Supplementary Material S1: Data Preparation and Analysis Code

For Whom Are School Digital Citizenship Programmes Associated With Adolescent Life Satisfaction? Person and Context Interaction Across Three Independent PISA Cohorts

This supplement describes how the analysis data were built from the PISA public-use files and how each result in the article can be reproduced. All steps were carried out in R.

S1.1 Data Sources

All data are public-use files released by the Organisation for Economic Co-operation and Development (OECD) and available from the PISA database file directory (https://webfs.oecd.org/pisa2022/index.html, which links to the 2018, 2022 and 2025 files). We used the SPSS versions of the following files.

Cycle	Student questionnaire file	School questionnaire file
PISA 2018	Student questionnaire data file (SPSS)	SPSS_SCH_QQQ.zip
PISA 2022	Student questionnaire data file (SPSS)	SCH_QQQ_SPSS.zip
PISA 2025	CY09_MS_STU_PUF.zip	CY09_MS_SCH_PUF.zip

The PISA 2015 school file (PUF_SPSS_COMBINED_CMB_SCH_QQQ.zip) was downloaded and inspected but not used, because it does not contain the programme item. The 2025 student file is compressed with the Deflate64 method, which the R functions `utils::unzip()`, `zip::unzip()` and `archive::archive_extract()` cannot read; it was extracted with the Windows file manager before being read into R. No data were altered outside R.

S1.2 Software

R 4.5.2 (R Core Team, 2025); grf 2.6.1 (Tibshirani et al., 2026); haven 2.5.5 (Wickham et al., 2025); dplyr 1.1.4; ggplot2 4.0.1; officer and flextable for the manuscript tables. The random seed for all forests was 20260923.

S1.3 Variables

Construct	2018	2022	2025	Coding
School programme for responsible internet behaviour	SC156Q05HA	SC190Q08JA	SC190Q08JA	1 = yes, 0 = no
Life satisfaction	ST016Q01NA	ST016Q01NA	ST016Q01NA	0 to 10
Feeling pressured to be online	not available	not available	ST322Q06JA	4-point agreement, standardised
Feeling anxious without a device	not available	not available	ST322Q07JA	4-point agreement, standardised
Socioeconomic status	ESCS	ESCS	ESCS	PISA index
Gender	ST004D01T	ST004D01T	ST004D01T	girl = 1
Immigrant background	IMMIG	IMMIG	IMMIG	codes 2 and 3 = 1; code 1 = 0

Computer for schoolwork at home	ST011Q04TA	ST250Q02JA	ST250Q02JA	yes = 1
Internet connection at home	ST011Q06TA	ST250Q05JA	ST250Q05JA	yes = 1
Home digital resources				sum of the two items above (0 to 2)
Grade	ST001D01T	ST001D01T	ST001D01T	as released
School type	SC013Q01TA	SC013Q01TA	SC013Q01TA	as released
School location	SC001Q01TA	SC001Q01TA	SC001Q01TA	as released
School mean ESCS				mean of ESCS within school
Final student weight	W_FSTUWT	W_FSTUWT	W_FSTUWT	sample weight
School identifier	CNTSCHID	CNTSCHID	CNTSCHID	cluster and merge key

The programme item was located in each school file by searching the variable labels for "responsible internet", and the value code for "yes" was read from the value labels rather than assumed.

S1.4 Data Preparation Steps

1. **Reading and missing codes.** Files were read with `haven::read_sav()`, which converts PISA user-defined missing codes to `NA` (95 to 99 for variables with two-digit ranges such as life satisfaction, and 5 to 9 for single-digit items). We checked the value range of every analysis variable after reading to confirm the conversion. Where the type of missingness mattered, the files were read again with `user_na = TRUE` so that "no response" (99) could be distinguished from "not administered" (97).
2. **Country selection (Gate 1).** Using the school files only, we computed for each country and cycle the number of schools answering the programme item and the proportion answering yes. A country was eligible if at least 30 schools responded and the proportion was between 0.10 and 0.90. The study was to proceed only if at least 15 countries were eligible in both 2018 and 2025. Result: 76 eligible countries in 2018, 75 in 2022 and 83 in 2025; 65 in both 2018 and 2025.
3. **Outcome availability.** Within eligible countries we computed the proportion of missing life satisfaction responses by country and cycle. Countries with 95% or more missing were treated as not having administered the item and were excluded (2018: AUS, BEL, CAN, DNK, ISR, NZL; 2022: AUS, BEL, CAN, ISR, NOR, USA; 2025: AUS, CAN, NOR, SWE, TUR, USA). This step looked only at missingness and never at outcome values by programme status.
4. **Merging.** For each cycle, student records were merged with school records by `CNTSCHID`. The merged files kept the variables in Section S1.3 and the 80 replicate weights.
5. **Recoding.** Programme = 1 if the school code equalled the "yes" code; girl = 1 if `ST004D01T` = 1; immigrant background = 1 if `IMMIG` was 2 or 3; home digital resources = number of "yes" answers to the computer and internet items; school mean ESCS = mean of ESCS among sampled students in the school.
6. **Analysis samples.** Students with missing data on the programme, life satisfaction, ESCS, immigrant background, home digital resources or, in the 2018 and 2022 analyses, gender were excluded from the model concerned. Remaining missing values in covariates (grade, school type, school location) were handled internally by grf.

7. **Non-response sensitivity.** Countries with more than 20% "no response" (code 99) to life satisfaction in 2018 or 2025 were identified (ARG, BGR, KSV, LBN, MKD, URY) and excluded in the sensitivity analysis.

S1.5 Scripts and the Results They Produce

Script	Purpose	Main outputs
`01_gate1_NOT.R`	Documents the Gate 1 rule and item codes	`gate1_ulke_kapsami.csv`, `gate1_ozet.txt`
`02_kesif_ve_tekrar.R`	2018 discovery forest (H1, H2, calibration); first 2025 replication	`sonuc_kesif_tekrar.rds`
`03_duyarlilik.R`	2022 cohort analysis, non-response sensitivity, first H4 run	`sonuc_duyarlilik.rds`
`05_H3_H4_duzeltilmis.R`	Replication (H3) and H4 with gender removed from the covariates, 51 countries	`sonuc_H3_duzeltilmis.rds`, `tablo4_H4_51ulke.csv`
`04_rapor.R`	Prints Tables 1 to 4 in the R console	console output
`07_sekiller.R`	Figures 2 to 4	`outputs/figures/*.png`
`06_makale_docx.R`	Builds the APA manuscript with all tables and figures	`Digital_Resilience_manuscript_APA.docx`
`08_ortak.R`	Shared set-up for the revision analyses (paths, country sets, helper functions)	none
`09a`	Respondents and non-respondents to the 2025 life satisfaction item, overall and by programme status	`a1_yanitsizlik_betimsel.rds`
`09b`	2025 forests after excluding countries with high non-response (sets A and B) and with inverse-probability weights for non-response	`a1_orman_A.rds`, `a1_orman_B.rds`, `a1_orman_IPW.rds`
`10_regfun.R`, `10a`	Weighted regression with country fixed effects and 80 replicate weights (PISA and equal-country weights), with and without interactions	`a2_brr_regresyon.rds`
`10b`	Country-specific associations and random-effects meta-analysis	`a4_ulke_duzeyi.rds`
`10c`	School comparisons and school-level propensity-score weighting	`a6_okul_ps.rds`
`11_h4_kontrast.R`	Direct contrasts between outcomes (H4) and the independent 2025 forest	`a5_h4_kontrast_blp25.rds`
`12_orman_esit_ulke.R`	Forests with equal-country weights	`a2_orman_esit_ulke.rds`
`13_esdegerlik.R`	Largest magnitude excluded by each interval and exploratory equivalence tests	`a3_esdegerlik.rds`

Each forest was estimated with `grf::causal_forest()` using `sample.weights = W_FSTUWT`, `clusters = CNTSCHID`, `equalize.cluster.weights = FALSE` and `num.trees = 2000`. Average associations were obtained with `average_treatment_effect()`, differentiation with `best_linear_projection()` on standardised characteristics, calibration with `test_calibration()` and replication with `rank_average_treatment_effect()` using priorities predicted from the 2018 forest.

S1.6 Reproducing the Results

Run the scripts in this order: the data preparation steps in S1.4, then `02`, `03`, `05`, `07` and `06`; the revision analyses are run with `08` to `13`. Each forest set took between 20 and 30 minutes on a Windows workstation with 32 GB of memory. Because the forests are estimated with a fixed seed, reruns reproduce the reported estimates up to the numerical tolerance of grf.

S1.7 Ethics

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