87*(2), 432–448. https://doi.org/10.1111/j.1540-6237.2006.00389.x

Helsper, E. J. (2012). A corresponding fields model for the links between social and digital exclusion. *Communication Theory, 22*(4), 403–426. https://doi.org/10.1111/j.1468-2885.2012.01416.x

Helsper, E. J., Veltri, G. A., & Livingstone, S. (2025). Parental mediation of children's online risks: The role of parental risk perception, digital skills and risk experiences. *New Media & Society, 27*(11), 5986–6005. https://doi.org/10.1177/14614448241261945

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70. https://www.jstor.org/stable/4615733

Hou, Y., Chen, S., & Lin, X. (2024). Parental digital involvement in online learning: Addressing the digital divide, not redressing digital reproduction. *European Journal of Education, 59*(2), Article e12635. https://doi.org/10.1111/ejed.12635

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2020). Citizens versus the Internet: Confronting digital challenges with cognitive tools. *Psychological Science in the Public Interest, 21*(3), 103–156. https://doi.org/10.1177/1529100620946707

Lewin, C., & Luckin, R. (2010). Technology to support parental engagement in elementary education: Lessons learned from the UK. *Computers & Education, 54*(3), 749–758. https://doi.org/10.1016/j.compedu.2009.08.010

Livingstone, S., & Helsper, E. (2007). Gradations in digital inclusion: Children, young people and the digital divide. *New Media & Society, 9*(4), 671–696. https://doi.org/10.1177/1461444807080335

Livingstone, S., & Helsper, E. J. (2008). Parental mediation of children's Internet use. *Journal of Broadcasting & Electronic Media, 52*(4), 581–599. https://doi.org/10.1080/08838150802437396

Livingstone, S., Ólafsson, K., Helsper, E. J., Lupiáñez-Villanueva, F., Veltri, G. A., & Folkvord, F. (2017). Maximizing opportunities and minimizing risks for children online: The role of digital skills in emerging strategies of parental mediation. *Journal of Communication, 67*(1), 82–105. https://doi.org/10.1111/jcom.12277

Loh, R. S. M., Kraaykamp, G., & van Hek, M. (2025). Plugging in at school: Do schools nurture digital skills and narrow digital skills inequality? *Computers & Education, 226*, Article 105195. https://doi.org/10.1016/j.compedu.2024.105195

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Manski, C. F. (1989). Anatomy of the selection problem. *The Journal of Human Resources, 24*(3), 343–360. https://doi.org/10.2307/145818

McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education, 46*(2), 165–193. https://doi.org/10.1080/00933104.2017.1416320

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535

Navarro, J. L., & Tudge, J. R. H. (2023). Technologizing Bronfenbrenner: Neo-ecological theory. *Current Psychology, 42*(22), 19338–19354. https://doi.org/10.1007/s12144-022-02738-3

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

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

OECD (2025a). *PISA 2025 parent questionnaire*. https://www.oecd.org/content/dam/oecd/en/data/datasets/pisa/pisa-2025-datasets/questionnaires/PARENT%20QUESTIONNAIRE%20PISA%202025.pdf

OECD (2025b). *PISA 2025 student questionnaire*. https://www.oecd.org/content/dam/oecd/en/data/datasets/pisa/pisa-2025-datasets/questionnaires/STUDENT%20QUESTIONNAIRE%20PISA%202025.pdf

OECD (2026). *PISA 2025 results (volume I): Future-ready students*. OECD Publishing. https://doi.org/10.1787/73451bc5-en

Robinson, L., Cotten, S. R., Ono, H., Quan-Haase, A., Mesch, G., Chen, W., Schulz, J., Hale, T. M., & Stern, M. J. (2015). Digital inequalities and why they matter. *Information, Communication & Society, 18*(5), 569–582. https://doi.org/10.1080/1369118X.2015.1012532

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher, 39*(2), 142–151. https://doi.org/10.3102/0013189X10363170

Scheerder, A., van Deursen, A., & van Dijk, J. (2017). Determinants of Internet skills, uses and outcomes. a systematic review of the second- and third-level digital divide. *Telematics and Informatics, 34*(8), 1607–1624. https://doi.org/10.1016/j.tele.2017.07.007

Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research, 22*(3), 278–295. https://doi.org/10.1177/0962280210395740

Song, X., Qu, H., Sun, L., Jin, J., Yi, X., Zhu, J., & Huang, H. (2026). Structuring the AI-enabled home learning environment: A gatekeeper model of digital capital, trust, and relational support. *Frontiers in Psychology, 17*, Article 1838910. https://doi.org/10.3389/fpsyg.2026.1838910

Su, J., & Yang, W. (2022). Artificial intelligence in early childhood education: A scoping review. *Computers and Education: Artificial Intelligence, 3*, Article 100049. https://doi.org/10.1016/j.caeai.2022.100049

van Deursen, A. J. A. M., & Helsper, E. J. (2015). The third-level digital divide: Who benefits most from being online? In L. Robinson, S. R. Cotten, J. Schulz, A. Williams, & T. M. Hale (Eds.), *Communication and information technologies annual: Digital distinctions & inequalities* (Vol. 10, pp. 29–52). Emerald Group Publishing Limited. https://doi.org/10.1108/S2050-206020150000010002

van Deursen, A. J. A. M., & van Dijk, J. A. (2014). The digital divide shifts to differences in usage. *New Media & Society, 16*(3), 507–526. https://doi.org/10.1177/1461444813487959

van de Werfhorst, H. G., Kessenich, E., & Geven, S. (2022). The digital divide in online education: Inequality in digital readiness of students and schools. *Computers and Education Open, 3*, Article 100100. https://doi.org/10.1016/j.caeo.2022.100100

Wang, B., Rau, P.-L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology, 42*(9), 1324–1337. https://doi.org/10.1080/0144929X.2022.2072768

Wang, R., Li, X., & Ng, D. T. K. (2026). Framing early childhood AI literacy: What did the literature review tell us? *AI, Brain and Child, 2*(1), Article 4. https://doi.org/10.1007/s44436-026-00028-4

Wineburg, S., & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record: The Voice of Scholarship in Education, 121*(11), 1–40. https://doi.org/10.1177/016146811912101102

Table 1 Sample flow across the sixteen education systems

System	D n	Pair n	Core n	Core schools	Retention %
United Arab Emirates	29,807	15,287	15,262	1,033	51.6
Belgium	3,646	2,831	2,788	119	74.9
Brazil	25,459	15,271	15,083	977	62.5
Brunei Darussalam	5,358	4,476	4,452	55	83.0
Colombia	6,581	5,057	4,987	256	74.8
Germany	5,477	2,174	2,163	213	38.5
Dominican Republic	5,859	3,931	3,909	250	66.3
Georgia	5,603	3,456	3,430	259	61.7
Hong Kong (China)	6,000	3,239	3,230	224	53.3
Ireland	6,078	4,976	4,957	172	81.4
Italy	8,555	6,327	6,295	279	75.1
Korea	6,669	5,448	5,448	196	81.7
Mongolia	6,779	6,010	5,979	191	87.9
Portugal	6,368	5,058	5,024	224	78.3
Romania	7,178	4,706	4,694	181	59.6
Chinese Taipei	7,091	6,015	5,986	286	84.3
Total	142,508	94,262	93,687	4,915	—

Counts are unweighted. D is the background complete parent option domain; Pair adds the two focal responses; Core also adds the ordinary parental rating. Weight retention is Core final student weight divided by D final student weight within a system. The equal system mean retention is 69.68%.

Table 2 Main association and fixed sensitivities

Specification	n	Schools	Difference pp	95% interval pp
Main nominal adjustment	93,687	4,915	-3.87	-4.73 to -3.02
Binary ordinary adjustment	93,687	4,915	-3.85	-4.67 to -3.02
All ten science plausible values	93,687	4,915	-3.90	-4.76 to -3.04
LPM in mixed outcome schools	91,190	4,396	-3.93	-4.80 to -3.06
LPM in response target schools	92,311	4,339	-3.78	-4.64 to -2.92
Response weighted LPM	92,311	4,339	-3.77	-4.65 to -2.88
Fixed effects logit probability contrast	91,190	4,396	Unavailable	Korea failed all fits

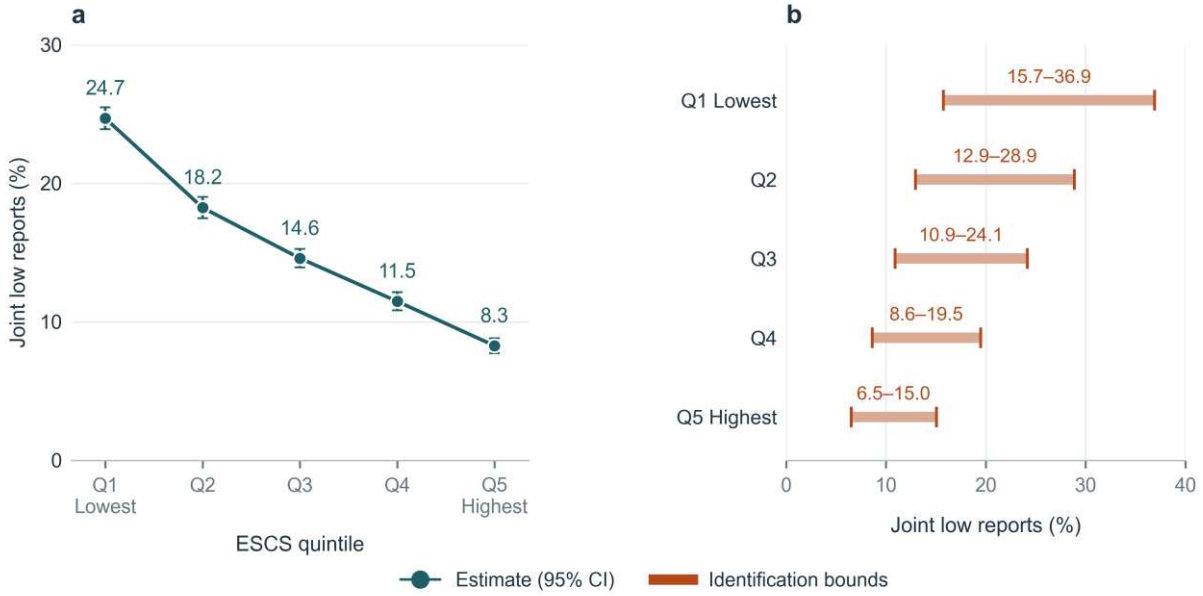
Each row is an equal system mean across the fixed sixteen systems. Differences compare a lower parental AI rating with effort or ease for reporting classroom AI evaluation at least sometimes. The linear models include school fixed effects, measured backgrounds, and nominal ordinary parental ratings unless stated. The science model uses all ten plausible values. Response weighted and matched response target models use the same 92,311 respondents in supported schools. The full sixteen system logit contrast is unavailable because Korea failed convergence. Interval conventions are defined in Methods.

Figure 1 Joint parental ratings and classroom practice



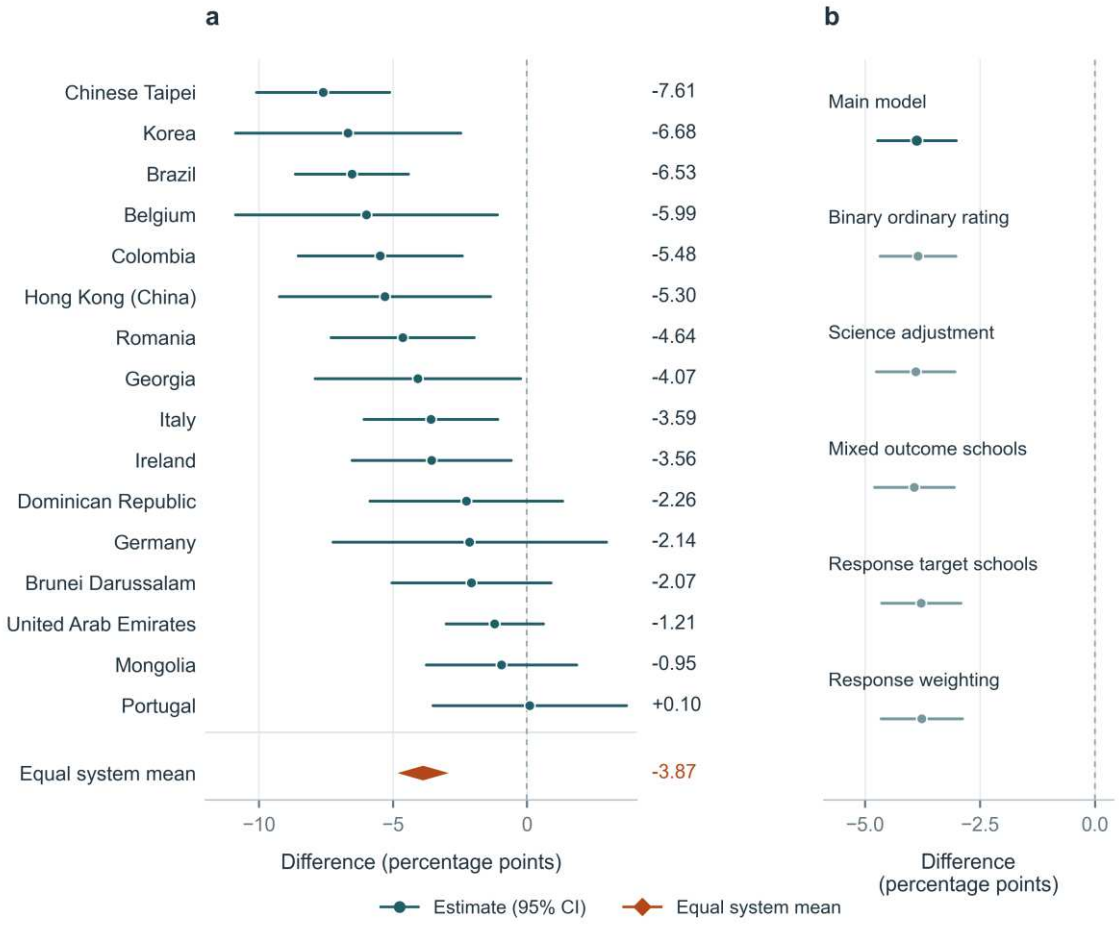
a, Joint distribution of all five parental AI evaluation categories and four classroom frequencies in Pair (n = 94,262). Cell values are percentages of all paired records within each system, averaged equally across sixteen systems; the twenty cells sum to 100% before rounding. Copper outlines identify the three cells comprising joint low reports (15.45%). Task unfamiliarity is displayed separately. b, Parental category shares obtained by summing the joint cells across classroom frequencies.

Figure 2 Socioeconomic gradient and nonresponse bounds



a, Joint low report prevalence across within system ESCS quintiles in Pair (n = 94,262), with pointwise 95% sampling intervals. Joint low reports combine parental inability, difficulty independently or task unfamiliarity with never or almost never classroom practice. b, Lower to upper identification bounds allowing arbitrary missing focal responses within D (n = 142,508). These ranges are identification intervals. Quintiles are separately defined within each sample and replicate. Both panels report equal system means.

Figure 3 Within school associations and specification checks



a, Adjusted probability differences for reporting classroom AI evaluation at least sometimes, comparing lower parental ratings with effort or ease. Systems are ordered by estimate. The diamond shows the equal system mean for Core (n = 93,687; 4,915 schools). b, Equal system means under alternative specifications, with corresponding samples in Table 2. Models include school fixed effects, measured backgrounds and ordinary parental ratings. Points and lines show estimates and pointwise 95% intervals. All intervals use the Fay convention in Methods. The full set logistic sensitivity is unavailable because Korea failed all 81 weight sets.

Supplementary Files

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

- Supplementaryinformation.docx