Journal of Pedagogical Research
Volume 10, Issue 5, 2026
https://doi.org/10.33902/JPR.202650050

Research Article

For whom are school digital citizenship programmes associated with adolescent life satisfaction? Person and context interaction across three independent PISA cohorts

Esen Spahi Kovaç[1] and Sezen Haskuka Abdullahi[2]

[1]*University "Ukshin Hoti" Prizren, Kosovo (ORCID: 0000-0002-3122-2592)*
[2]*Department of Preschool Education, Faculty of Education, Trakya University, Türkiye (ORCID: 0000-0001-6727-3298)*

Schools increasingly run programmes that prepare students for responsible internet behaviour, yet little is known about whether such programmes are associated with how adolescents evaluate their lives, or for whom. Drawing on the bioecological model and the differential susceptibility to media effects model, we examined associations between a school-level indicator of digital citizenship provision and adolescent life satisfaction in three independent cohorts of the Programme for International Student Assessment (2018, 2022 and 2025). School principals reported whether their school had a specific programme to prepare students for responsible internet behaviour. We used causal forests to estimate conditional associations with life satisfaction and to test whether these associations differed by socioeconomic status, home digital access, gender and immigrant background, and we tested whether the pattern found in 2018 appeared in the independent 2025 cohort. Students in schools with a programme did not report meaningfully higher life satisfaction than comparable students in schools without one, and the programme indicator was not associated with feeling pressured to be online or anxious without a device. In the primary analyses the association did not differ by the four student characteristics, and the 2018 ranking of students did not identify stronger associations in 2025. Analyses that took the survey design, country weighting, missing data and school-level differences into account led to the same overall conclusion, although conventional models that weighted countries equally showed small and inconsistent differences across student characteristics. The confidence intervals ruled out average associations larger than a few hundredths of a standard deviation. A principal's report that a programme exists tells us little about adolescent wellbeing; measures of programme content, implementation and students' experience are needed.

Keywords: Adolescence; Differential susceptibility; Digital citizenship; Life satisfaction; PISA; School context

Article History: Submitted 3 September; Revised 27 October 2026; Accepted 6 November 2026; Published online 16 November 2026

1. Introduction

Adolescents now conduct a large share of their friendships, schoolwork and leisure online. Schools have responded by introducing programmes that teach students to behave safely and responsibly on the internet, and in several countries such programmes have become a standard part of school provision. Whether these programmes go together with how adolescents feel about their lives remains an open question. Existing studies of digital technology and adolescent wellbeing have

Address of Corresponding Author

Sezen Haskuka, PhD, Department of Preschool Education, Faculty of Education, Trakya University, İsmail Hakkı Tonguç Campus, 22030 Edirne, Türkiye.

✉ haskukasezen82@gmail.com

How to cite: Spahi Kovac, E., & Haskuka Abdullahi, S. (2026). For whom are school digital citizenship programmes associated with adolescent life satisfaction? Person and context interaction across three independent PISA cohorts. *Journal of Pedagogical Research, 10*(5), 135–163. https://doi.org/10.33902/JPR.202650050

Copyright © 2026 by the author(s). This is an **open-access** article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).

mostly examined how much time young people spend online, or what they do there, rather than the school settings in which they learn to use digital tools (Odgers & Jensen, 2020; Orben & Przybylski, 2019).

A second question may matter more than the first. Developmental theories hold that one and the same context is experienced differently by different young people. If that is right, an average association between a school programme and wellbeing, whether positive or null, may conceal groups of students for whom the programme is associated with better or worse outcomes. The present study asks for whom school digital citizenship programmes are associated with adolescent life satisfaction, and whether the answer holds across cohorts of adolescents surveyed before, during and after the coronavirus disease 2019 [COVID-19] pandemic.

1.1. The School as a Developmental Context

Development is described by the bioecological model as resulting from proximal processes, the regular and increasingly complex interactions between a person and the people, objects and symbols in their immediate environment (Bronfenbrenner & Morris, 2006). A strong claim about variation is advanced by the model. Proximal processes are expected to differ systematically in form, strength and direction according to the characteristics of the person, the features of the context, and the historical period in which development takes place. Internet behaviour programmes in schools represent a feature of the school microsystem. Home digital access available to a student link the home and the school, forming part of the mesosystem. Viewed from this perspective, a uniform association between any school feature and adolescent wellbeing would constitute the exception rather than the rule.

Life satisfaction, the cognitive component of subjective wellbeing, forms the focus of this study and refers to a person's overall judgement of their life (Diener, 1984). In adolescence it is sensitive to school experiences and peer relationships and serves as a broad indicator of adjustment (Proctor et al., 2009). Self-determination theory (Ryan & Deci, 2000) and stage-environment fit theory (Eccles et al., 1993) give a further reason why a digital citizenship programme might be associated with life satisfaction: a school that helps students feel competent and in control in digital settings could support their basic psychological needs and match the developmental needs of adolescents whose social lives are increasingly online. We treat these two perspectives as supporting rationales and build the hypotheses on the bioecological and differential susceptibility perspectives.

1.2. Differential Susceptibility

People are proposed by differential susceptibility theory to differ in the strength of their response to environments, with some individuals responding more strongly than others to both negative and positive conditions (Belsky & Pluess, 2009). The positive end of this variation is described by vantage sensitivity, in which some individuals benefit more than others from supportive experiences such as a helpful programme (Pluess & Belsky, 2013). Application of this reasoning to digital settings is made by the differential susceptibility to media effects model (Valkenburg & Peter, 2013). Distinctions are drawn by the model among dispositional susceptibility, which rests on stable characteristics such as gender; developmental susceptibility, which depends on age and developmental stage; and social susceptibility, which arises from family, peer and cultural circumstances. This model guided the choice of the student characteristics along which associations might differ. Dispositional susceptibility is represented by gender. Social susceptibility is represented by socioeconomic status, home digital access and immigrant background. Examination of developmental susceptibility was not possible because the Programme for International Student Assessment [PISA] samples students of a single age.

Opposite predictions about socioeconomic status and home digital access are made by two accounts. Students with fewer resources at home are held by the compensation account to have more to gain from what a school offers, because the school fills a gap that the home leaves open. This way of reading the evidence on adolescents' online communication has been advanced (Valkenburg & Peter, 2007). The reverse is held by the cumulative advantage account, namely that

students who already have more resources are better placed to use new opportunities, so that those who start ahead gain most (Kraut et al., 2002). Consistency with compensation is shown by evidence that online risks tend to reflect and amplify existing offline vulnerabilities (Odgers & Jensen, 2020), since a protective programme would matter most for the most vulnerable students. A further possibility is added by evidence that digital skills increase exposure to both opportunities and risks (Livingstone & Helsper, 2010). Greater worry rather than greater wellbeing could, for some students, accompany a programme that raises awareness of online risks. With respect to immigrant background, a higher risk of adverse school experiences is shown by recent international evidence for students with an immigrant background (Organisation for Economic Co-operation and Development [OECD], 2026a), which would lead us to expect a stronger association in this group if the programme is protective.

1.3. What a School-Level Programme Indicator Can and Cannot Capture

Four levels need to be kept apart: whether a programme is available at a school, how it is implemented (its content, intensity and quality), whether and how much individual students are exposed to it, and whether it is effective for particular outcomes. Reviews of prevention and promotion programmes find that the level of implementation is related to the outcomes that programmes obtain (Durlak & DuPre, 2008), so schools that report the same programme label may offer students very different experiences. Programmes also differ in their targets. Interventions aimed at cyberbullying, for example, can reduce cyberbullying perpetration and victimisation (Gaffney et al., 2019), but these outcomes are narrower and closer to the programme than global life satisfaction. A systematic review of psychological wellbeing in education likewise describes a multidimensional picture in which digital literacy appears alongside institutional support, and excessive social media use appears among the risk factors (Sarzhanova & Nurgabdeshov, 2025).

The bioecological model places development in proximal processes, that is, in what students actually experience and do. Work on instructional features points in the same direction: studies of teacher humour model students' perceptions of instruction as the route through which a feature of teaching is related to their emotions and motivation (Banaruee & Bieg, 2026; Bieg et al., 2026). Studies of social media show that responses to the same digital exposure differ markedly from one adolescent to another (Beyens et al., 2020) and across developmental windows (Orben et al., 2022). Neither kind of variation can be seen in a single school-level report.

A principal's report that the school has a programme records availability only. The data in this study therefore allow us to ask whether this distal indicator is associated with life satisfaction and whether the association varies across the measured student characteristics. They do not allow us to test the proximal processes through which differential susceptibility is proposed to operate. We therefore frame the hypotheses below as tests of the indicator and not of those processes.

1.4. Stability across Cohorts

That the same context can operate differently across historical periods is reminded to us by the time component of the bioecological model. Different digital worlds were those in which the adolescents surveyed in 2018, 2022 and 2025 grew up. Surveying of the first cohort took place before the COVID-19 pandemic, of the second during a period of school closures and remote learning, and of the third as generative artificial intelligence tools became widely available. Should differences in how students relate to a school programme reflect stable processes linking persons and contexts, a pattern of differentiation found in one cohort should appear again in an independent cohort. Should the pattern reflect the circumstances of one cohort, or chance, it should not. A developmental question about the stability of person and context interaction, and not only a statistical one, is therefore posed by the question of replication across cohorts. Reliance on a single sample characterises most studies of heterogeneity in associations between digital contexts and wellbeing, which leaves open whether the subgroup differences they report would be found again.

1.5. The Present Study

Associations between school digital citizenship programmes and adolescent life satisfaction were examined in three independent cohorts of PISA. The hypotheses were derived from the theories reviewed above.

Hypothesis 1 (H1). Students in schools with a programme to prepare students for responsible internet behaviour report higher life satisfaction than comparable students in schools without such a programme. We expected this average association to be small.

Hypothesis 2 (H2). The association differs by student characteristics that precede the programme. It is stronger for students with lower socioeconomic status and fewer home digital access (H2a, compensation) or, alternatively, for students with higher socioeconomic status and more home digital access (H2b, cumulative advantage); it differs by gender (H2c) and by immigrant background (H2d); and it may be close to zero or negative in some groups (H2e).

Hypothesis 3 (H3). The pattern of differentiation identified in the 2018 cohort is reproduced in the independent 2025 cohort.

Hypothesis 4 (H4). In the 2025 cohort, the programme is more strongly associated with proximal digital wellbeing outcomes, namely feeling pressured to be online and feeling anxious without a device, than with life satisfaction. Association strength is judged in the favourable direction for all three outcomes, that is, higher life satisfaction, less pressure to be online and less anxiety without a device.

We treated the three cohorts as three independent tests of the same question rather than as a time series, and we drew no inference about change over time.

The study makes two contributions. First, existing studies of digital technology and adolescent wellbeing have mostly examined time spent online or online activities rather than school provision (Odgers & Jensen, 2020; Orben & Przybylski, 2019); this study adds evidence on a school-level indicator across three independent cohorts and not a single sample. Second, it separates two questions that are often merged, namely whether the indicator is associated with wellbeing on average and whether this association varies across students, and it tests the second question in an independent cohort. The study is not a test of programme effectiveness, because the indicator records availability only.

2. Method

2.1. Research Design

The study is a secondary analysis of public-use data from three cycles of PISA, conducted by the OECD in 2018, 2022 and 2025. PISA assesses 15-year-old students enrolled in school and collects questionnaire data from students and from school principals (OECD, 2026b). Sampling follows a two-stage stratified design in which schools are sampled with probability proportional to size and students are then sampled within schools. Each cycle draws a new sample of schools and students, so the three cycles are independent cross-sectional samples of different cohorts. No school or student can be followed across cycles. We downloaded the student and school questionnaire files for each cycle from the OECD PISA database website. The 2025 files were released on 8 September 2026. The 2015 cycle was examined but excluded because its school questionnaire did not include the programme item. The design is repeated cross-sectional, with students nested in schools and schools in countries. The programme indicator is measured at school level and the outcomes at student level. The target population is 15-year-old students enrolled in school in the participating countries. Unless stated otherwise, each student contributes in proportion to the PISA final student weight, so larger countries carry more weight in pooled estimates.

2.2. Participants

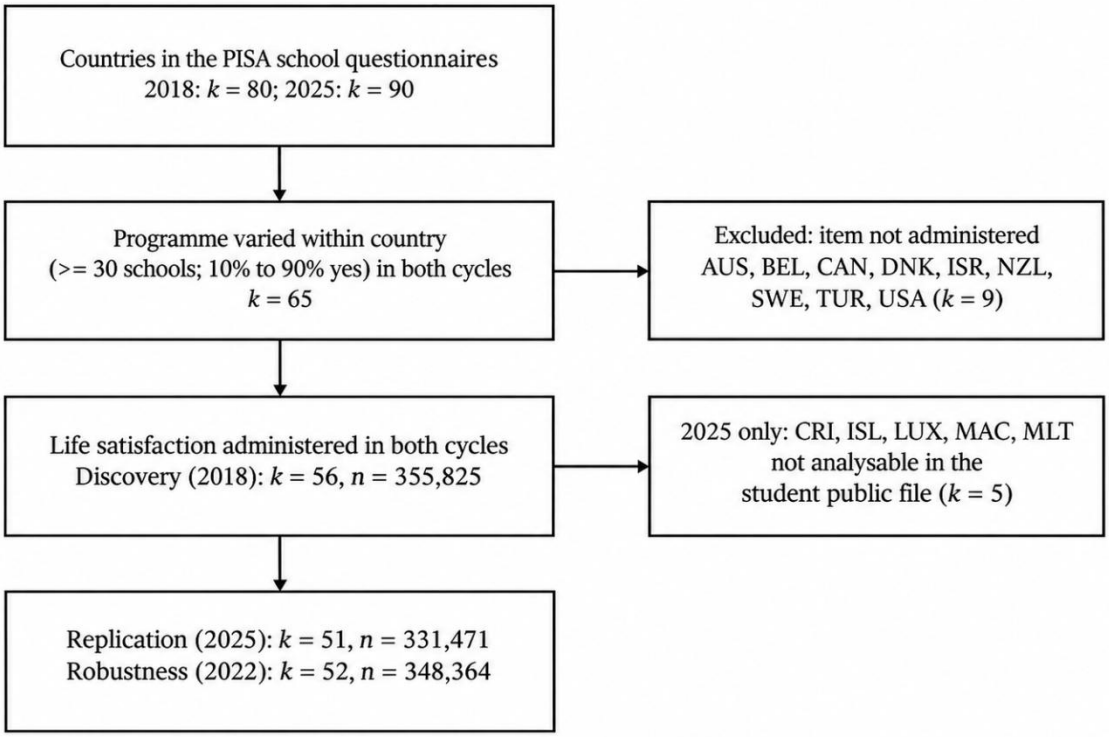
2.2.1. Country selection

Countries entered the analysis if they met two conditions. First, the programme item had to vary within the country, so that students in schools with and without a programme could be compared under the same national conditions. A country met this condition if at least 30 schools answered the item and between 10% and 90% of responding schools reported having a programme. Second, the life satisfaction item had to be administered in the cycle, which we operationalised as fewer than 95% of student responses being missing.

Figure 1 shows the selection process. The school questionnaires contained the programme item for 80 countries in 2018 and 90 countries in 2025. Of these, 76 met the first condition in 2018, 75 in 2022 and 83 in 2025, and 65 met it in both 2018 and 2025. Nine of these 65 countries did not administer the life satisfaction item in 2018 or 2025 (Australia, Belgium, Canada, Denmark, Israel, New Zealand, Sweden, Türkiye and the United States), leaving 56 countries for the discovery analysis. In the 2025 replication, five further countries could not be analysed. For Iceland, Luxembourg, Macao (China) and Malta, the 2025 student public-use file contained no student records at all, although their schools appear in the school file, and for Costa Rica the socioeconomic, immigrant background and home resource variables were absent. The replication analysis therefore used 51 countries. The 2022 cohort analysis used the 52 countries that met both conditions in 2022 and had complete data on the analysis variables. Appendix Table A1 lists the countries in each analysis.

We checked the absence of the four countries against the school identifiers, which begin with the numeric country code. The 2025 student file contains 729,963 students from 85 countries and economies, and each country code is paired with a single identifier prefix; none of the identifiers carried the prefixes of Iceland, Luxembourg, Macao and Malta (352, 442, 446 and 470). The absence is therefore one of records and not of a mismatch between country codes. We could not establish from the documentation why these records are not in the public-use release.

Figure 1
Selection of Countries for the Analyses



Note. k = number of countries; n = number of students with complete data on the analysis variables.

2.2.2. Analysis samples

Table 1 describes the analysis samples. After the country-selection rules were applied and students with missing data on the programme item, life satisfaction or the four student characteristics were excluded, the 2018 discovery sample comprised 355,825 students in 14,012 schools in 56 countries, the 2022 sample 348,364 students in 12,586 schools in 52 countries, and the 2025 replication sample 331,471 students in 13,061 schools in 51 countries. The cohorts are independent cross-sectional samples, and schools and students cannot be linked across cycles.

2.3. Data Collection

The data come from the student and school questionnaires of PISA 2018, 2022 and 2025. The instruments are described in turn in the following. The PISA documentation gives their coding, and we checked the value labels and missing-value codes of every variable used.

2.3.1. School digital citizenship programme

School principals reported whether their school had "a specific programme to prepare students for responsible internet behaviour" (yes or no). The item was part of the school questionnaire in all three cycles, with the variable name SC156Q05HA in 2018 and SC190Q08JA in 2022 and 2025. The wording is the same across cycles, but in 2018 the item sits in a different block of questions. We treat this as a constraint of the data rather than as a finding and make no claim about the measurement comparability of the item across cycles. Students were coded 1 if their principal reported a programme and 0 otherwise. Principals' responses therefore indicate programme availability only. The item does not record the content, duration, frequency or quality of the programme, whether students took part, or how principals interpreted "a specific programme". We accordingly speak of programme availability and avoid language that implies implementation, student exposure or effectiveness.

2.3.2. Life satisfaction

Students answered the question "Overall, how satisfied are you with your life as a whole these days?" on an 11-point scale from 0 (not at all satisfied) to 10 (completely satisfied) (ST016Q01NA). Single-item measures of global life satisfaction are standard in large-scale surveys of adolescents and are sensitive to school experiences (Proctor et al., 2009). Because the programme item and the life satisfaction item are single items, their internal consistency cannot be estimated. The life satisfaction item has the same wording and response format in the three cycles, and its comparability across countries is assumed here and not tested.

2.3.3. Digital wellbeing

The 2025 student questionnaire included two items that capture the aspects of digital life that a responsible internet behaviour programme most directly addresses. Students indicated how much they agreed that they "feel pressured to be online and respond to messages immediately" (ST322Q06JA) and that they "feel nervous or anxious when [their] device is not near [them]" (ST322Q07JA), on a four-point agreement scale. These items were not available in earlier cycles and were used only for H4. These two items are closer to digital life than life satisfaction, but they are not necessarily outcomes targeted by every digital citizenship programme, since some programmes focus on privacy, security, misinformation, cyberbullying or responsible communication. We therefore treat them as proximal indicators of digital wellbeing and not as programme targets.

2.3.4. Student characteristics preceding the programme

Following the differential susceptibility to media effects model, we examined four characteristics. Socioeconomic status was measured with the PISA index of economic, social and cultural status [ESCS], a composite of parental education, parental occupation and home possessions. Gender was coded from the student questionnaire (girl = 1). Immigrant background was coded 1 for first- and

second-generation students and 0 for students without an immigrant background (variable IMMIG). Home digital access was the number of two items the student endorsed, namely having a computer that can be used for schoolwork and having an internet connection at home (range 0 to 2). These items are ST011Q04TA and ST011Q06TA in 2018 and ST250Q02JA and ST250Q05JA in 2022 and 2025, and they belong to the core student questionnaire in all three cycles. We did not use the information and communication technology availability index (variable ICTHOME) as the main measure, because in 2018 it comes from an optional questionnaire and 39% to 64% of values were missing in the analysis samples. The two-item home digital access index captures basic access to a computer and an internet connection and not broader digital resources or digital capital, so it may have limited sensitivity in contexts where almost all students have such access.

We deliberately excluded bullying victimisation, loneliness and school belonging from the set of characteristics. A school programme that addresses online behaviour could itself be associated with these experiences, and estimating how the programme association varies across a characteristic that the programme may alter would bias the estimates. Life satisfaction therefore cannot be examined for students who were bullied with this design, a point we return to in the Discussion.

2.3.5. Covariates

All models included the student's grade, the school mean of ESCS, school type (public or private; SC013Q01TA), school location on a five-point scale from village to large city (SC001Q01TA) and indicators for each country. School size and the ratio of computers to students were available in 2018 and 2022 but not in 2025 and were therefore not used, so that the discovery and replication models contained the same covariates. These covariates were chosen because they precede, or are unlikely to be changed by, the programme and are available in all three cycles. They are not a complete adjustment set: leadership, safeguarding culture, school resources and prior digital problems were not measured (see Limitations).

2.3.6. Data preparation and missing data

All data preparation and analysis were carried out in R version 4.5.2 (R Core Team, 2025). The SPSS files were read with the haven package, version 2.5.5 (Wickham et al., 2025), which converts the user-defined missing value codes of PISA (95 to 99 for variables with two-digit ranges and 5 to 9 for single-digit items) to missing values. We verified this conversion for every analysis variable by checking value ranges and value labels. The 2025 student file is compressed with the Deflate64 method, which the R extraction functions cannot read, and was therefore extracted with the Windows file manager before being read into R. For each cycle, student records were merged with school records by the international school identifier (CNTSCHID). Students with missing data on the programme item, life satisfaction or any of the four student characteristics were excluded from each model. Within the analysed countries, the programme item was missing for 4.1% of students in 2018 and 6.9% in 2022.

Missing life satisfaction data required closer attention. In the countries retained for analysis, the median proportion of missing responses was 5.6% in 2018 and 4.1% in 2022. In 2025 it was higher. Across the 56 discovery countries, 86.5% of 2025 responses were valid, 11.7% were coded as no response and 1.5% as not administered, with a country median of 9.3% no response and values above 20% in six countries (Argentina, Bulgaria, Kosovo, Lebanon, North Macedonia and Uruguay). We therefore ran a sensitivity analysis that excluded countries with more than 20% no response to the life satisfaction item in 2018 or 2025. That analysis was run in the 2018 discovery sample; sensitivity analyses in the 2025 sample were added during revision and are described below. The 2025 public-use file lacks the gender variable for eight countries (Argentina, Chile, Colombia, Germany, Ireland, the Netherlands, Spain and Uruguay). Because gender was needed only to test H2c, which was tested in 2018, gender was left out of the replication models so that these countries could be retained. Students with missing data on the programme item, life satisfaction or the four student characteristics were excluded from the model concerned. Missing

values in the remaining covariates (grade, school type and school location) were handled internally by the forest algorithm, which sends observations with a missing value to either side of a split.

2.4. Data Analysis

2.4.1. Primary analyses

We estimated conditional associations between the programme and each outcome with causal forests (Wager & Athey, 2018), implemented in the grf package, version 2.6.1 (Athey et al., 2019; Tibshirani et al., 2026). A causal forest is an ensemble of decision trees that estimates, for each student, the expected difference in the outcome between being in a school with and without a programme, given the student's characteristics and covariates. The method does not require the analyst to specify in advance how this difference varies across characteristics. Each forest contained 2,000 trees, used the PISA final student weight (W_FSTUWT) as the sample weight and treated schools as clusters, so that standard errors reflect the clustering of students within schools. Cluster weights were not equalised, because equalising them cannot be combined with sample weights in grf. Country indicators were included among the covariates, but because a forest can split on other covariates before it splits on country, this does not by itself guarantee strictly within-country comparisons; country fixed-effects models and country-specific estimates were therefore added (see Robustness and Sensitivity Analyses). The random seed was set to 20260923. Results are reported with standard errors [*SE*] and 95% confidence intervals [CI], and standardised quantities are expressed in standard deviations [*SD*] of life satisfaction. Causal forests were preferred to a single parametric interaction model because they do not require the form of the heterogeneity to be specified and can examine several characteristics together. Because the four student characteristics were chosen on theoretical grounds, a conventional interaction model is also reported as a transparent, direct test of the same hypotheses (Appendix Table A2). The pooled estimates describe an average across countries whose school systems, programmes and interpretation of the programme item may differ, and the country-specific analyses below examine how far this average hides differences between countries. The target population was 15-year-old students in the participating countries, with each country contributing in proportion to its PISA final student weights. Because these weights give larger countries more influence, we also report analyses in which each country received equal weight.

These estimates can be read as programme effects only if schools with and without a programme are comparable once the measured characteristics are taken into account, if every country contains schools with and without a programme, and if one school's programme is unrelated to outcomes in other schools. The first assumption cannot be tested with cross-sectional data. We therefore report all estimates as conditional associations. Because the estimates were close to zero, E-values (VanderWeele & Ding, 2017) would equal one and add no information; we instead report the size of association that each confidence interval excludes.

The analysis had three steps. In the discovery step, we estimated the average association across students in the 2018 cohort (H1) and projected the student-level estimates onto the four characteristics with a best linear projection, which gives the change in the conditional association per standard deviation of each characteristic (H2; Athey & Wager, 2019). We also report the calibration test of the forest, in which a differential prediction coefficient near one indicates that the forest captures real variation in associations, and the proportion of student-level estimates below zero (H2e). In the replication step, we used the 2018 forest to rank the 2025 students by their predicted association and asked whether this ranking identified students with stronger associations in an independently estimated 2025 forest. This was assessed with the rank-weighted average treatment effect [RATE], using the area under the targeting operator characteristic curve [AUTOC] (Yadlowsky et al., 2025). A value near zero indicates that the 2018 ranking does not separate students with stronger and weaker associations in 2025 (H3). In the third step, we repeated the discovery analysis in the 2022 cohort. This cohort was surveyed during the COVID-19

pandemic, and we drew no conclusions about the pandemic from the comparison. For the replication step, the forest that ranks students was refitted in the 2018 data without gender, in the 56 discovery countries and with the same covariates as the 2025 forest, because gender was unavailable for eight countries in 2025. This refitted forest, and not the discovery forest used for H1 and H2, produced the predicted associations for the 2025 students. In addition, we fitted a forest in the 2025 data and projected its student-level estimates on socioeconomic status, home digital access and immigrant background, which gives an independent 2025 assessment of the three moderators that could be examined in both cohorts.

For H4, we estimated separate forests in the 2025 cohort for life satisfaction and for the two digital wellbeing items, each standardised to a mean of zero and a standard deviation of one, and compared the resulting average associations, and we later tested the contrasts between outcomes directly (see below). As a sensitivity analysis, we repeated the discovery analysis after excluding countries with more than 20% no response to the life satisfaction item in 2018 or 2025.

2.4.2. Robustness and sensitivity analyses

Several further sets of analyses examined whether the conclusions depended on the estimator, the weighting of students and countries, the comparability of schools or the treatment of missing outcome data. They were added during revision in response to reviewers' comments and are exploratory.

First, to benchmark the forests against a conventional survey-aware estimator, we regressed each outcome on the programme indicator and the covariates of the forests with country fixed effects. Standard errors came from the 80 replicate weights supplied by the OECD, using Fay's variant of balanced repeated replication [BRR] (OECD, 2024; Rutkowski et al., 2010). Country fixed effects were chosen instead of random intercepts because they guarantee that comparisons are made within countries and because the countries are not a random sample of countries. A second model added interactions between the programme indicator and the four standardised student characteristics. Both models were estimated with the PISA final student weight and with weights that gave each country equal total weight, which changes the target population from all 15-year-old students in the participating countries to the average country. Benjamini and Hochberg's (1995) procedure was applied to the eleven interaction terms of the equal-country model, and forests with equal-country weights were also fitted.

Second, country-specific associations were estimated with the same regression within each country, with replicate-weight standard errors, and combined in a random-effects meta-analysis fitted by restricted maximum likelihood (Viechtbauer, 2010). We report the mean association, the between-country standard deviation (τ), the proportion of variation due to heterogeneity (I^2), the 95% prediction interval, the number of countries with an interval that excluded zero and the range of the pooled estimate when each country was left out in turn.

Third, schools with and without a programme were compared on school mean ESCS, school type (public or private), school location, mean grade and the shares of immigrant students, girls and students with home digital access, using standardised mean differences [SMD]. A school-level propensity score was estimated by logistic regression of programme availability on school mean ESCS, school type, school location and country. Scores were truncated at .05 and .95, and the regression was re-estimated with inverse-propensity weights multiplied by the PISA weights (Austin & Stuart, 2015).

Fourth, we compared respondents and non-respondents to the 2025 life satisfaction item on observed characteristics, separately by programme status, and repeated the 2025 forest analysis, including the ranking by the 2018 forest, after excluding (a) the six countries with more than 20% no response, Argentina, Bulgaria, Kosovo, Lebanon, North Macedonia and Uruguay, and (b) the eleven countries in which more than 20% of students lacked a life satisfaction score for any reason (these six plus Brazil, Chile, Germany, the Dominican Republic and Peru). We also weighted respondents by the inverse of their estimated probability of responding, which was modelled from

the programme indicator, ESCS, immigrant background, home digital access, grade, school mean ESCS, school type, school location and country (Seaman & White, 2013). This approach assumes that missingness depends only on these observed variables.

Fifth, H4 was tested directly. The three outcomes were aligned so that higher values were favourable, which required multiplying feeling pressured to be online and feeling anxious without a device by −1, and were standardised and analysed for the same students. The differences between the programme associations for the outcomes were estimated from the doubly robust scores of the forests, with variance that is robust to clustering of students within schools and therefore reflects the dependence of the outcomes within students, and from the replicate-weight regression. H4 predicts a positive difference between the association for a digital wellbeing outcome and the association for life satisfaction.

Sixth, to assess the moderators independently in 2025, we fitted a forest in the 2025 data and projected its student-level estimates on ESCS, home digital access and immigrant background (gender was unavailable), and we report 95% confidence intervals for all projections in all cohorts.

Seventh, each average association and its 95% confidence interval were expressed in standard deviations of life satisfaction, using the weighted standard deviation of each cohort, and we report the largest magnitude that the interval excludes. As an exploratory complement, we ran two one-sided tests [TOST] of equivalence against bounds of ±0.05 and ±0.10 standard deviations (Lakens, 2017). No smallest effect size of interest [SESOI] was set before the analyses, and bounds taken from benchmarks are the weakest justification for a SESOI (Lakens et al., 2018). The bounds were chosen from effect-size conventions for very small associations (Funder & Ozer, 2019; Kraft, 2020): a difference of 0.10 standard deviations between groups of about equal size corresponds to a correlation of about .05. The tests are therefore descriptive, and we do not claim equivalence on their basis.

This study used anonymised, publicly available secondary data and did not require ethical approval.

3. Results

3.1. Sample Characteristics

Table 1 describes the three analysis samples. The discovery sample comprised 355,825 students in 14,012 schools across 56 countries. The replication sample comprised 331,471 students in 13,061 schools across 51 countries, and the 2022 sample 348,364 students in 12,586 schools across 52 countries. Just over half of the students in 2018 (53.7%) attended a school whose principal reported a programme to prepare students for responsible internet behaviour, rising to 63.1% in 2022 and

Table 1
Characteristics of the Analysis Samples by PISA Cohort

Characteristic	*2018 (discovery)*	*2022 (comparison)*	*2025 (replication)*
Countries (k)	56	52	51
Schools	14,012	12,586	13,061
Students	355,825	348,364	331,471
Programme, % of students	53.7	63.1	64.9
Programme, % of schools	53.0	59.2	58.8
Life satisfaction, M (*SD*)	7.26 (2.58)	6.96 (2.61)	7.63 (2.38)
ESCS, M	−0.83	−0.83	−0.83
Girls, %	50.9	50.9	51.7
Immigrant background, %	3.5	3.7	4.1
Computer and internet at home, %	58.2	59.1	60.2

Note. Percentages and means are weighted with the PISA final student weight. ESCS = PISA index of economic, social and cultural status. In 2025 the percentage of girls is based on students with gender information, because the public-use file lacks gender for eight countries.

64.9% in 2025. At the school level the corresponding figures were 53.0%, 59.2% and 58.8%. Mean life satisfaction was 7.26 (*SD* = 2.58) in 2018, 6.96 (*SD* = 2.61) in 2022 and 7.63 (*SD* = 2.38) in 2025. About three in five students had both a computer for schoolwork and an internet connection at home, and between 3.5% and 4.1% had an immigrant background. The low share of students with an immigrant background reflects the exclusion of several countries with large immigrant populations that did not administer the life satisfaction item.

3.2. Average Association (H1)

In the 2018 discovery cohort, students in schools with a programme did not report higher life satisfaction than comparable students in schools without one (Table 2 and Figure 2). The estimated difference was 0.002 points on the 0 to 10 scale (*SE* = 0.035, 95% CI [−0.067, 0.070], *p* = .962). The confidence interval excludes differences larger than about 0.07 points, which corresponds to 0.03 standard deviations of life satisfaction (the upper limits of the intervals correspond to 0.04 standard deviations in 2022 and 0.02 in 2025). H1 was not supported. The average association was also close to zero in the 2025 cohort (0.006 points, *SE* = 0.025, 95% CI [−0.042, 0.054], *p* = .807) and in the 2022 cohort (0.050 points, *SE* = 0.032, 95% CI [−0.013, 0.114], *p* = .120). Conventional regression and propensity-score models, reported under Robustness Analyses (Table 5), placed the 2025 association slightly above zero.

Table 2

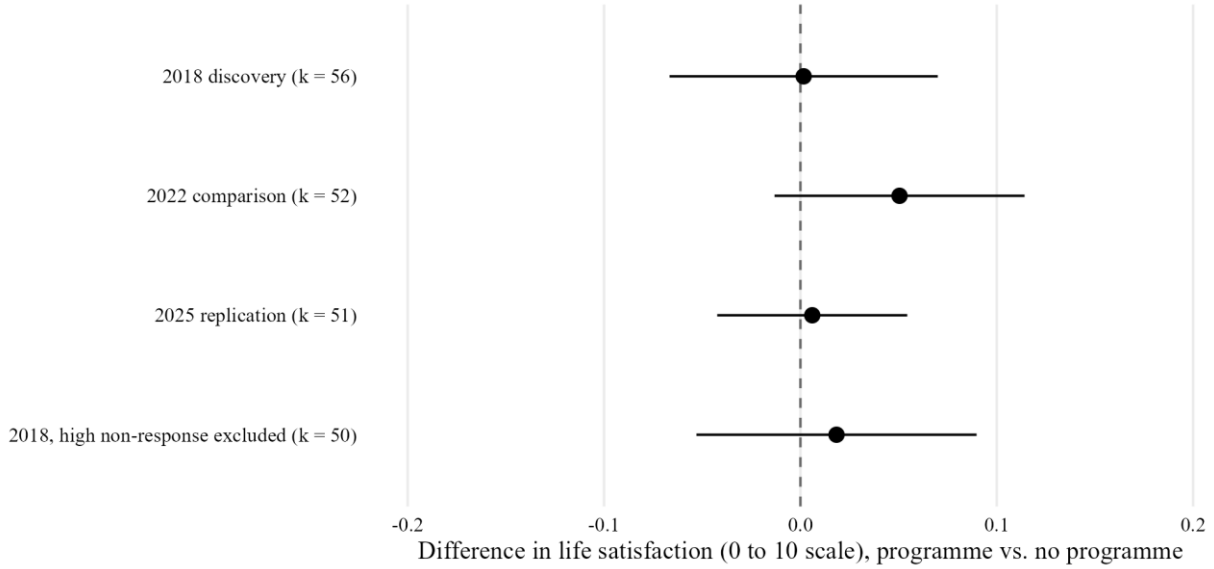
Average Associations Between the School Programme and Life Satisfaction and the Replication Test

Analysis	*Cohort*	*k*	*n*	*Estimate*	*SE*	*95% CI*	*p*
H1: average association	2018	56	355,825	0.002	0.035	[−0.067, 0.070]	.962
H3: RATE (AUTOC), 2018 ranking	2025	51	331,471	−0.0003	0.026	[−0.052, 0.052]	.992
Average association	2025	51	331,471	0.006	0.025	[−0.042, 0.054]	.807
Average association	2022	52	348,364	0.050	0.032	[−0.013, 0.114]	.120
Sensitivity: high non-response excluded	2018	50	324,262	0.018	0.036	[−0.053, 0.090]	.613

Note. Estimates for average associations are differences in life satisfaction (0 to 10 scale) between students in schools with and without a programme, conditional on covariates and country. RATE = rank-weighted average treatment effect; AUTOC = area under the targeting operator characteristic curve. *k* = number of countries; *SE* = standard error; CI = confidence interval.

Figure 2

Conditional Association Between the School Programme and Life Satisfaction, by Cohort



Note. Points show the estimated difference in life satisfaction (0 to 10 scale) between students in schools with and without a programme; bars show 95% confidence intervals. The dashed line marks zero.

3.3. Differentiation by Student Characteristics (H2)

None of the four student characteristics was associated with the size of the conditional association (Table 3 and Figure 3). For each standard deviation increase in the characteristic, the projected change in the association was 0.037 points for ESCS (*SE* = 0.037, *p* = .310), −0.033 points for home digital access (*SE* = 0.037, *p* = .362), −0.005 points for girls (*SE* = 0.025, *p* = .838) and 0.008 points for immigrant background (*SE* = 0.025, *p* = .733). The compensation account (H2a) predicted a negative projection for ESCS and home digital access, and the cumulative advantage account (H2b) a positive one. The two estimates pointed in different directions and neither differed from zero, so neither account was supported. There was no evidence of differentiation by gender (H2c) or immigrant background (H2d). Table 9 gives 95% confidence intervals for all projections. They are wide enough to admit moderation of a size that could matter, so the absence of significant projections is not evidence that moderation is absent.

The calibration test gave no reliable evidence that the forest captured real variation in associations (differential prediction coefficient = 0.19, *p* = .085). About half of the student-level estimates were negative (52%), which is the pattern expected when estimates vary around a true value of zero. H2e, which allowed for negative associations in some groups, could therefore not be distinguished from random variation.

Table 3

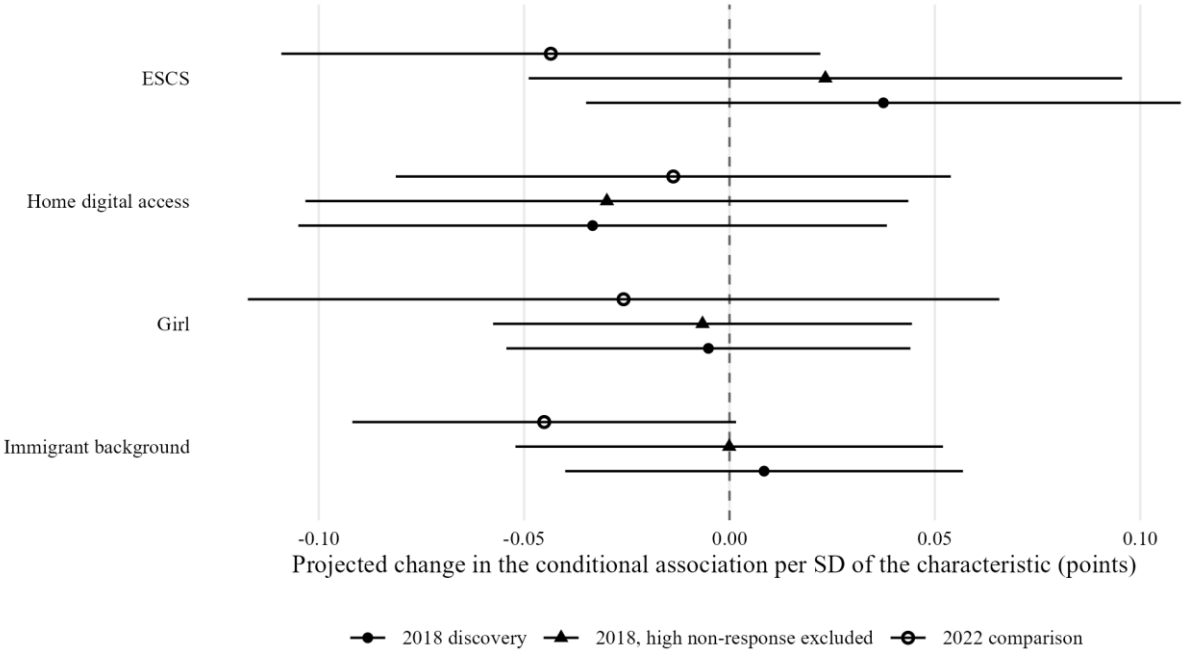
Best Linear Projection of the Conditional Association on Student Characteristics

Characteristic	*2018: b (SE)*	*2018: p*	*2022: b (SE)*	*2022: p*	*Sensitivity: b (SE)*	*Sensitivity: p*
ESCS	0.037 (0.037)	.310	−0.043 (0.033)	.194	0.023 (0.037)	.526
Home digital access	−0.033 (0.037)	.362	−0.014 (0.034)	.691	−0.030 (0.037)	.425
Girl	−0.005 (0.025)	.838	−0.026 (0.047)	.581	−0.007 (0.026)	.801
Immigrant background	0.008 (0.025)	.733	−0.045 (0.024)	.058	−0.0001 (0.027)	.998

Note. Coefficients are projected changes in the conditional association (points on the 0 to 10 life satisfaction scale) per standard deviation of each characteristic, estimated jointly. Standard errors are cluster-robust. Sensitivity = 2018 discovery cohort after excluding six countries with more than 20% non-response to the life satisfaction item. ESCS = PISA index of economic, social and cultural status. Ninety-five percent confidence intervals for all projections are given in Table 9.

Figure 3

Best Linear Projection of the Conditional Association on Student Characteristics



Note. Points show the projected change in the conditional association per standard deviation of each characteristic; bars show 95% confidence intervals.

3.4. Replication across Cohorts (H3)

Ranking the 2025 students by the associations predicted from the 2018 forest did not identify students with stronger associations in the independently estimated 2025 forest (rank-weighted average treatment effect = −0.0003, *SE* = 0.026, 95% CI [−0.052, 0.052]). A first version of this analysis, run before the gender variable was found to be missing for eight countries, used 43 countries and gave the same result (−0.0024, *SE* = 0.021). H3 was not supported. Because the discovery analysis found no reliable differentiation to begin with, this result is better read as the absence of a pattern to replicate than as evidence that an existing pattern is unstable across cohorts. The 2018 forest used for the ranking was refitted without gender and in the 56 discovery countries so that the same covariates were available in both cohorts (see Method). An independent assessment of ESCS, home digital access and immigrant background in a forest fitted to the 2025 data alone is reported in Table 9.

3.5. Proximal and Distal Outcomes (H4)

In the 2025 cohort, the programme was not associated with the two digital wellbeing items any more than with life satisfaction (Table 4 and Figure 4). In standard deviation units, the estimates were 0.004 for feeling pressured to be online (*SE* = 0.012, 95% CI [−0.019, 0.027]), 0.008 for feeling anxious without a device (*SE* = 0.012, 95% CI [−0.016, 0.032]) and 0.002 for life satisfaction (*SE* = 0.011, 95% CI [−0.019, 0.023]). H4 was not supported. The outcomes were aligned so that higher values are favourable for the direct comparison of associations reported under Robustness Analyses (Table 10).

Table 4

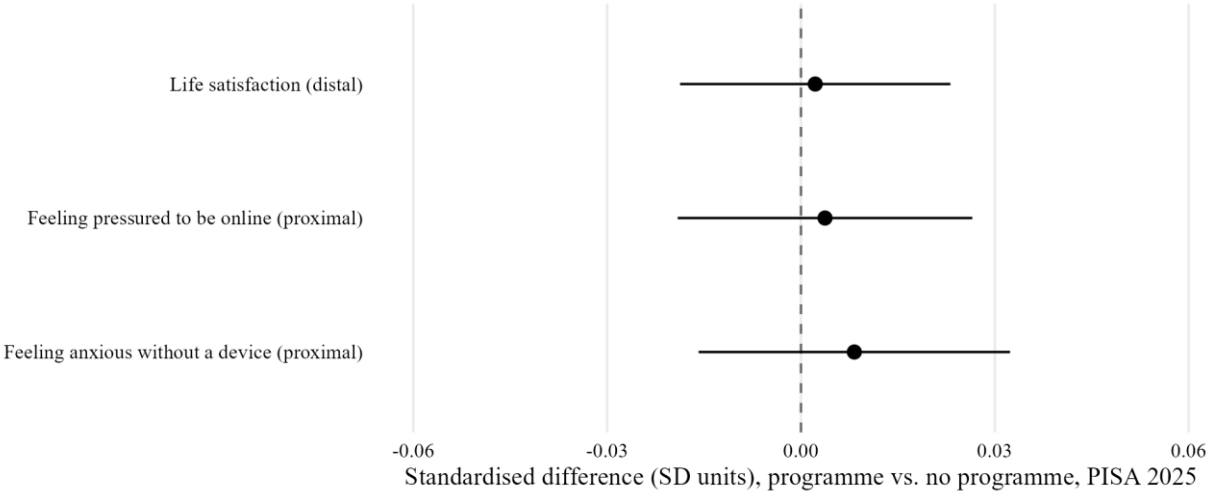
Standardised Associations of the School Programme With Proximal and Distal Outcomes, PISA 2025

Outcome	*n*	*Estimate*	*SE*	*95% CI*	*p*
Life satisfaction (distal)	331,471	0.002	0.011	[−0.019, 0.023]	.836
Feeling pressured to be online (proximal)	317,334	0.004	0.012	[−0.019, 0.027]	.749
Feeling anxious without a device (proximal)	315,049	0.008	0.012	[−0.016, 0.032]	.502

Note. Outcomes were standardised to a mean of zero and a standard deviation of one. Estimates are average differences between students in schools with and without a programme, conditional on covariates and country; *k* = 51 countries. *SE* = standard error; CI = confidence interval.

Figure 4

Standardised Associations of the School Programme With Proximal and Distal Outcomes, PISA 2025



Note. Points show standardised differences between students in schools with and without a programme; bars show 95% confidence intervals.

3.6. Sensitivity and 2022 Cohort Analyses

Excluding the six countries in which more than 20% of students gave no response to the life satisfaction item left the discovery results unchanged. In the remaining 50 countries (324,262 students), the average association was 0.018 points (*SE* = 0.036, 95% CI [−0.053, 0.090], *p* = .613), and none of the four projections differed from zero (Table 3). The 2022 cohort showed the same pattern. The average association was 0.050 points (*p* = .120), and none of the projections reached significance. The projection for immigrant background came closest (−0.045, *SE* = 0.024, *p* = .058), but it was one of twelve tests of differentiation, pointed in the opposite direction to the corresponding 2018 estimate and did not appear in the sensitivity analysis. Further sensitivity analyses of the 2025 cohort are reported under Robustness Analyses.

3.7. Robustness Analyses

The analyses in this section examine whether the conclusions depend on the estimator, the weighting of students and countries, the comparability of schools, the treatment of missing outcome data or the way the hypotheses were tested. They were added during revision, are exploratory, and use the samples of the primary analyses unless stated otherwise.

3.7.1. Survey design and country weighting

Table 5 compares the primary forest estimates with conventional estimators of the average association: a regression with country fixed effects and the covariates of the forests, estimated with the PISA final student weight and with standard errors from the 80 replicate weights; the same regression with equal weight for each country; and a regression weighted by the inverse of a school-level propensity score (see School Comparability and Overlap). Estimates are given in life satisfaction points and in standard deviations.

The conventional estimators agreed with the forests that the average association was small. With PISA weights, the regression estimates were 0.024, 0.026 and 0.041 points in 2018, 2022 and 2025 (0.009 to 0.017 standard deviations), and the 95% confidence interval included zero in every cohort (in 2025 the lower limit was −0.001). Their replicate-weight standard errors (0.021 to 0.027 points) were similar to or smaller than the school-clustered standard errors of the forests (0.025 to 0.035), so the inference of the primary analysis does not hinge on how the sampling design was handled. Giving each country the same weight lowered the 2018 and 2022 estimates (0.010 and 0.019 points) and left the 2025 estimate small but positive (0.035 points, 95% CI [0.012, 0.059], or 0.015 standard deviations). The propensity-score weighted regression gave a 2025 estimate of 0.048 points (0.020 standard deviations), and the forest with equal-country weights gave −0.015 points (*SE* = 0.016) in 2025, −0.024 (*SE* = 0.015) in 2018 and 0.010 (*SE* = 0.016) in 2022. These conventional estimates do not overturn the conclusion that the average association is small, but they do not support a strict null in 2025: an association of about 0.015 to 0.02 standard deviations is compatible with the 2025 data under several specifications. We therefore describe the 2025 result as an association that the forest analysis cannot distinguish from zero and that conventional models place slightly above zero, and we do not claim that it is exactly zero.

Table A2 in the Appendix reports the eleven interaction terms of the regression with the four student characteristics. With PISA weights, none was significant at the .05 level (smallest *p* = .052, immigrant background in 2025). With equal-country weights, six of the eleven were nominally significant, and five remained so after Benjamini and Hochberg's (1995) correction: ESCS in 2018 (0.044 points per standard deviation, *p* = .002) and in 2025 (0.027, *p* = .017), gender in 2022 (0.051, *p* < .001), home digital access in 2022 (0.036, *p* = .011) and immigrant background in 2025 (−0.021, *p* = .010). The gender term in 2018 (0.022, *p* = .036) did not survive the correction. These terms were small, at most about 0.05 points per standard deviation of the characteristic, or 0.02 standard deviations of life satisfaction. They did not agree in direction across cohorts: the ESCS term was positive in 2018 and 2025 but close to zero in 2022, and the home digital access term was positive in 2022 and negative, although not significant, in 2018 and 2025. They also appeared only when

Table 5

Average Associations Between the School Programme and Life Satisfaction Under Alternative Estimators

Estimator	*Estimate*	*SE*	*95% CI*	*Estimate (SD)*	*95% CI (SD)*
2018 (discovery): $k = 56$, $n = 355{,}825$					
Causal forest, PISA weights (primary)	0.002	0.035	[−0.067, 0.070]	0.001	[−0.026, 0.027]
Causal forest, equal-country weights	−0.024	0.015	[−0.054, 0.006]	−0.009	[−0.021, 0.002]
Regression, PISA weights	0.024	0.027	[−0.029, 0.077]	0.009	[−0.011, 0.030]
Regression, equal-country weights	0.010	0.013	[−0.015, 0.035]	0.004	[−0.006, 0.013]
Regression, propensity-score weighted	0.030	0.026	[−0.022, 0.081]	0.011	[−0.008, 0.031]
2022 (comparison): $k = 52$, $n = 348{,}364$					
Causal forest, PISA weights (primary)	0.050	0.032	[−0.013, 0.114]	0.019	[−0.005, 0.044]
Causal forest, equal-country weights	0.010	0.016	[−0.021, 0.042]	0.004	[−0.008, 0.016]
Regression, PISA weights	0.026	0.024	[−0.022, 0.074]	0.010	[−0.008, 0.028]
Regression, equal-country weights	0.019	0.012	[−0.005, 0.043]	0.007	[−0.002, 0.017]
Regression, propensity-score weighted	0.023	0.026	[−0.029, 0.075]	0.009	[−0.011, 0.029]
2025 (replication): $k = 51$, $n = 331{,}471$					
Causal forest, PISA weights (primary)	0.006	0.025	[−0.042, 0.054]	0.003	[−0.018, 0.023]
Causal forest, equal-country weights	−0.015	0.016	[−0.045, 0.016]	−0.006	[−0.019, 0.007]
Regression, PISA weights	0.041	0.021	[−0.001, 0.082]	0.017	[0.000, 0.034]
Regression, equal-country weights	0.035	0.012	[0.012, 0.059]	0.015	[0.005, 0.025]
Regression, propensity-score weighted	0.048	0.022	[0.004, 0.092]	0.020	[0.002, 0.039]

Note. Estimates are differences in life satisfaction (0 to 10 scale) between students in schools with and without a programme, conditional on covariates and country. The regressions include country fixed effects, and their standard errors (*SE*) come from the 80 PISA replicate weights. Equal-country weights give each country the same total weight. Units of standard deviation use the weighted *SD* of life satisfaction in each cohort (2.58, 2.61 and 2.38). The propensity-score weights are described in the text. k = number of countries; CI = confidence interval.

countries were weighted equally, which gives more weight to students in countries with fewer students. We therefore do not read them as support for H2a or H2b. They show instead that small differences by student characteristics cannot be excluded and that conclusions about differentiation depend on whether larger or smaller countries carry more weight.

The forests with equal-country weights, which are comparable with the primary forests except for the weighting, gave 4 projections with $p < .05$ out of 11 (home digital access in 2018: −0.030 points per standard deviation, $SE = 0.015$, $p = .048$; gender in 2018: 0.027 points per standard deviation, $SE = 0.012$, $p = .020$; home digital access in 2022: 0.029 points per standard deviation, $SE = 0.014$, $p = .042$; immigrant background in 2025: −0.027 points per standard deviation, $SE = 0.012$, $p = .022$). The calibration test of these forests gave a differential prediction coefficient of 0.36 ($SE = 0.08$, $p < .001$), 0.14 ($SE = 0.07$, $p = .032$), 0.04 ($SE = 0.09$, $p = .319$) in 2018, 2022 and 2025. The equal-country forests therefore captured some variation in associations in the first two cohorts, which the four characteristics explained only in part, whereas the calibration test of the primary 2018 forest gave no reliable evidence of such variation. We report this as an exploratory result that depends on how countries are weighted.

3.7.2. Country-specific associations

A pooled association close to zero could conceal positive associations in some countries and negative ones in others. Table 6 therefore summarises regressions estimated separately in each country (country estimates are in Appendix Table A3 and Figure A1), combined in a random-effects meta-analysis.

Table 6

Random-Effects Meta-Analysis of Country-Specific Associations Between the School Programme and Life Satisfaction

Cohort	*k*	*Mean [95% CI]*	τ	I^2 (%)	*95% prediction interval*	*Sig. + / −*	*Leave-one-out range*
2018	56	0.001 [−0.023, 0.025]	0.030	11	[−0.064, 0.065]	0 / 1	[−0.003, 0.009]
2022	52	0.014 [−0.011, 0.038]	0.042	22	[−0.072, 0.099]	3 / 2	[0.009, 0.018]
2025	51	0.018 [−0.006, 0.043]	0.040	20	[−0.063, 0.100]	3 / 1	[0.014, 0.023]

Note. Estimates are in life satisfaction points. τ = between-country standard deviation of the association; I^2 = percentage of variation due to heterogeneity between countries; Sig. + / − = number of countries with a 95% confidence interval above or below zero. The leave-one-out range is the range of the pooled estimate when each country is omitted in turn.

The mean country-specific association was 0.001 points in 2018, 0.014 in 2022 and 0.018 in 2025, and no 95% confidence interval excluded zero. The between-country standard deviation of the association (τ) was 0.030, 0.042 and 0.040 points, with I^2 of 11%, 22% and 20%, so most of the variation between country estimates was sampling error. The 95% prediction intervals, which describe where the association in a further country would be expected to fall, ran from −0.064 to 0.065 in 2018, from −0.072 to 0.099 in 2022 and from −0.063 to 0.100 in 2025, or roughly −0.03 to 0.04 standard deviations. A significant estimate in either direction was found in 1, 5 and 4 countries in the three cohorts, against about three expected by chance among 51 to 56 countries at the .05 level. Heterogeneity was clearest in 2022 (Q test $p = .044$) and absent in 2018 ($p = .193$) and 2025 ($p = .229$). Omitting any one country changed the pooled estimate very little (Table 6), and the most influential countries were Peru (2018), Macao (2022) and Brazil (2025). The near-zero pooled association is therefore not the average of large associations of opposite sign in different countries. It does not follow that the programme item means the same thing in every country, and the pooled estimate should be read as an average across potentially different national implementations.

3.7.3. School comparability and overlap

Schools do not adopt programmes at random, so we compared schools with and without a programme (Table 7).

Table 7
Standardised Differences Between Schools With and Without a Programme, Before and After Propensity-Score Weighting

Characteristic	*2018 raw*	*2018 weighted*	*2022 raw*	*2022 weighted*	*2025 raw*	*2025 weighted*
School mean ESCS	0.30	−0.01	0.11	0.02	0.04	−0.02
Private school	0.17	−0.01	0.30	0.03	0.17	−0.02
School location	0.02	−0.01	−0.05	−0.01	0.01	0.00
Mean grade	0.03	−0.03	0.01	0.03	−0.21	−0.12
Share of immigrant students	0.12	0.02	0.05	−0.03	0.08	0.02
Mean home digital access	0.14	−0.02	0.00	0.09	−0.01	0.02
Share of girls	0.03	0.03	0.00	−0.03	–	–

Note. Entries are standardised mean differences [SMD] (schools with a programme minus schools without), computed on school-level values weighted by student weights. Raw = unadjusted; weighted = within country, after inverse-propensity weighting. The propensity model included school mean ESCS, school type, school location and country; the other characteristics were not in the model. A dash means the characteristic was not available (gender was not available in the 2025 analysis).

Before adjustment, schools with a programme had higher mean ESCS (SMD = 0.30) and were more often private (0.17) in 2018, the largest difference in 2022 concerned private schools (0.30), and in 2025 students in programme schools were in lower grades (−0.21). After weighting by the inverse of the school-level propensity score, the absolute differences were below 0.03 for the characteristics in the propensity model and below 0.09 for the other characteristics, except for grade in 2025 (−0.12), which was not in the model. The share of student weight in schools with an estimated propensity outside .05 to .95 was 0.15% in 2018, 0.04% in 2022 and 0.00% in 2025, so schools with and without a programme overlapped well on the observed characteristics. The propensity-weighted regressions gave 0.030, 0.023 and 0.048 points (Table 5), close to the unweighted regression estimates. Balance on measured characteristics says nothing about unmeasured differences such as school leadership, safeguarding culture or earlier digital problems, and better adjustment would not by itself establish a causal effect. We therefore keep the interpretation of all estimates as conditional associations.

3.7.4. Missing Life Satisfaction Data in 2025

The 2025 cohort carries the replication test and had more missing life satisfaction responses than earlier cycles. In the 51 countries of the 2025 analysis, 368,662 students had valid data on the programme indicator, ESCS, immigrant background and home digital access, and 331,471 of them answered the life satisfaction item. The weighted non-response rate was 8.8%: 10.9% in schools without a programme and 7.6% in schools with one, so respondents over-represented programme schools slightly (by 0.8 percentage point). Table A4 compares respondents and non-respondents. Non-respondents had lower ESCS (SMD = −0.24), less home digital access (−0.21), lower grade (−0.21) and more often an immigrant background (0.12, positive values meaning a higher level among non-respondents), and they attended programme schools less often (−0.19). The differences in ESCS and home digital access were present both in schools with and in schools without a programme, so the missing responses are related to the observed characteristics but not in a way that is confined to one programme group.

To examine whether the 2025 results depended on these students, we repeated the 2025 analysis after excluding countries with the most missing data and after weighting respondents by the inverse of their estimated probability of responding (see Table 8). The 2025 estimates ranged from −0.002 to 0.015 points (−0.001 to 0.006 standard deviations) across the three missing-data analyses, compared with 0.006 in the primary analysis. Excluding the six countries with more than 20% no response gave 0.000 points (SE = 0.027). Excluding the eleven countries with more than 20% missing for any reason gave 0.015 (SE = 0.029). Inverse-probability weighting gave −0.002 (SE =

Table 8
Sensitivity of the 2025 Average Association to Missing Life Satisfaction Data

Analysis	*k*	*n*	*Estimate*	*SE*	*95% CI*	*Estimate (SD)*	*RATE (SE)*
All 51 countries (primary)	51	331,471	0.006	0.025	[−0.042, 0.054]	0.003	0.000 (0.026)
Six countries with > 20% no response excluded	45	306,072	0.000	0.027	[−0.053, 0.052]	0.000	−0.004 (0.022)
Eleven countries with > 20% missing for any reason excluded	40	267,843	0.015	0.029	[−0.042, 0.072]	0.006	0.019 (0.031)
All 51 countries, inverse-probability weighting for non-response	51	331,471	−0.002	0.025	[−0.052, 0.047]	−0.001	–

Note. Estimates are conditional differences in life satisfaction (0 to 10 scale) between students in schools with and without a programme, estimated with causal forests. RATE = rank-weighted average treatment effect for the ranking by the 2018 forest (area under the targeting operator characteristic curve); it was not computed for the weighted analysis. Estimate (*SD*) divides the estimate by the weighted *SD* of life satisfaction in 2025 (2.38). *k* = number of countries; *SE* = standard error; CI = confidence interval.

0.025). The ranking by the 2018 forest again failed to identify students with stronger associations (no RATE differed from zero), and none of the 9 projections on ESCS, home digital access and immigrant background in the reduced and weighted samples was significant. The results of the 2025 replication therefore do not depend on the countries with the highest non-response or on observed differences between respondents and non-respondents, although they cannot rule out bias from missingness that depends on unobserved variables.

3.7.5. Confidence intervals for the moderator projections and an independent 2025 assessment

Table 9 gives 95% confidence intervals for every projection of the conditional association on the student characteristics, including projections from a forest fitted only to the 2025 data, which assesses ESCS, home digital access and immigrant background independently of the 2018 forest.

Table 9
Projections of the Conditional Association on Student Characteristics, With 95% Confidence Intervals

Characteristic	2018	2022	2018, sensitivity	2025 (independent)
ESCS	0.037 [−0.035, 0.110]	−0.043 [−0.109, 0.022]	0.023 [−0.049, 0.096]	0.011 [−0.043, 0.065]
Home digital access	−0.033 [−0.105, 0.038]	−0.014 [−0.081, 0.054]	−0.030 [−0.103, 0.044]	−0.024 [−0.071, 0.023]
Girl	−0.005 [−0.054, 0.044]	−0.026 [−0.117, 0.066]	−0.007 [−0.058, 0.044]	–
Immigrant background	0.008 [−0.040, 0.057]	−0.045 [−0.092, 0.002]	0.000 [−0.052, 0.052]	−0.022 [−0.066, 0.021]

Note. Entries are projected changes in the conditional association (points on the 0 to 10 life satisfaction scale) per standard deviation of the characteristic, with 95% confidence intervals in brackets, estimated jointly. The sensitivity column is the 2018 cohort after excluding six countries with more than 20% non-response. The 2025 forest did not include gender, which was unavailable for eight countries; a dash marks the missing projection. ESCS = PISA index of economic, social and cultural status.

The intervals are wide enough to admit moderation of a size that could matter. In 2018, the projection for ESCS ranged from −0.035 to 0.110 points per standard deviation (−0.014 to 0.043 standard deviations of life satisfaction) and that for home digital access from −0.105 to 0.038. The absence of a significant projection should therefore not be read as proof that moderation is absent; the data exclude only changes in the association larger than about 0.11 points per standard deviation of ESCS or home digital access.

In the independent 2025 forest, the projections were 0.011 for ESCS (p = .686), −0.024 for home digital access (p = .322) and −0.022 for immigrant background (p = .314). None of the three differed from zero, and their signs for ESCS and home digital access did not match those of 2018, so H2a and H2b were not supported in the independent cohort either. The calibration test of this forest

gave a differential prediction coefficient of 0.26 (*SE* = 0.11, *p* = .008), which indicates some variation in 2025 associations that ESCS, home digital access and immigrant background did not account for.

3.7.6. Direct contrasts between outcomes (H4)

Hypothesis 4 concerns differences between associations, so we tested the differences directly. The outcomes were aligned so that higher values were favourable, which means that feeling pressured to be online and feeling anxious without a device were reversed, and were analysed in the 313,974 students with all three outcomes (see Table 10).

Table 10
Differences Between the Associations With Digital Wellbeing Outcomes and With Life Satisfaction, PISA 2025

Contrast (favourable direction positive)	*Forest: difference [95% CI]*	*p*	*Regression: difference [95% CI]*	*p*
Feeling pressured to be online minus life satisfaction	−0.009 [−0.039, 0.021]	.553	−0.009 [−0.032, 0.013]	.417
Feeling anxious without a device minus life satisfaction	−0.009 [−0.042, 0.024]	.604	−0.010 [−0.034, 0.013]	.377
Mean of the two digital wellbeing outcomes minus life satisfaction	−0.009 [−0.038, 0.020]	.550	−0.010 [−0.031, 0.011]	.352
Feeling pressured minus feeling anxious	0.000 [−0.024, 0.023]	.968	0.001 [−0.018, 0.020]	.914

Note. Entries are differences between the standardised associations of the programme with two outcomes, with 95% confidence intervals in brackets. All outcomes were standardised and aligned so that higher values are favourable (pressure and anxiety were multiplied by −1); a positive difference means that the association is more favourable for the first outcome. Forest differences come from the doubly robust scores, with variance robust to clustering of students within schools. Regression differences use country fixed effects and replicate-weight standard errors.

The association with feeling pressured to be online did not differ from the association with life satisfaction (difference = −0.009 standard deviations, 95% CI [−0.039, 0.021] in the forests), nor did the association with feeling anxious without a device (−0.009, 95% CI [−0.042, 0.024]). The mean of the two digital wellbeing outcomes differed from life satisfaction by −0.009 (95% CI [−0.038, 0.020]), and the regression contrasts gave the same conclusion (Table 10). The intervals exclude a difference larger than about 0.02 standard deviations in favour of the proximal outcomes. H4 was not supported, and the comparison is also limited by the fact that these two items are not necessarily targeted by every digital citizenship programme.

3.7.7. What the confidence intervals exclude

Table 11 expresses the average associations in standard deviations of life satisfaction and gives the largest magnitude that each 95% confidence interval excludes, together with exploratory equivalence tests. Appendix Table A5 gives the same quantities for all analyses.

In the primary forests, the largest association excluded by the 95% confidence interval was 0.027 standard deviations in 2018, 0.044 in 2022 and 0.023 in 2025, and 0.032 or less for the two digital wellbeing items. Across all 22 analyses of the average association with life satisfaction (Appendix Table A5), the largest magnitude excluded was 0.044 standard deviations (Forest 2022, primary). Equivalence within ±0.10 standard deviations was supported (TOST $p < .05$) in 22 of the 22 analyses and equivalence within ±0.05 standard deviations in 22. Because no smallest effect size of interest was set before the analyses, these tests describe what the data are compatible with and do not establish that the association is negligible.

Table 11
Largest Associations Excluded by the 95% Confidence Intervals and Exploratory Equivalence Tests

Analysis	*Estimate (SD)*	*95% CI (SD)*	*Largest magnitude excluded (SD)*	*TOST p, ±0.05 SD*	*TOST p, ±0.10 SD*
Forest 2018 (primary)	0.001	[−0.026, 0.027]	0.027	< .001	< .001
Forest 2022 (primary)	0.019	[−0.005, 0.044]	0.044	.007	< .001
Forest 2025 (primary)	0.003	[−0.018, 0.023]	0.023	< .001	< .001
Forest 2018, six high non-response countries excluded	0.007	[−0.021, 0.035]	0.035	.001	< .001
Feeling pressured to be online, 2025 (reversed)	−0.004	[−0.027, 0.019]	0.027	< .001	< .001
Feeling anxious without a device, 2025 (reversed)	−0.008	[−0.032, 0.016]	0.032	< .001	< .001

Note. Estimates are average associations between the programme and the outcome in standard deviations. The largest magnitude excluded is the larger absolute limit of the 95% confidence interval. TOST = two one-sided tests of equivalence; p is the larger of the two one-sided p values, so $p < .05$ means that the association lies within the stated bounds. The bounds were not fixed before the analyses, and the tests are exploratory. For the digital wellbeing outcomes the signs are reversed so that positive values are favourable.

4. Discussion

We asked for whom school digital citizenship programmes are associated with adolescent life satisfaction, and whether any pattern of differentiation holds across cohorts. Across three independent PISA cohorts comprising more than one million students, the answer was consistent. Students in schools whose principal reported a programme to prepare students for responsible internet behaviour did not report higher life satisfaction than comparable students in schools without one, and they did not report less pressure to be online or less anxiety without a device. This held across levels of socioeconomic status and home digital access, for girls and boys and for students with and without an immigrant background. None of the four hypotheses received support in the primary analyses. Analyses that used survey replicate weights, equal-country weights, country-specific estimates, school-level propensity weighting and alternative treatments of missing outcome data led to the same overall conclusion, with two qualifications: conventional models placed the 2025 average association slightly above zero (up to about 0.02 standard deviations), and models that weighted countries equally showed small and inconsistent differences by student characteristics.

4.1. What the Confidence Intervals Exclude

These results are not simply non-significant. How much they rule out depends on the analysis. In the primary forests, the upper limits of the 95% confidence intervals for life satisfaction corresponded to 0.03 standard deviations in 2018, 0.04 in 2022 and 0.02 in 2025, and to 0.03 or less for the two digital wellbeing items; across all analyses (Appendix Table A5) the largest magnitude excluded was 0.04 standard deviations. For comparison, even the associations between adolescents' digital technology use and wellbeing, which have attracted wide public concern, are small (Orben & Przybylski, 2019). Whether an association of the size that remains possible would matter for a school-level programme cannot be decided from these data. No smallest effect size of interest was fixed before the analyses, so we do not claim equivalence and report the equivalence tests as exploratory (Table 11). Benchmarks from other fields should not be applied unchanged (Kraft, 2020), and small associations can matter at the level of a population: Funder and Ozer (2019) describe a correlation of .05 as very small but potentially consequential when it accumulates across many people. Narrow confidence intervals also describe precision under the fitted model. They do not remove unmeasured confounding or measurement error, and they do not show that the association is the same in every school or country. The findings therefore say something

definite about what the programme indicator does not show. It does not mark a group of schools whose students are on average much more satisfied with their lives or more at ease online.

4.2. Implications for Differential Susceptibility

Differential susceptibility theory and the differential susceptibility to media effects model lead us to expect that features of the context matter more for some adolescents than for others (Belsky & Pluess, 2009; Valkenburg & Peter, 2013). We selected the characteristics to test from these models. We found no differentiation along any of them in the primary forest analysis, although conventional models and forests that weighted countries equally showed small differences by student characteristics that were inconsistent across cohorts. There are at least two ways to read this result, and the present data cannot separate them. The first reading is that the mere presence of a programme is a distal feature of the school microsystem. In bioecological terms, what matters for development are proximal processes, the regular interactions that students take part in (Bronfenbrenner & Morris, 2006). Knowing that a school has a programme says nothing about whether a given student attended it, how often, or with what quality of teaching. On this reading, the absence of differentiation reflects the absence of a proximal process in the measure, not the absence of differential susceptibility.

The second reading is that a single yes or no report from a principal cannot capture what a programme does. Programmes labelled in the same way can differ in content, intensity and duration, and principals may interpret "a specific programme" in different ways. If the item mixes intensive, well-run programmes with brief or nominal ones, any association with student outcomes would be diluted. The absence of an association with the proximal digital wellbeing items is informative here. If the programme indicator captured programmes that work on the outcomes they target but whose benefits do not extend to overall life satisfaction, we would have expected an association with feeling pressured to be online or anxious without a device. We found none, which fits the second reading better than the view that programmes help with proximal outcomes but fail to reach life satisfaction. This inference is limited, because the two digital wellbeing items are closer to digital life than life satisfaction but are not necessarily outcomes that every digital citizenship programme targets.

These results do not challenge differential susceptibility theory or the bioecological model. Both concern how individuals respond to environments they experience, and the data contain only a school-level report that a programme is available. What the results speak to is the associational relevance of a distal indicator. Susceptibility is better tested with measures of experienced environments, such as studies of instructional features whose relation to emotion and motivation runs through students' perceptions (Banaruee & Bieg, 2026; Bieg et al., 2026), and with data on individual differences in responses to digital exposure (Beyens et al., 2020). A role for self-determination processes would likewise require measures of how students experience a programme, and we do not test it.

4.3. Compensation and Cumulative Advantage

The competing predictions about socioeconomic status and home digital access both failed, and the two estimates did not even point in a consistent direction. This does not show that digital inequality is unrelated to adolescent wellbeing. Evidence that online risks reflect and amplify offline vulnerabilities concerns the experiences of students, not whether their school has a programme (Odgers & Jensen, 2020). What our results show is that the programme indicator is not a point at which such inequality becomes visible in students' life satisfaction. Whether programmes that are more intensive, or that are designed for students with fewer resources, would show the compensatory pattern is a question for studies that measure programme content. The 2018 projection for ESCS, for example, had a 95% confidence interval from −0.035 to 0.110 points per standard deviation (Table 9), and the data exclude only changes in the association larger than about 0.11 points, roughly 0.04 standard deviations of life satisfaction. In conventional models that weighted countries equally, the ESCS interaction was small and positive in 2018 and 2025, the

direction predicted by cumulative advantage, but this was not found in the forests or in 2022, and the size of the terms is too small to bear on either account.

4.4. Stability across Cohorts

We designed the study to test whether patterns of differentiation hold across cohorts, on the grounds that single-sample estimates of heterogeneity are often not reproduced. The consistency we found was a consistency of small average associations, and this consistency, more than the unsuccessful replication of differentiation, may be the stronger contribution of the study. The 2022 cohort, surveyed during the pandemic, and the 2025 cohort, surveyed as generative artificial intelligence spread, gave the same answer as the 2018 cohort, which was surveyed before either development. The share of students in schools that reported a programme was higher in 2025 than in 2018 (53.7% and 64.9%), but because the countries and measurement conditions differ across cohorts we do not read this as a trend. In the 48 countries common to all three cohorts the share was 53.8% in 2018, 63.7% in 2022 and 65.0% in 2025, so the difference does not arise from changes in the country sample, although we cannot establish that principals interpreted the item in the same way in every cycle. The replication test itself was uninformative, because there was no pattern of differentiation in 2018 to carry forward. This outcome illustrates why replication across independent samples is worth building into studies of heterogeneity. A single-sample analysis could easily have reported the marginal immigrant background estimate in 2022 as a subgroup finding. Seen alongside the other cohorts and the sensitivity analysis, it is better read as noise.

5. Limitations

Several limitations qualify these conclusions. The design is cross-sectional within each cohort, and schools choose whether to run a programme, so the estimates are conditional associations that rest on the assumption that schools with and without a programme are comparable once measured characteristics are accounted for. Unmeasured differences between schools could mask a true association as well as create a spurious one. The programme was measured with a single item reported by the principal, which carries no information about content, intensity or student participation. Life satisfaction was measured with a single item, and the digital wellbeing items were available only in 2025. The life satisfaction item was not administered in nine countries in 2018 or 2025, several of which have large immigrant populations, so the results may not generalise to these countries and the analysis of immigrant background rests on a smaller share of immigrant students than in the full PISA samples. Non-response to the life satisfaction item was about three times higher in 2025 than in earlier cycles; the sensitivity analysis did not change the results, but it cannot rule out bias from responses that are missing for reasons related to both the programme and wellbeing. The 2025 public-use file lacks the gender variable for eight countries and student records for four others, which reduced the replication sample. Finally, variables that a programme could plausibly alter, such as bullying victimisation and loneliness, were excluded as moderators to avoid bias. As a result, the question of whether programmes are associated with better outcomes for students who are bullied cannot be answered with this design.

The revision added several further limitations. First, adjusting for school type, location, mean ESCS, grade and country cannot capture school leadership, resources, safeguarding culture, earlier digital problems or school climate, which may be related both to adopting a programme and to students' wellbeing. The regression and propensity-score analyses address observed differences only, and a more flexible estimator does not remove this limitation. Second, we could not test whether the programme item has the same meaning in every country; country-specific estimates show near-zero associations in most countries, but this is not evidence of measurement equivalence. Third, life satisfaction is a broad indicator that may be conceptually distant from what programmes aim to change, such as knowledge of online safety, digital literacy, privacy practices or experiences of cyberbullying. A programme could change these outcomes without a detectable association with life satisfaction, and the two digital wellbeing items are not targets of every

programme. Fourth, the many tests of differentiation and the many robustness analyses raise the risk of chance findings in both directions; we corrected the eleven interaction terms of the regression for multiplicity but not the other analyses. Fifth, the robustness and equivalence analyses were added during revision and are exploratory. Sixth, the country samples differ across cohorts, so differences between cohorts cannot be read as change over time, and the data do not allow us to examine developmental windows.

6. Implications for Research and Practice

For researchers, the main lesson is that an indicator of whether a school has a programme is too coarse to test theories about who benefits from school-based digital education. Tests of differential susceptibility in this area will need measures of what programmes contain, how much of them students receive and how students experience them, ideally with repeated measurement of the same students. For schools and policymakers, the results do not show that digital citizenship programmes are without value. They show that having such a programme, as reported by a principal, is not in itself associated with students' life satisfaction or digital wellbeing, and that the presence of a programme should not be taken as a marker of student wellbeing. Such studies should also include outcomes that programmes target, such as digital literacy, online safety behaviour and experiences of cyberbullying, in addition to global wellbeing.

7. Conclusion

Whether a school reports having a programme for responsible internet behaviour tells us little about how satisfied its students are with their lives or how pressured they feel online. This held for students of different socioeconomic backgrounds, with different digital resources at home, for girls and boys and for students with and without an immigrant background, and it held in cohorts surveyed before, during and after the COVID-19 pandemic. Understanding whether and for whom such programmes matter will require measures of what the programmes contain and of how students experience them. The confidence intervals of the primary analyses exclude associations larger than about 0.04 standard deviations, although small positive associations of up to about 0.02 standard deviations in 2025 cannot be ruled out.

Author contributions: The authors declare that they contributed equally to the manuscript.

Data availability: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of interest: The authors declare no competing of interest.

Ethics declaration: This study used publicly available, anonymized secondary data from the Programme for International Student Assessment (PISA) 2018, 2022, and 2025. The researchers did not recruit participants, collect primary data, or access personally identifiable information. As the study involved secondary analysis of de-identified, publicly available datasets, additional ethical approval and informed consent were not required. The original PISA data collection was conducted in accordance with the ethical and data protection procedures established by the Organisation for Economic Co-operation and Development (OECD).

Funding: The authors received no financial support for this research.

References

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics, 47*(2), 1148–1178. https://doi.org/10.1214/18-AOS1709

Athey, S., & Wager, S. (2019). Estimating treatment effects with causal forests: An application. *Observational Studies, 5*(2), 37–51. https://doi.org/10.1353/obs.2019.0001

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine, 34*(28), 3661–3679. https://doi.org/10.1002/sim.6607

Banaruee, H., & Bieg, S. (2026). Teacher humour as a predictor of student emotions and motivation: The mediating role of socio-emotional and motivational factors. *Motivation and Emotion*. Advance online publication. https://doi.org/10.1007/s11031-026-10252-3

Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: Differential susceptibility to environmental influences. *Psychological Bulletin, 135*(6), 885–908. https://doi.org/10.1037/a0017376

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Beyens, I., Pouwels, J. L., van Driel, I. I., Keijsers, L., & Valkenburg, P. M. (2020). The effect of social media on well-being differs from adolescent to adolescent. *Scientific Reports, 10*, Article 10763. https://doi.org/10.1038/s41598-020-67727-7

Bieg, S., Banaruee, H., & Dresel, M. (2026). The impact of teacher humour on teaching quality and student learning: An experimental approach. *Learning and Instruction, 102*, Article 102311. https://doi.org/10.1016/j.learninstruc.2025.102311

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (6th ed., pp. 793–828). Wiley. https://doi.org/10.1002/9780470147658.chpsy0114

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*(3), 542–575. https://doi.org/10.1037/0033-2909.95.3.542

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*(3–4), 327–350. https://doi.org/10.1007/s10464-008-9165-0

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist, 48*(2), 90–101. https://doi.org/10.1037/0003-066X.48.2.90

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Gaffney, H., Farrington, D. P., Espelage, D. L., & Ttofi, M. M. (2019). Are cyberbullying intervention and prevention programs effective? A systematic and meta-analytical review. *Aggression and Violent Behavior, 45*, 134–153. https://doi.org/10.1016/j.avb.2018.07.002

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., & Crawford, A. (2002). Internet paradox revisited. *Journal of Social Issues, 58*(1), 49–74. https://doi.org/10.1111/1540-4560.00248

Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science, 1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Livingstone, S., & Helsper, E. (2010). Balancing opportunities and risks in teenagers' use of the internet: The role of online skills and internet self-efficacy. *New Media & Society, 12*(2), 309–329. https://doi.org/10.1177/1461444809342697

Odgers, C. L., & Jensen, M. R. (2020). Annual research review: Adolescent mental health in the digital age: Facts, fears, and future directions. *Journal of Child Psychology and Psychiatry, 61*(3), 336–348. https://doi.org/10.1111/jcpp.13190

OECD. (2024). *PISA 2022 technical report*. OECD Publishing. https://doi.org/10.1787/01820d6d-en

OECD. (2026a). *Bullying in education: Prevalence, impact and responses across countries* (OECD Education Working Papers No. 341). https://doi.org/10.1787/d9f8bd9f-en

OECD. (2026b). *PISA 2025 assessment and analytical framework*. https://doi.org/10.1787/86c36975-en

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour, 3*(2), 173–182. https://doi.org/10.1038/s41562-018-0506-1

Orben, A., Przybylski, A. K., Blakemore, S.-J., & Kievit, R. A. (2022). Windows of developmental sensitivity to social media. *Nature Communications, 13*, Article 1649. https://doi.org/10.1038/s41467-022-29296-3

Pluess, M., & Belsky, J. (2013). Vantage sensitivity: Individual differences in response to positive experiences. *Psychological Bulletin, 139*(4), 901–916. https://doi.org/10.1037/a0030196

Proctor, C. L., Linley, P. A., & Maltby, J. (2009). Youth life satisfaction: A review of the literature. *Journal of Happiness Studies, 10*(5), 583–630. https://doi.org/10.1007/s10902-008-9110-9

R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.5.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

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

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sarzhanova, G., & Nurgabdeshov, A. (2025). Mapping psychological well-being in education: A systematic review of key dimensions and an integrative conceptual framework. *Journal of Pedagogical Research, 9*(3), 327–349. https://doi.org/10.33902/JPR.202534832

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

Tibshirani, J., Athey, S., Sverdrup, E., & Wager, S. (2026). *grf: Generalized random forests* (Version 2.6.1) [Computer software]. https://doi.org/10.32614/CRAN.package.grf

Valkenburg, P. M., & Peter, J. (2007). Preadolescents' and adolescents' online communication and their closeness to friends. *Developmental Psychology, 43*(2), 267–277. https://doi.org/10.1037/0012-1649.43.2.267

Valkenburg, P. M., & Peter, J. (2013). The differential susceptibility to media effects model. *Journal of Communication, 63*(2), 221–243. https://doi.org/10.1111/jcom.12024

VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: Introducing the E-value. *Annals of Internal Medicine, 167*(4), 268–274. https://doi.org/10.7326/M16-2607

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association, 113*(523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839

Wickham, H., Miller, E., & Smith, D. (2025). *haven: Import and export 'SPSS', 'Stata' and 'SAS' files* (Version 2.5.5) [Computer software]. https://doi.org/10.32614/CRAN.package.haven

Yadlowsky, S., Fleming, S., Shah, N., Brunskill, E., & Wager, S. (2025). Evaluating treatment prioritization rules via rank-weighted average treatment effects. *Journal of the American Statistical Association, 120*(549), 38–51. https://doi.org/10.1080/01621459.2024.2393466

Appendix

Table A1

Countries Included in Each Analysis

Analysis	*k*	*Countries*
Discovery (2018)	56	ALB, ARG, AUT, BGR, BRA, BRN, CHE, CHL, COL, CRI, CZE, DEU, DOM, ESP, EST, FIN, FRA, GEO, GRC, HKG, HRV, HUN, IDN, IRL, ISL, ITA, JOR, JPN, KAZ, KOR, KSV, LBN, LTU, LUX, LVA, MAC, MAR, MDA, MEX, MKD, MLT, MNE, MYS, NLD, PER, PHL, POL, PRT, ROU, SAU, SRB, SVK, SVN, TAP, THA, URY
Replication (2025)	51	ALB, ARG, AUT, BGR, BRA, BRN, CHE, CHL, COL, CZE, DEU, DOM, ESP, EST, FIN, FRA, GEO, GRC, HKG, HRV, HUN, IDN, IRL, ITA, JOR, JPN, KAZ, KOR, KSV, LBN, LTU, LVA, MAR, MDA, MEX, MKD, MNE, MYS, NLD, PER, PHL, POL, PRT, ROU, SAU, SRB, SVK, SVN, TAP, THA, URY
Comparison (2022)	52	ALB, ARG, AUT, BGR, BRA, BRN, CHE, CHL, COL, CRI, CZE, DEU, DOM, ESP, EST, FIN, FRA, GEO, GRC, HKG, HRV, HUN, IDN, ISL, ITA, JOR, JPN, KAZ, KOR, KSV, LTU, LVA, MAC, MAR, MDA, MEX, MKD, MLT, MNE, MYS, NLD, PER, PHL, POL, PRT, ROU, SAU, SRB, SVK, SVN, TAP, THA, URY

Note. Countries are listed by their PISA three-letter codes. The 2022 list shows the 53 countries that met the selection rules; one of them had no students with complete data on the analysis variables, leaving 52 in the analysis.

Table A2

Interactions Between the Programme and Student Characteristics in the Regression With Country Fixed Effects

Cohort	*Characteristic*	*PISA weights: b (SE)*	*p*	*Equal-country weights: b (SE)*	*p*	*Adjusted p*
2018	ESCS	0.052 (0.033)	.114	0.044 (0.014)	.002	.011
2018	Girl	0.010 (0.020)	.632	0.022 (0.011)	.036	.066
2018	Home digital access	−0.032 (0.026)	.226	−0.011 (0.015)	.456	.511
2018	Immigrant background	0.031 (0.025)	.217	0.013 (0.012)	.275	.377
2022	ESCS	−0.012 (0.026)	.647	−0.009 (0.013)	.495	.511
2022	Girl	0.018 (0.026)	.485	0.051 (0.011)	< .001	< .001
2022	Home digital access	−0.007 (0.024)	.778	0.036 (0.014)	.011	.031
2022	Immigrant background	−0.033 (0.023)	.161	−0.006 (0.010)	.511	.511
2025	ESCS	0.002 (0.019)	.931	0.027 (0.011)	.017	.037
2025	Home digital access	−0.021 (0.019)	.273	−0.014 (0.013)	.255	.377
2025	Immigrant background	−0.036 (0.019)	.052	−0.021 (0.008)	.010	.031

Note. Coefficients are changes in the association between the programme and life satisfaction (points) per standard deviation of the characteristic, estimated jointly with replicate-weight standard errors (*SE*). Adjusted p is the *p* value for the equal-country weights model after the Benjamini–Hochberg correction for the eleven terms. ESCS = PISA index of economic, social and cultural status.

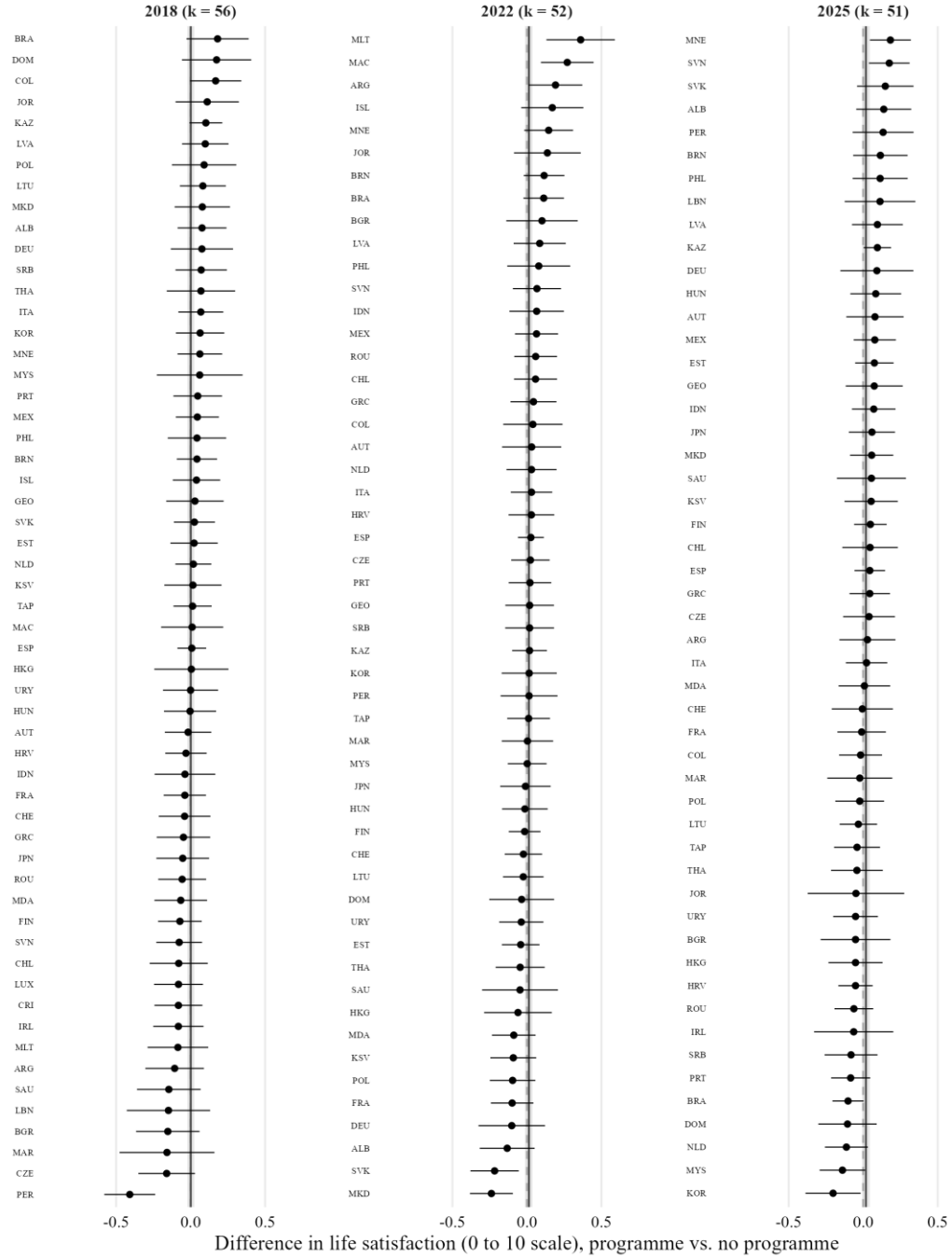
Table A3
Country-Specific Associations Between the School Programme and Life Satisfaction

Country	*2018: estimate (SE)*	*2022: estimate (SE)*	*2025: estimate (SE)*
ALB	0.077 (0.084)	−0.132 (0.094)	0.138 (0.094)
ARG	−0.107 (0.100)	0.192 (0.092)	0.028 (0.096)
AUT	−0.017 (0.079)	0.032 (0.101)	0.079 (0.098)
BGR	−0.153 (0.109)	0.101 (0.122)	−0.051 (0.119)
BRA	0.181 (0.106)	0.113 (0.069)	−0.102 (0.053)
BRN	0.042 (0.069)	0.115 (0.070)	0.115 (0.093)
CHE	−0.041 (0.088)	−0.024 (0.064)	−0.006 (0.105)
CHL	−0.080 (0.099)	0.058 (0.074)	0.046 (0.095)
COL	0.168 (0.088)	0.040 (0.101)	−0.018 (0.073)
CRI	−0.082 (0.082)	–	–
CZE	−0.161 (0.097)	0.024 (0.066)	0.039 (0.089)
DEU	0.076 (0.107)	−0.102 (0.114)	0.092 (0.125)
DOM	0.174 (0.119)	−0.035 (0.111)	−0.105 (0.100)
ESP	0.008 (0.049)	0.027 (0.044)	0.044 (0.052)
EST	0.023 (0.081)	−0.042 (0.064)	0.075 (0.066)
FIN	−0.072 (0.075)	−0.015 (0.054)	0.048 (0.056)
FRA	−0.039 (0.072)	−0.099 (0.073)	−0.010 (0.083)
GEO	0.029 (0.098)	0.018 (0.083)	0.074 (0.098)
GRC	−0.048 (0.092)	0.044 (0.079)	0.044 (0.069)
HKG	0.005 (0.127)	−0.060 (0.116)	−0.051 (0.093)
HRV	−0.031 (0.071)	0.030 (0.078)	−0.052 (0.059)
HUN	−0.004 (0.089)	−0.014 (0.078)	0.084 (0.087)
IDN	−0.038 (0.104)	0.065 (0.093)	0.071 (0.075)
IRL	−0.082 (0.086)	–	−0.064 (0.136)
ISL	0.039 (0.081)	0.170 (0.106)	–
ITA	0.068 (0.077)	0.031 (0.071)	0.023 (0.071)
JOR	0.111 (0.108)	0.137 (0.114)	−0.049 (0.165)
JPN	−0.053 (0.090)	−0.010 (0.086)	0.059 (0.079)
KAZ	0.102 (0.056)	0.018 (0.060)	0.095 (0.047)
KOR	0.064 (0.083)	0.015 (0.094)	−0.202 (0.094)
KSV	0.015 (0.098)	−0.091 (0.079)	0.053 (0.091)
LBN	−0.149 (0.143)	–	0.113 (0.121)
LTU	0.082 (0.079)	−0.024 (0.069)	−0.032 (0.064)
LUX	−0.081 (0.084)	–	–
LVA	0.098 (0.080)	0.086 (0.089)	0.095 (0.087)
MAC	0.010 (0.106)	0.271 (0.090)	–
MAR	−0.159 (0.162)	0.004 (0.088)	−0.023 (0.111)
MDA	−0.066 (0.090)	−0.088 (0.075)	0.008 (0.088)
MEX	0.045 (0.074)	0.064 (0.074)	0.078 (0.072)
MKD	0.079 (0.095)	−0.238 (0.073)	0.057 (0.074)
MLT	−0.085 (0.104)	0.361 (0.117)	–
MNE	0.062 (0.076)	0.146 (0.084)	0.183 (0.070)
MYS	0.061 (0.147)	0.002 (0.066)	−0.139 (0.078)
NLD	0.018 (0.062)	0.031 (0.086)	−0.113 (0.073)
PER	−0.409 (0.087)	0.015 (0.097)	0.134 (0.104)
PHL	0.043 (0.100)	0.079 (0.108)	0.114 (0.094)
POL	0.090 (0.110)	−0.096 (0.078)	−0.023 (0.084)
PRT	0.048 (0.083)	0.021 (0.073)	−0.084 (0.067)
ROU	−0.057 (0.082)	0.059 (0.073)	−0.063 (0.066)
SAU	−0.146 (0.109)	−0.046 (0.130)	0.055 (0.118)
SRB	0.071 (0.088)	0.018 (0.084)	−0.081 (0.091)
SVK	0.025 (0.070)	−0.217 (0.083)	0.148 (0.097)
SVN	−0.077 (0.078)	0.067 (0.082)	0.175 (0.070)
TAP	0.013 (0.065)	0.011 (0.073)	−0.041 (0.079)
THA	0.069 (0.117)	−0.045 (0.084)	−0.042 (0.089)
URY	−0.001 (0.094)	−0.038 (0.076)	−0.051 (0.077)

Note. Estimates are differences in life satisfaction (0 to 10 scale) between students in schools with and without a programme, estimated within each country with the covariates of the main analysis and replicate-weight standard errors (*SE*). A dash means the country was not in the analysis for that cohort.

Figure A1

Country-Specific Associations Between the School Programme and Life Satisfaction, by Cohort



Note. Points show the estimated difference in life satisfaction (0 to 10 scale) between students in schools with and without a programme, with 95% confidence intervals; the vertical line marks zero, and the grey band and solid line mark the 95% confidence interval and mean of the random-effects meta-analysis. Countries are ordered by estimate within each cohort; axis limits are truncated at ±0.9.

Table A4
Respondents and Non-Respondents to the Life Satisfaction Item, PISA 2025

Characteristic	*Respondents*	*Non-respondents*	*SMD*	*SMD, no programme*	*SMD, programme*
ESCS	−0.83	−1.12	−0.24	−0.32	−0.17
Immigrant background (share)	0.04	0.07	0.12	0.09	0.15
Home digital access (mean)	1.50	1.36	−0.21	−0.27	−0.18
Grade (mean)	9.75	9.60	−0.21	–	–
School mean ESCS	−0.83	−1.04	−0.25	–	–
Programme (share of students)	0.65	0.56	−0.19	–	–

Note. Entries are weighted means (ESCS, home digital access, grade, school mean ESCS, immigrant background and programme as proportions) and standardised mean differences [SMD] (non-respondents minus respondents), in the 51 countries of the 2025 analysis. SMD by programme status is given for the characteristics that could be compared within groups. ESCS = PISA index of economic, social and cultural status.

Table A5
Largest Associations Excluded by the 95% Confidence Intervals, All Analyses of the Average Association With Life Satisfaction

Analysis	*Estimate (SD)*	*95% CI (SD)*	*Largest magnitude excluded (SD)*	*TOST p, ±0.05 SD*	*TOST p, ±0.10 SD*
Forest 2018 (primary)	0.001	[−0.026, 0.027]	0.027	< .001	< .001
Forest 2022 (primary)	0.019	[−0.005, 0.044]	0.044	.007	< .001
Forest 2025 (primary)	0.003	[−0.018, 0.023]	0.023	< .001	< .001
Forest 2018, six high non-response countries excluded	0.007	[−0.021, 0.035]	0.035	.001	< .001
Regression 2018, PISA weights	0.009	[−0.011, 0.030]	0.030	< .001	< .001
Regression 2018, equal-country weights	0.004	[−0.006, 0.013]	0.013	< .001	< .001
Regression 2022, PISA weights	0.010	[−0.008, 0.028]	0.028	< .001	< .001
Regression 2022, equal-country weights	0.007	[−0.002, 0.017]	0.017	< .001	< .001
Regression 2025, PISA weights	0.017	[0.000, 0.034]	0.034	< .001	< .001
Regression 2025, equal-country weights	0.015	[0.005, 0.025]	0.025	< .001	< .001
Forest 2025, six countries excluded	0.000	[−0.022, 0.022]	0.022	< .001	< .001
Forest 2025, eleven countries excluded	0.006	[−0.018, 0.030]	0.030	< .001	< .001
Forest 2025, inverse-probability weighting	−0.001	[−0.022, 0.020]	0.022	< .001	< .001
Forest 2018, equal-country weights	−0.009	[−0.021, 0.002]	0.021	< .001	< .001
Forest 2022, equal-country weights	0.004	[−0.008, 0.016]	0.016	< .001	< .001
Forest 2025, equal-country weights	−0.006	[−0.019, 0.007]	0.019	< .001	< .001
Propensity-score weighted regression 2018	0.011	[−0.008, 0.031]	0.031	< .001	< .001
Propensity-score weighted regression 2022	0.009	[−0.011, 0.029]	0.029	< .001	< .001
Propensity-score weighted regression 2025	0.020	[0.002, 0.039]	0.039	< .001	< .001
Country-level meta-analysis 2018	0.000	[−0.009, 0.010]	0.010	< .001	< .001
Country-level meta-analysis 2022	0.005	[−0.004, 0.015]	0.015	< .001	< .001
Country-level meta-analysis 2025	0.008	[−0.003, 0.018]	0.018	< .001	< .001

Note. Estimates are in standard deviations of life satisfaction (weighted standard deviation of each cohort). The largest magnitude excluded is the larger absolute limit of the 95% confidence interval. TOST = two one-sided tests of equivalence; p is the larger of the two one-sided *p* values. The bounds were not fixed before the analyses, and the tests are exploratory.