

Teen Distress Pilot

Two books offer rival explanations for the rise in adolescent anxiety and depression after 2010: Jonathan Haidt's *The Anxious Generation* (smartphones and social media) and Peter Gray's account of intensified schooling. This report asks what large-scale public microdata can say about them today, before any licensed or restricted data are requested, and whether that gets us closer to an answer than reading either book. Every figure is computed from IPUMS microdata (ACS, ATUS, MICS, IPUMS International) and public files. This is the print edition of the interactive report; the charts are static images and the data tables behind each are listed in the project archive.

The four hypotheses

The study was designed around four candidate explanations, and every test below is meant to separate at least two of them.

Hypothesis	Claim	What it predicts in these data
H1 · Phones and social media (Haidt, Twenge)	Smartphone-based social media, arriving 2010–2013, caused the rise, most of all in girls.	Distress rises first where and when phones and fast mobile data arrive; girls far more than boys; rises in phone-first countries as coverage spreads; heavier users worse off; school phone bans help.
H2 · School pressure (Gray)	Accountability testing and reduced free play made school more stressful from about 2010, and that caused the rise.	No difference by phone timing; girls somewhat more than boys; strongest in the high-stakes years (15–17) and in the school year; distress rises where and when schoolwork pressure rises; policy shocks (Race to the Top, Common Core testing) show up in state outcomes.
H3 · Interaction	Phones did the damage mainly in combination with high school demands: social media running through the same hours that carry the tests and grades, so that neither alone explains the rise and removing phones from the school day brings relief out of proportion to the hours removed.	Both pressure and phone use rise after 2010; phone-timing effects larger under high pressure; school phone bans help, and help most for students under most academic pressure; out-of-school use matters less.

Hypothesis	Claim	What it predicts in these data
H0 • Neither, or measurement	The rise is mostly changed reporting, diagnosis-seeking, or survey drift.	Adults, including the 65+ placebo comparison, drift the same way; behavioral and self-report measures disagree; no geographic or exposure level pattern at all.

Summary: what the public microdata support today

Each row is one observable from the design, grouped by the hypothesis it bears on most. Details, methods and qualifications are in the sections that follow.

What was measured	What the pilot found	Relevant to
Is there a real rise, and is the outcome usable? (H0)		
65+ placebo comparison and adult drift	The ACS cognitive-difficulty item falls monotonically for people 65+ (12.2% → 9.4%, 2008–19) and is flat at 50–64 while rising for everyone under 50. No generic response drift.	against H0
Agreement across measures	Self-report depression (NSDUH), sadness (YRBS), and the ACS item all turn up in 2011–2013 and are concentrated in girls in every source; state-level changes move together. The ACS item captures only about a sixth of the depression rise.	real rise; ACS usable but weak
Who rose most	Steepest among affluent, private-school, Hispanic, Asian and foreign-born teens; smallest among poor teens. Fits early affluent adoption and diagnosis-seeking equally.	ambiguous
Phones and social media (H1)		
Sex pattern	Girls' relative rise exceeds boys' everywhere: ACS +17% vs +9% (2008–10 → 2017–19), YRBS sadness +26% vs +19%; after 2020 girls 45–55% vs boys 16–23%.	H1 / H3
Timing in early vs late LTE markets	No timing effect. Metros that got LTE early and late track each other through 2013; the event study is 0.02 pp (SE 0.15) on 279 dated metros, –0.04 (0.19) for girls; no effect again at county level in six state surveys.	not H1 via rollout timing
U.S. adoption exposure level	Teen-household smartphone-internet share predicts YRBS girls' sadness within states (+0.9 pp per 10 points, SE 0.44) but not at county level (+0.1, SE 0.1). Suggestive at most.	weak H1

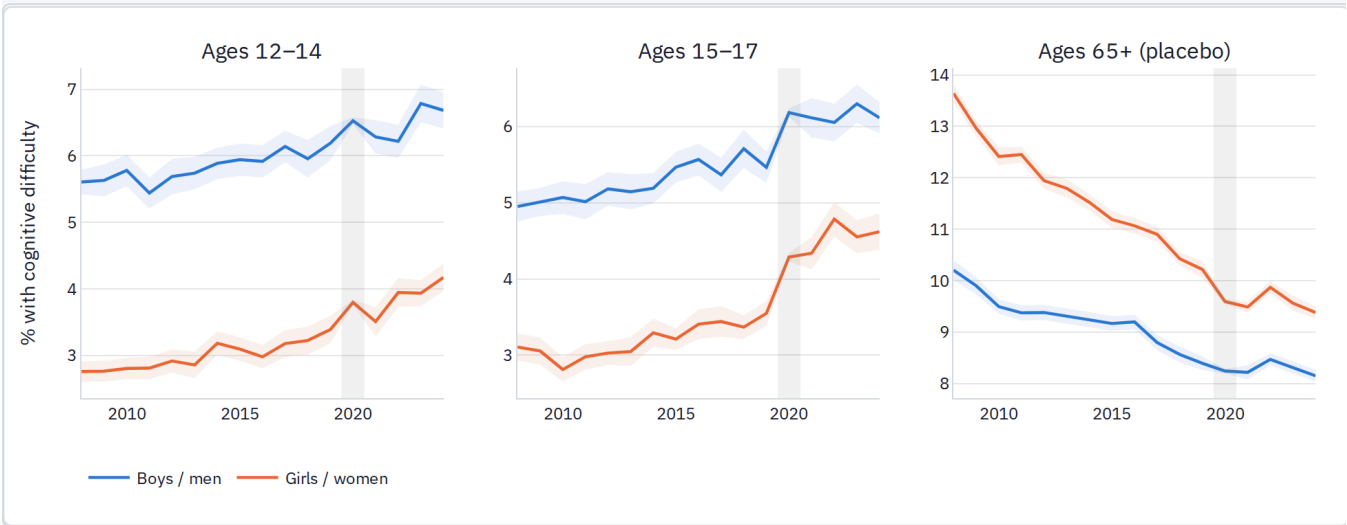
What was measured	What the pilot found	Relevant to
Phone-first countries, use-based exposure	MICS (68 surveys): caregiver-reported anxiety and depression show no relation to cluster connectivity or a girl's own daily phone/internet use; the one positive coefficient is matched by the 5–9-year-olds (a placebo comparison, since they are not the exposed group). GSHS (39 countries): no relation to internet penetration, mobile-broadband subscriptions or 3G/4G coverage. Census cognitive item uninformative.	not H1 on adult- or cluster-measured use
Self-reported phone hours	GSHS 2022–23 (five countries): girls on a phone 3+ h/day report loneliness +8 pp and suicidal ideation +10 pp, rising with hours; boys about half. Nepal 2022 PHQ-9/GAD-7 point the same way for 15–19-year-olds. One-time self-report, so the direction of cause is open.	H1 pattern, not timing
School pressure (H2)		
Age pattern and seasonality	Through 2019 the 12–14 and 15–17 rises are similar and 18–24 rises more, so no high-stakes-years signature. Seasonality is not testable in public ACS. ATUS: homework on school weekdays up about 40% (52 → 71 min) between 2003–08 and 2015–19, sleep flat, in-person socializing down.	school demand rose (H2 premise)
HBSC schoolwork pressure	Pressure and psychosomatic complaints both rose after 2009/10 across ~45 countries, and complaints rose more where pressure rose more (slope 0.14–0.24). But the slope is the same in the 2002–2010 placebo comparison period and most of the rise is common to all countries.	H2 partly pressure is a real correlate, not the break
State accountability policy	Waivers, Common Core testing, teacher evaluation and exit exams show no relation to ACS, YRBS or NSDUH outcomes; the apparent Race to the Top effect disappears when traced year by year.	not H2 via U.S. policy shocks
Interaction (H3)		
LTE × school pressure	Early-LTE markets in high-pressure states did not rise more; where the comparison can be made, the interaction is negative for girls.	not H3 via state policy
School phone bans	PISA 2022, 80 systems: ban schools have 12 pp less daily in-school use; girls report 1.4 pp less loneliness and 2 pp less device anxiety, +0.02–0.03 SD belonging and life satisfaction; 3.5 pp less reported classroom distraction. Cross-school comparison, so selection cannot be ruled out. U.S. statewide bans (2024–26) not yet testable. Nationally, school prohibitions fell 91% → 66% during 2009–16 and then re-tightened while sadness first rose and then fell.	H1 / H3 direction, small; U.S. test pending
Bans × academic pressure	PISA 2022: the ban's association with less loneliness is larger for students with high schoolwork anxiety (girls –1.2 vs –0.3 pp; boys –0.8 vs +1.1), but the difference is within its standard error and the pressure measure is a proxy.	H3 direction, not established

Part 1 • The rise, and whether public data can measure it

Before either hypothesis can be tested, the outcome has to be established: is there a rise, is it real rather than a change in reporting, and which public measures track it. This part uses the American Community Survey, the national depression and sadness surveys, and the state youth surveys.

The ACS cognitive-difficulty item, 2008–2024

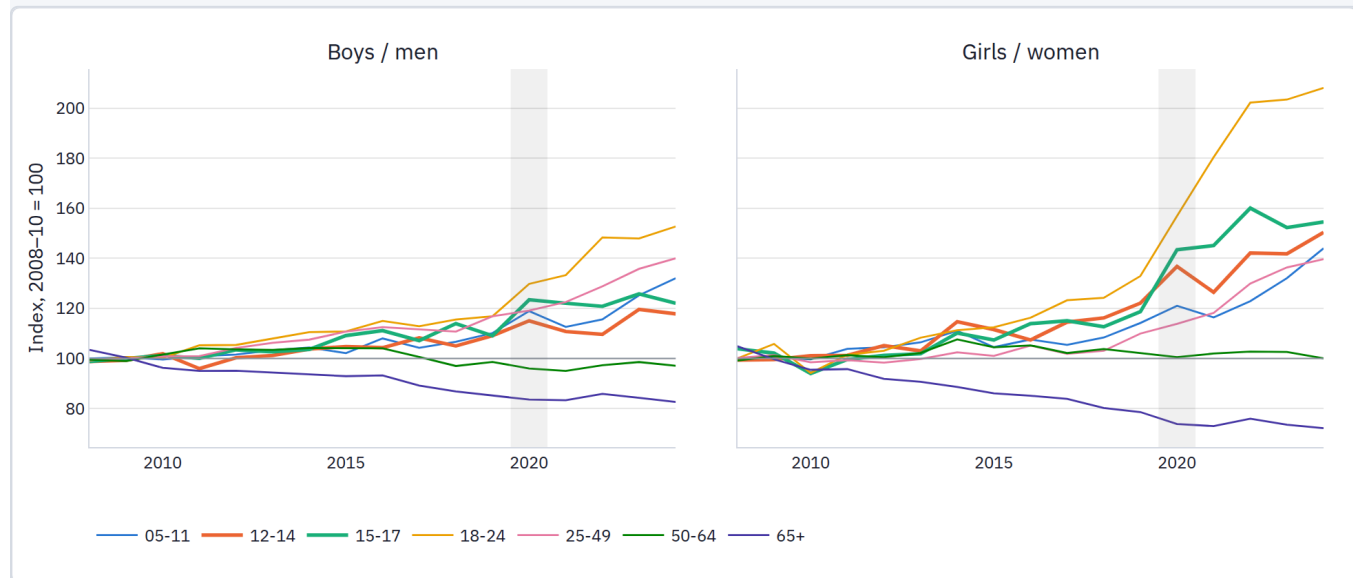
The American Community Survey asks, of everyone 5 and older, whether the person has "serious difficulty concentrating, remembering, or making decisions" because of a physical, mental, or emotional condition. The question has been stable since 2008 (2005–07 used different wording and is excluded). Each year samples about 240,000 12–17-year-olds. Standard errors use the 80 successive-difference replicate weights.



Shaded bands are 95% confidence intervals. The grey band marks the 2020 experimental ACS release, which had severe nonresponse problems.

Age	Sex	2008–10 %	2017–19 %	Rel. change	2022–24 %	Rel. change
12-14	boys	5.67	6.10	+7%	6.56	+16%
12-14	girls	2.78	3.27	+18%	4.02	+45%
15-17	boys	5.01	5.51	+10%	6.16	+23%
15-17	girls	2.99	3.45	+15%	4.65	+56%

Age gradient



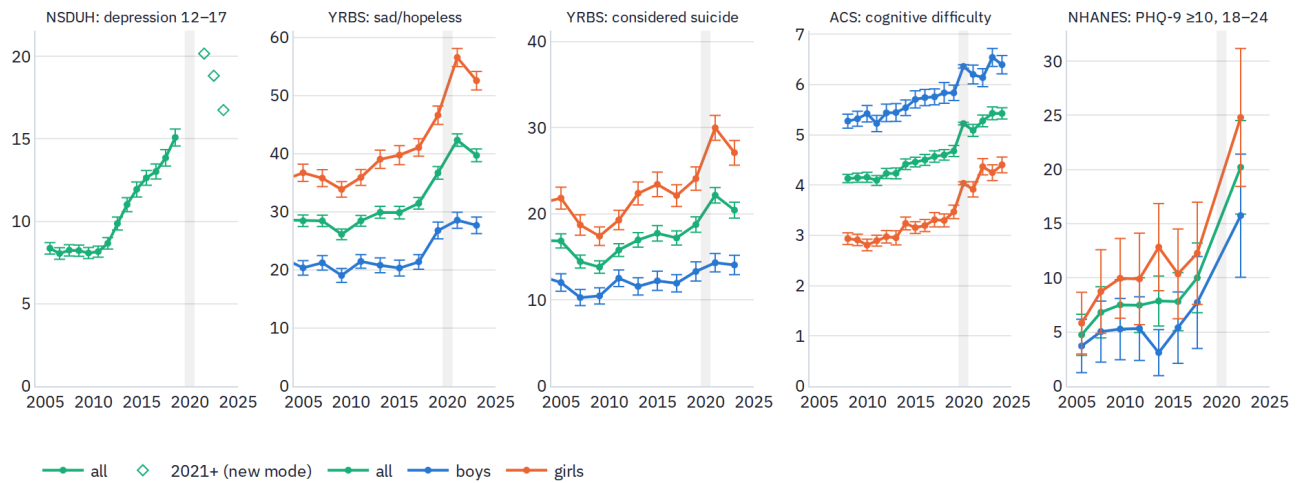
Each line is the group's rate divided by its own 2008–10 mean. Thick lines are the adolescent groups.

Two things stand out. The 2010s rise is real but modest in this item (about half a percentage point on a base of 3–5%), and it is shared by every age under 50, with 18–24 rising most. The post-2020 acceleration is far larger and is concentrated in girls and young women. The item is a disability screener, so it will understate mood disorders that do not impair concentration; the next section tests whether it moves with the depression measures at all.

Does the ACS item measure adolescent distress?

Three checks: (a) does the ACS item move with a real depression measure at the state level, (b) is its rise concentrated in girls, (c) does its national timing match the survey series. The public NHANES depression screener covers ages 18+ only (the adolescent file is restricted), so NHANES enters as an 18–24 check.

National timing



All series in percent, with 95% CIs. NSDUH points sit at the midpoint of each two-year period; the 2021+ NSDUH estimates use a new web mode and DSM-5 wording and are not comparable to earlier years. YRBS sad/hopeless was first asked in 1999.

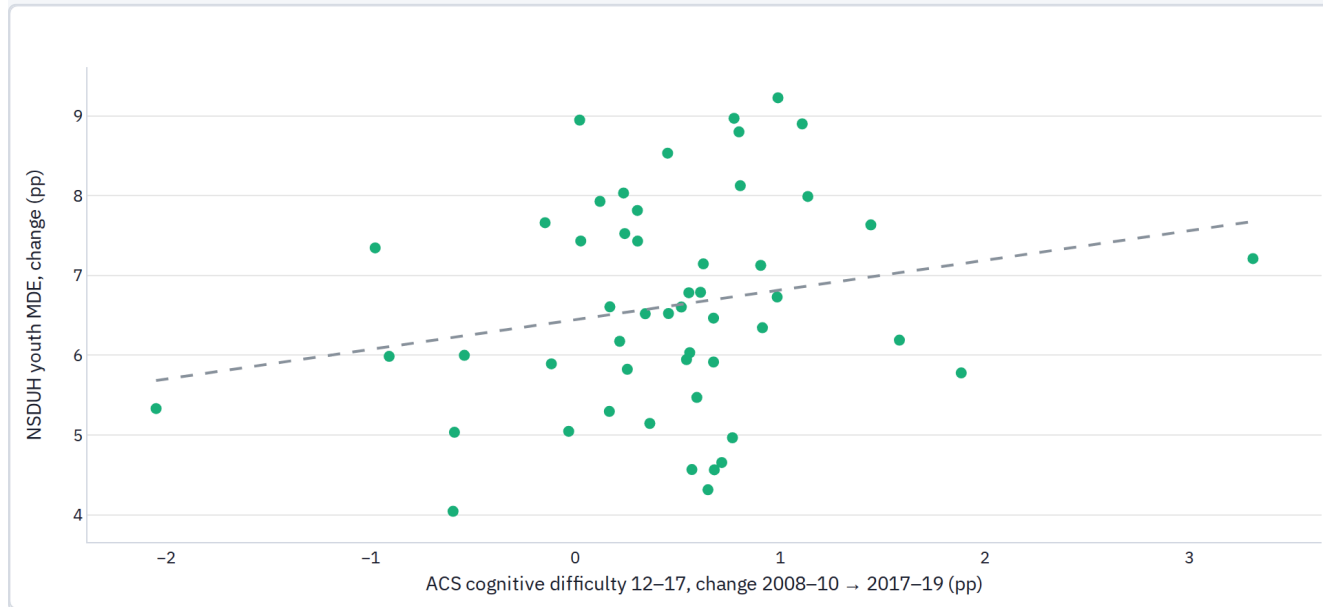
NSDUH youth depression was flat at about 8% from 2005 through 2010–11 and then rose almost every period to 15% in 2018–19. YRBS sadness was flat from 1999 to 2013, then rose sharply in 2017–2019. The ACS item rises gradually from 2012 and accelerates after 2019. The ACS series turns up at roughly the right time but its amplitude is a fraction of NSDUH's: about +13% relative versus +85%.

Sex pattern (check b)

Series	Sex	2009–11	2017–19	Relative change
ACS cognitive difficulty 12–17	boys	5.3	5.8	+9%
ACS cognitive difficulty 12–17	girls	2.9	3.4	+17%
YRBS sad/hopeless	boys	20.3	24.1	+19%
YRBS sad/hopeless	girls	34.9	43.9	+26%
YRBS considered suicide	boys	11.5	12.6	+10%
YRBS considered suicide	girls	18.3	23.1	+26%
NHANES PHQ-9 ≥10, 18–24	boys	5.3	7.7	+46%
NHANES PHQ-9 ≥10, 18–24	girls	9.9	12.3	+24%

In every source the girls' relative rise exceeds the boys'. The ACS item passes this check: if it were mainly picking up ADHD diagnoses, the boys' change would dominate.

Do states move together? (check a)



51 states. Each point is the state's change in both measures between 2008–10 and 2017–19; dashed line is the OLS fit (slope 0.37, SE 0.18, $r = 0.22$).

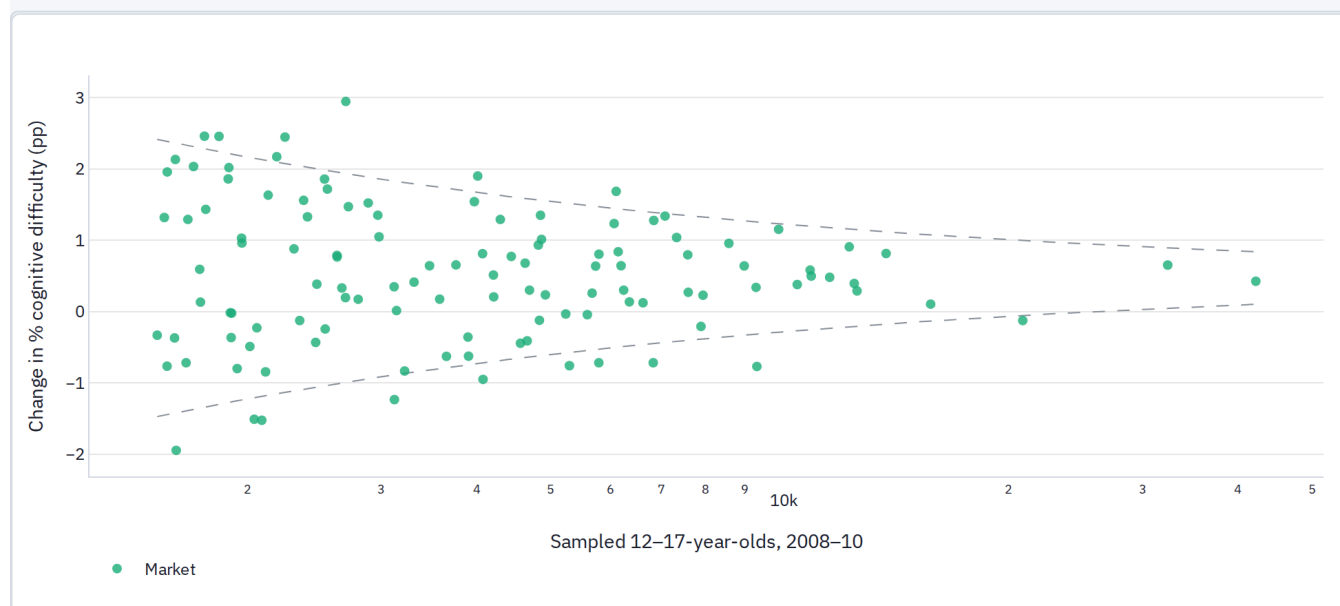
Across states there is no relation between the levels of the two measures (the ACS item is a disability screener, NSDUH is a diagnostic interview), but their *changes* are related: a state-and-period fixed-effects regression of NSDUH youth MDE on the ACS rate gives a slope of 0.14 (SE 0.05, $p = 0.01$) across 714 state-by-period observations, and the decade-long changes correlate at 0.22 (slope 0.37, SE 0.18). Against the YRBS state files (39 states, by sex), the ACS rate moves together with "sad or hopeless" at a slope of 0.35 (SE 0.18) with state-by-sex and year-by-sex effects, and not at all with "considered suicide."

Conclusion. The ACS item passes the sex check and moves, weakly but genuinely, with depression within states. It can serve as one main outcome, with two caveats: it captures perhaps a sixth of the depression rise, and the part of its change that is specific to girls is small relative to sampling error. The state youth surveys (the YRBS state files here, and the county surveys later in the report) deserve equal standing rather than a supporting role.

Geographic variation: is there anything to explain?

For a rollout design to work, the change among 12–17-year-olds has to differ across places by more than sampling noise. Using metropolitan areas (plus each state's non-metro remainder) as the unit, the average change from 2008–10 to 2017–19 is about half a percentage point, the true spread across metros is about the same size, and the sampling error for a typical metro is a bit larger still. In other words, real differences between places exist, but no single metro's change can be trusted

on its own; the metros have to be analysed together. (Statistics: weighted mean +0.47 pp, between-metro SD 0.46 pp, median sampling error 0.64 pp, Cochran's Q 218 on 116 df, $I^2 \approx 0.47$.)



117 markets with at least 1,500 sampled 12-17-year-olds in 2008-10. Dashed lines are the ± 1.96 SE funnel around the pooled mean for a market of that size.

States, by sex

At the state level the boys' change varies a lot from state to state, but the girls' change is nearly the same everywhere ($I^2 \approx 0.46$ for boys versus 0.05 for girls). Since the girls' component is the one that tracks distress, there is little geographic variation for a U.S. design to work with, which is the main weakness of the U.S. part of the study.



Who is driving the rise

Relative changes by subgroup, ages 12-17. The steepest rises are among the most advantaged teens.

By family income relative to poverty line

Group	Sex	2008-10 %	2017-19 %	Rel. change to 2017-19	Rel. change to 2022-24	Share of 12-17, 2008-10
100-199%	boys	6.67	7.05	+6%	+15%	11%
100-199%	girls	3.50	4.30	+23%	+46%	10%

Group	Sex	2008–10 %	2017–19 %	Rel. change to 2017–19	Rel. change to 2022–24	Share of 12–17, 2008–10
200–399%	boys	4.55	5.17	+14%	+31%	16%
200–399%	girls	2.38	2.91	+22%	+78%	16%
400%+	boys	2.95	3.73	+26%	+48%	16%
400%+	girls	1.55	2.04	+32%	+89%	15%
poor <100%	boys	8.61	9.02	+5%	+11%	8%
poor <100%	girls	4.88	5.45	+12%	+36%	8%

By school type

Group	Sex	2008–10 %	2017–19 %	Rel. change to 2017–19	Rel. change to 2022–24	Share of 12–17, 2008–10
not in school	boys	8.32	8.54	+3%	+15%	1%
not in school	girls	6.21	5.82	-6%	+25%	1%
private	boys	4.53	5.91	+30%	+44%	5%
private	girls	2.57	3.61	+40%	+70%	5%
public	boys	5.33	5.71	+7%	+17%	45%
public	girls	2.82	3.26	+16%	+50%	43%

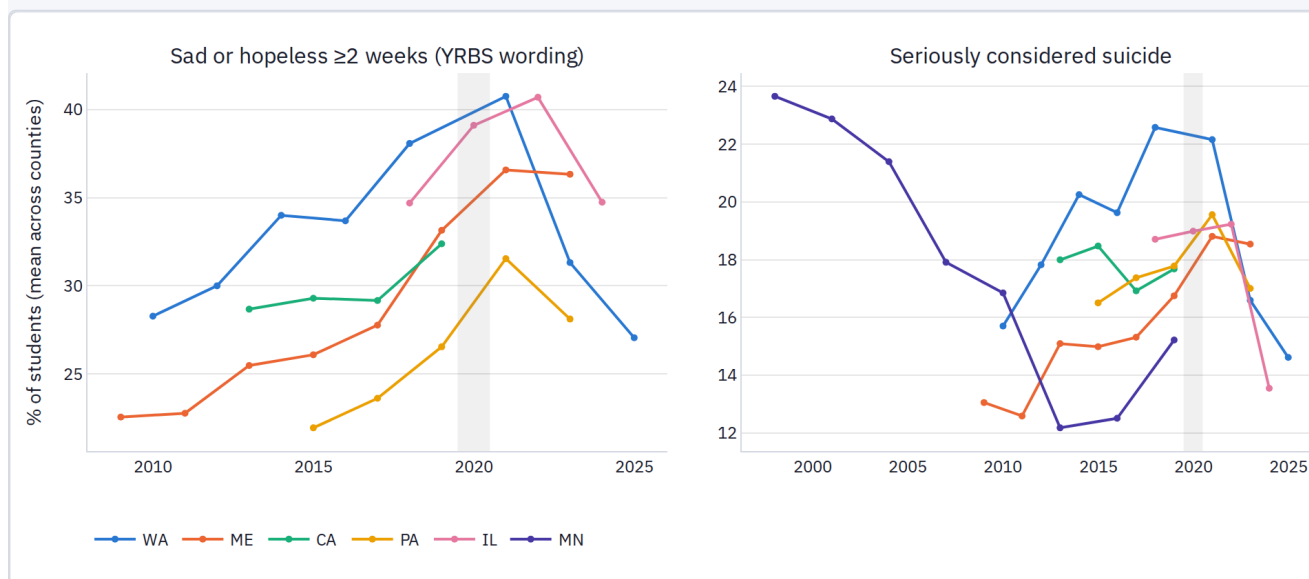
By race / ethnicity

Group	Sex	2008–10 %	2017–19 %	Rel. change to 2017–19	Rel. change to 2022–24	Share of 12–17, 2008–10
Asian-PI NH	boys	1.66	2.24	+35%	+90%	2%
Asian-PI NH	girls	1.10	1.47	+33%	+107%	2%
Black NH	boys	6.67	7.14	+7%	+13%	7%
Black NH	girls	3.69	3.94	+7%	+26%	7%
Hispanic	boys	4.09	5.16	+26%	+48%	10%
Hispanic	girls	2.37	2.95	+24%	+74%	10%

Group	Sex	2008–10 %	2017–19 %	Rel. change to 2017–19	Rel. change to 2022–24	Share of 12–17, 2008–10
Other-multi NH	boys	7.64	7.96	+4%	-8%	2%
Other-multi NH	girls	4.50	4.73	+5%	+7%	2%
White NH	boys	5.54	5.88	+6%	+12%	29%
White NH	girls	2.87	3.44	+20%	+51%	28%

State youth surveys at county level

The state youth surveys are the self-report outcome the ACS lacks. County tables were compiled for six states: Minnesota (1998–2019, 87 counties), Washington (2010–2025, 38), Maine (2009–2023, 16), California (2013–2019, 54, by sex), Pennsylvania (2015–2023, 60) and Illinois (2018–2024, 78), 34,439 county rows with source and wording recorded for each. Minnesota's pre-2013 sadness item has different wording and is excluded from the sadness series.



Unweighted county means over the grades each state reports in every survey year (Washington 8, 10 and 12; Maine's and Pennsylvania's all-grade rows; Minnesota grade 9; California 7, 9 and 11; Illinois 8, 10 and 12). Minnesota's 2022 county tables sit behind a CAPTCHA and are not yet included.

Three things stand out. Sadness rose in every state from about 2013 to a 2021 peak (Washington 28% → 41%, Maine 22% → 37%, Illinois 35% → 41%). Minnesota's suicidal-ideation series (grade 9), the only one that reaches back to the 1990s, *fell* from 24% in 1998 to 12% in 2013 before turning up. And the post-2021 declines are large: Washington's 2025 sadness rate (27%) is below its 2010 level and Illinois' 2024 ideation rate (13.5%) is the lowest in its series.

Part 2 · Phones and social media (H1)

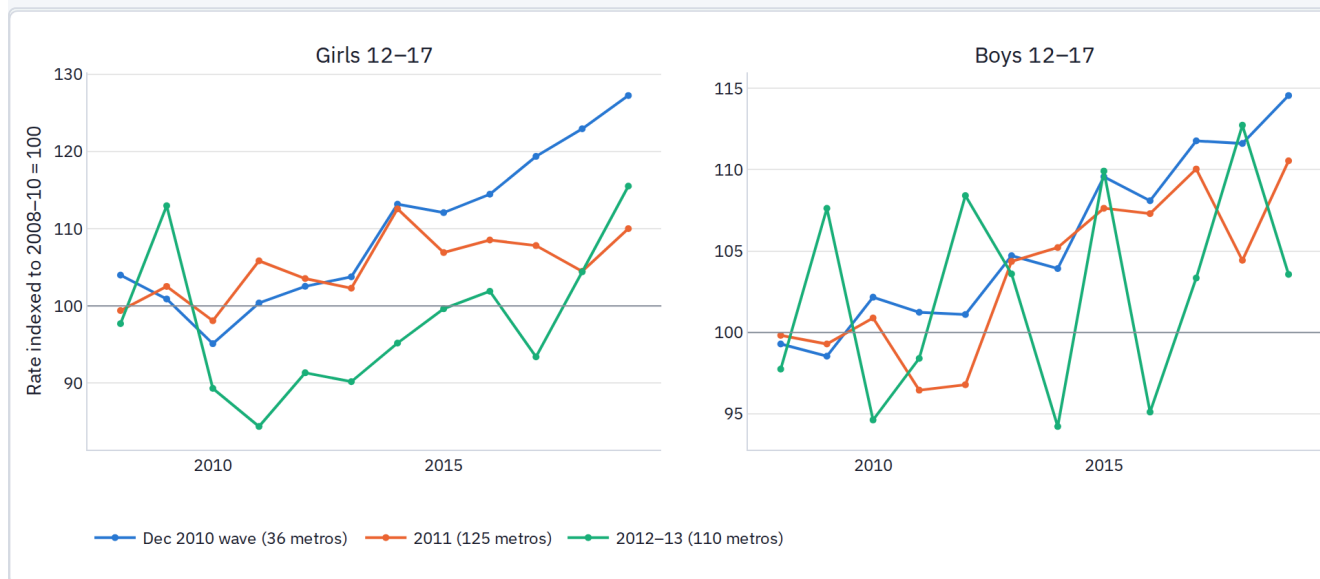
The phone hypothesis predicts that distress rose first where and when smartphones and mobile data arrived, that heavier users are worse off, and that the same pattern appears in countries that went straight to phones. The tests run from the coarsest exposure (when a metro got 4G) to the finest (an adolescent's own reported hours).

4G rollout across U.S. metros

The central test of the phone hypothesis: if fast mobile data mattered, teenagers in metros that got 4G first should have turned worse first.

LTE launch dates were compiled from carrier press releases and their press reprints: 967 carrier-market launches (Verizon 442, AT&T 243, Sprint 234, T-Mobile 47), mapped to 2013 CBSA codes, with a first-launch date for 98 of the top 100 metros and for 279 of the 311 metros the ACS identifies. Verizon's December 2010 launch covered 36 of these metros; 125 more launched in 2011, 96 in 2012, 14 in 2013. Onset is assigned to the school year that starts after the launch. Coverage gaps (mainly Verizon's 2013 small-market waves and about 44 AT&T dates among the top 150) are listed in the methods note.

Descriptive: onset cohorts

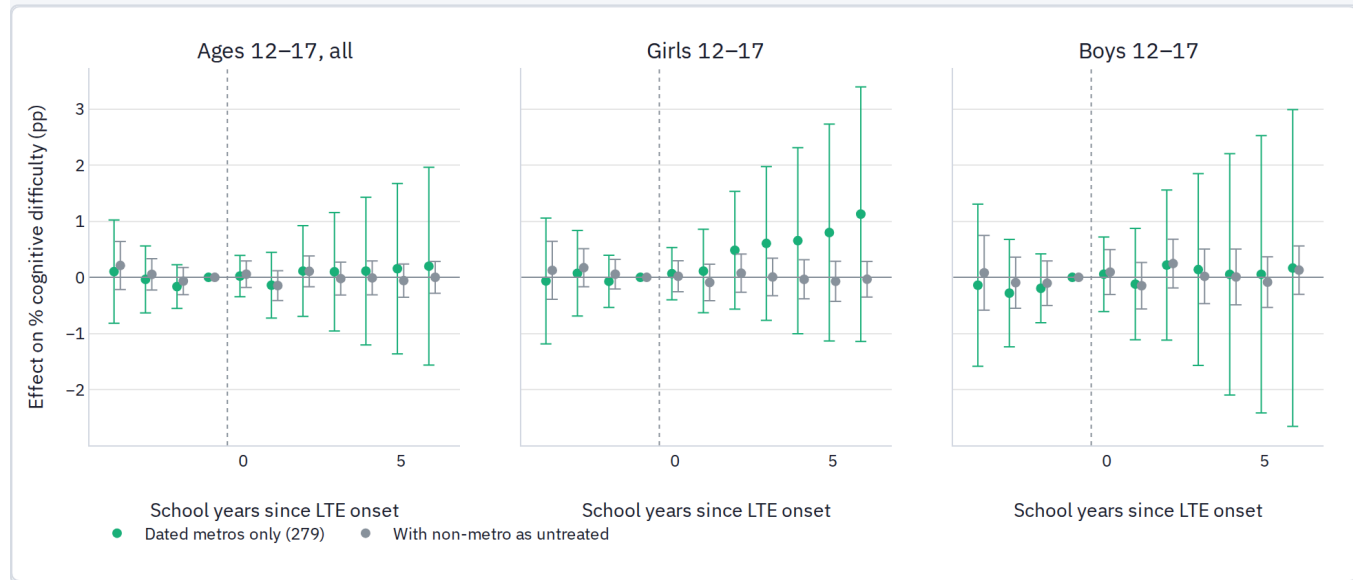


Population-weighted ACS cognitive-difficulty rate for 12-17-year-olds in metros grouped by first LTE launch, indexed to each cohort's 2008-10 mean.

The simplest check: metros that got LTE first should have pulled ahead of the later ones around 2011-2013. Through 2013 it does not; the cohorts track each other. From 2014 the December 2010 cohort does drift higher for girls (index 127 by 2019 versus 110-116), but this cohort is the 36 largest metros, where the affluent-teen composition effect is strongest, and the gap opens three

years after onset. The event-study coefficients for girls at four to six years after onset are +0.6 to +1.1 pp with standard errors near 1 pp.

Event study



Estimated effect by school year relative to the LTE launch (the year before launch is the reference), market and year fixed effects, population weights, standard errors clustered by market. Green: the 279 metros with a known launch date, so only the timing differences are used. Grey: also includes the 32 metros without a date and the non-metro parts of each state as if never treated (they are really late adopters, so this version is an outer bound).

Onset definition	Sample	12-17 all	Girls	Boys	65+ placebo comparison	25-49
First carrier	dated metros	+0.02 (0.15)	-0.04 (0.19)	+0.03 (0.27)	+0.03 (0.15)	+0.11 (0.08)
Verizon	dated metros	+0.01 (0.16)	-0.11 (0.20)	+0.05 (0.29)	+0.07 (0.16)	+0.11 (0.08)
Second carrier	dated metros	-0.01 (0.15)	-0.11 (0.17)	+0.20 (0.23)	+0.28 (0.14)	+0.09 (0.07)
First carrier	+ non-metro as untreated	0.00 (0.09)	-0.08 (0.10)	+0.11 (0.13)	+0.12 (0.11)	-0.10 (0.06)

Static difference-in-differences coefficient on "LTE in force", percentage points, SE in parentheses. Random-onset placebo comparison runs give SE $\approx$ 0.07 for the full panel.

What this shows. LTE timing across U.S. metros does not predict the timing of the rise in the ACS item. The confidence interval for girls excludes effects larger than about +0.3 pp on the dated sample and +0.1 pp with the untreated controls, against a total 2010–2019 rise of about 0.5 pp. Three things keep this from being a rejection of H1: LTE timing is compressed into 2011–2012 for most large metros, so there is little pre-period for the early cohort; social-media exposure travels through peers and national platforms, which pushes a market-level design toward zero; and the ACS item captures only part of the distress rise. It does mean the rollout design cannot get anywhere with the ACS alone: it needs the state youth surveys or clinical outcomes at the metro level. The one rollout-style study that reaches into the Instagram years, Elaine Guo's comparison of British Columbia mountain communities with and without high-speed wireless internet, does find a large effect for girls on a clinical outcome; the reconciliation section sets the two designs side by side.

Rollout and adoption at county level, state youth surveys

LTE onset at county level

Counties mapped to 2013 CBSAs and to the LTE first-launch date. The unit is a county-by-grade series, with county-by-grade and state-by-year fixed effects, so a state changing which grades it surveys cannot register as a change in distress; non-metro and undated counties as untreated (a bound, since rural LTE arrived 2013–14); standard errors clustered by metro area (by county outside metro areas). In the year-by-year version the three years before onset are the reference, because the surveys are biennial or triennial and no single pre-onset year is observed for every metro.

Panel	DiD on "LTE in force" (pp)
Sad/hopeless, WA + ME counties, 2009-2019 (54 counties, 26 dated-metro, 130 county-grade series)	+0.64 (SE 0.87)
Considered suicide, WA + ME + MN counties, 1998-2019 (141 counties, 59 dated-metro, 217 county-grade series)	-0.72 (SE 0.43)
Considered suicide, MN counties, 1998-2019 (87 counties, 33 dated-metro, 87 county-grade series)	-1.82 (SE 0.67)

Smartphone adoption at the metro level

Outcome	Sex	Effect per +10 pp teen-household smartphone internet (SE)	Counties / CBSAs	States
sad hopeless	all	+0.13 (0.13)	89 / 42	CA, IL, ME, PA, WA
sad hopeless	female	+0.22 (0.25)	32 / 21	CA, ME
sad hopeless	male	-0.11 (0.14)	32 / 21	CA, ME
considered suicide	all	+0.06 (0.15)	105 / 46	CA, IL, ME, MN, PA, WA
considered suicide	female	-0.04 (0.33)	31 / 21	CA, ME
considered suicide	male	-0.39 (0.13)	31 / 21	CA, ME

NTIA CBSA shares interpolated between waves (only areas with 30 or more respondents), county-by-grade and state-by-year fixed effects, SE clustered by CBSA, 2011–2019.

What this shows. With the self-report outcome measured at county level, neither LTE onset nor smartphone adoption predicts where sadness or suicidal thoughts rose. The one coefficient beyond two standard errors, Minnesota's suicidal ideation after LTE (–1.8 pp, SE 0.7), has the wrong sign for the phone hypothesis, and its pre-onset coefficients lean the same way, so it reads as the metro counties' earlier decline continuing rather than as an effect of LTE in either direction. The state-level result for girls reported above does not reappear here (+0.2 pp per 10 points, SE 0.3, in the two states with county data by sex), so it should be read as suggestive at best. The rise was broad, happened at the same time in counties with very different connectivity histories, and, in the latest waves, is reversing.

Adoption rather than rollout: smartphone use against the state youth survey

If rollout timing is too crude a measure, the next step is how many teenagers actually had smartphone internet. Using the rebuilt NTIA series (share of persons in households with a 12–17-year-old who use a smartphone for internet, by state, interpolated between the 2011–2019 waves) against the YRBS state files, with state and year fixed effects, 39 states, 2011–2019:

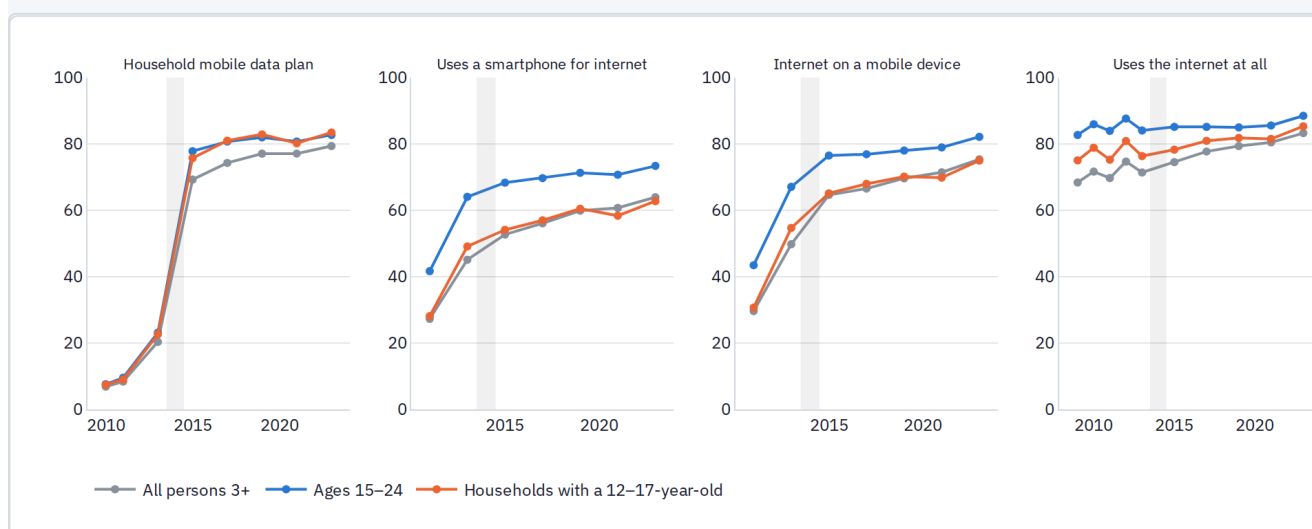
Outcome	Measure of adoption	Effect per +10 pp of adoption (SE)
YRBS sad/hopeless, girls	teen-household smartphone internet	+0.90 (0.44)
YRBS sad/hopeless, boys	same	+0.28 (0.41)

Outcome	Measure of adoption	Effect per +10 pp of adoption (SE)
YRBS sad/hopeless, girls	household mobile data plan	+0.50 (0.51)
ACS cognitive difficulty 12–17, girls	teen-household smartphone internet	–0.04 (0.07)

This has the pattern the phone hypothesis predicts (girls, not boys; the self-report sadness item, not the ACS disability item), and it is fragile: five biennial YRBS waves, an interpolated exposure level, and a p-value near 0.04 with 39 clusters. The county-level version in the next subsection does not reproduce it, so it stays suggestive.

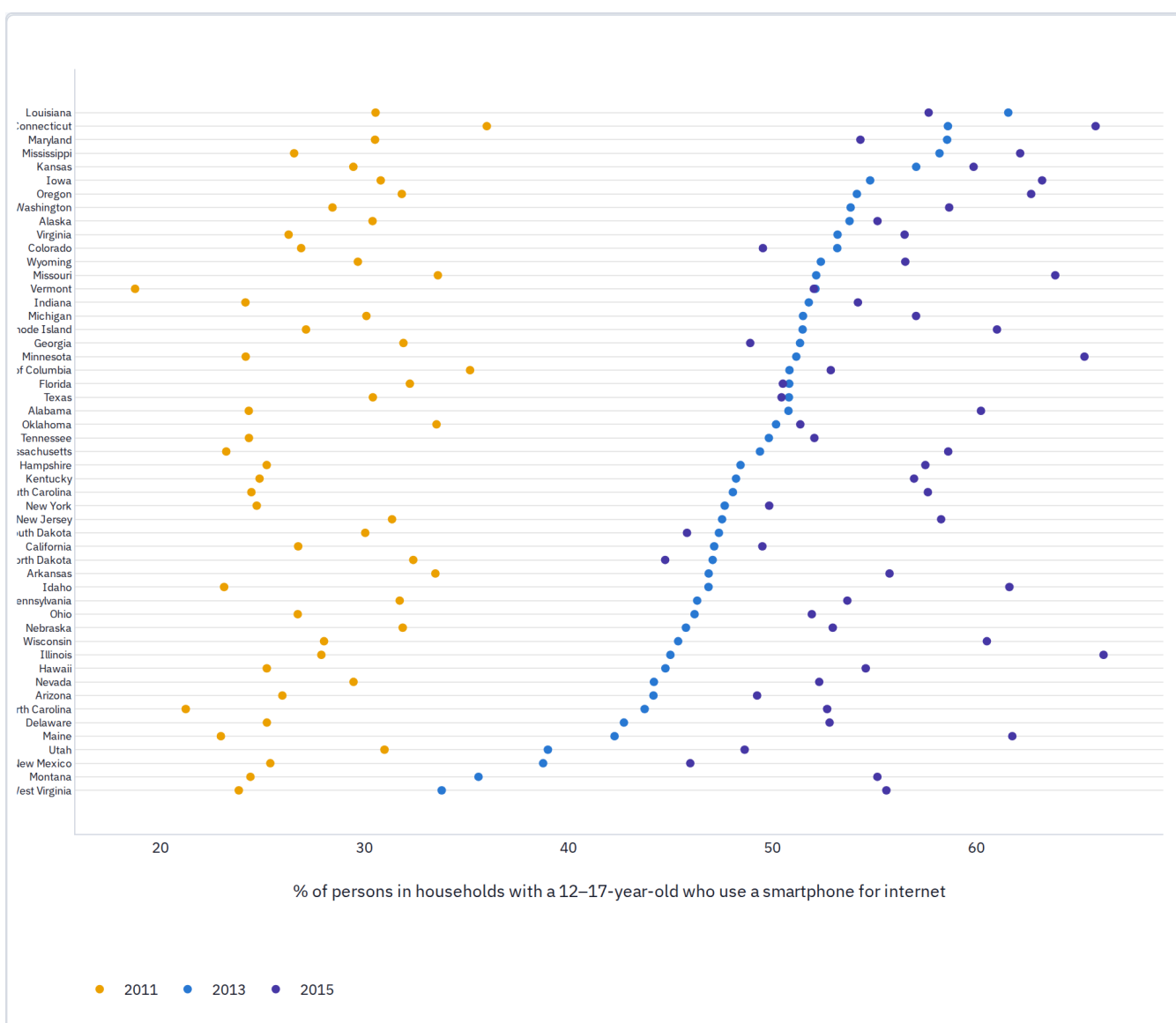
Smartphone adoption in the United States, 2009–2023

IPUMS CPS carries only part of the Computer and Internet Use Supplement (no mobile items), so the raw Census supplement files for October 2009 through November 2023 were rebuilt following NTIA's own harmonization code. The series below are the pre-2016 adoption measures the ACS cannot supply. The grey band marks the 2015 questionnaire redesign, which breaks the household mobile-plan series (at-home mobile broadband before, any-location data plan after).



Weighted shares by wave. Smartphone-internet use: "use a cellular phone or smartphone to access the internet" (2011, 2013, all persons), NTIA's internet-enabled-phone definition from 2015. 2012 phone items were asked of the primary respondent only and are omitted here.

Smartphone internet use in households with teenagers went from roughly 30% (2011) to 45% (2013) to 55% (2015), which is the adoption window the design needs; by 2017 it is near saturation. State-level shares in 2011–2015 vary by 20–30 points across states, which is the variation in adoption available for the U.S. tests.

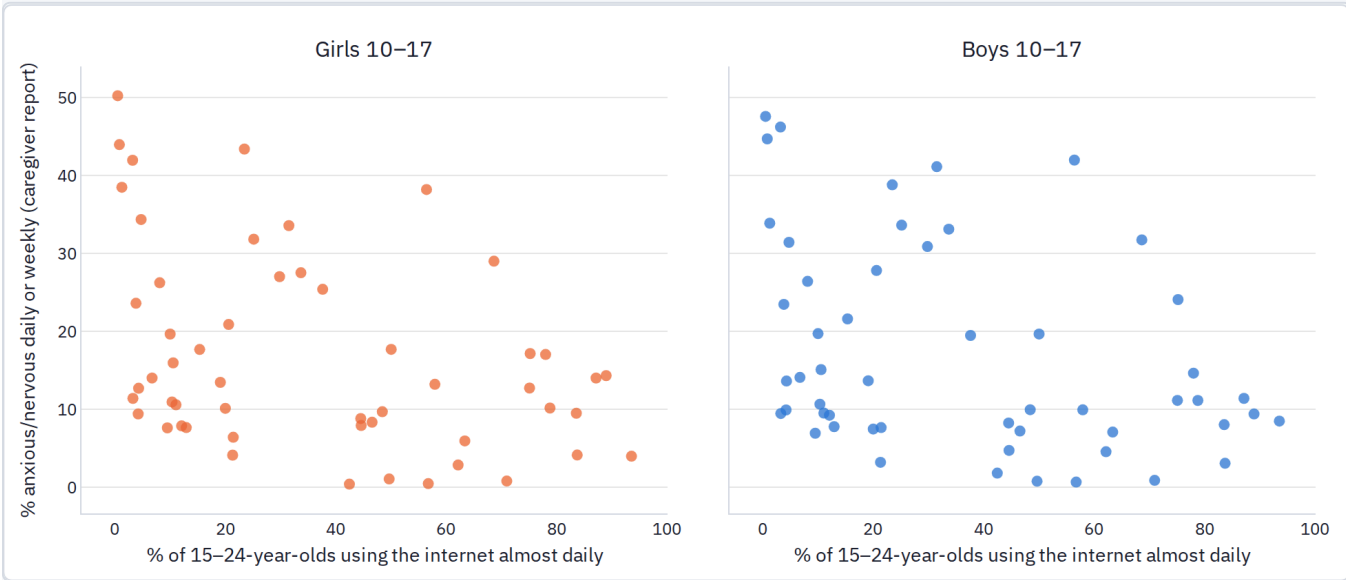


States ordered by the 2013 share. CBSA-level shares for 260–295 metros per wave are in the extract, but some have few respondents (check the unweighted counts before use).

Phone-first countries: MICS Child Functioning Module (68 surveys, 2017–2024)

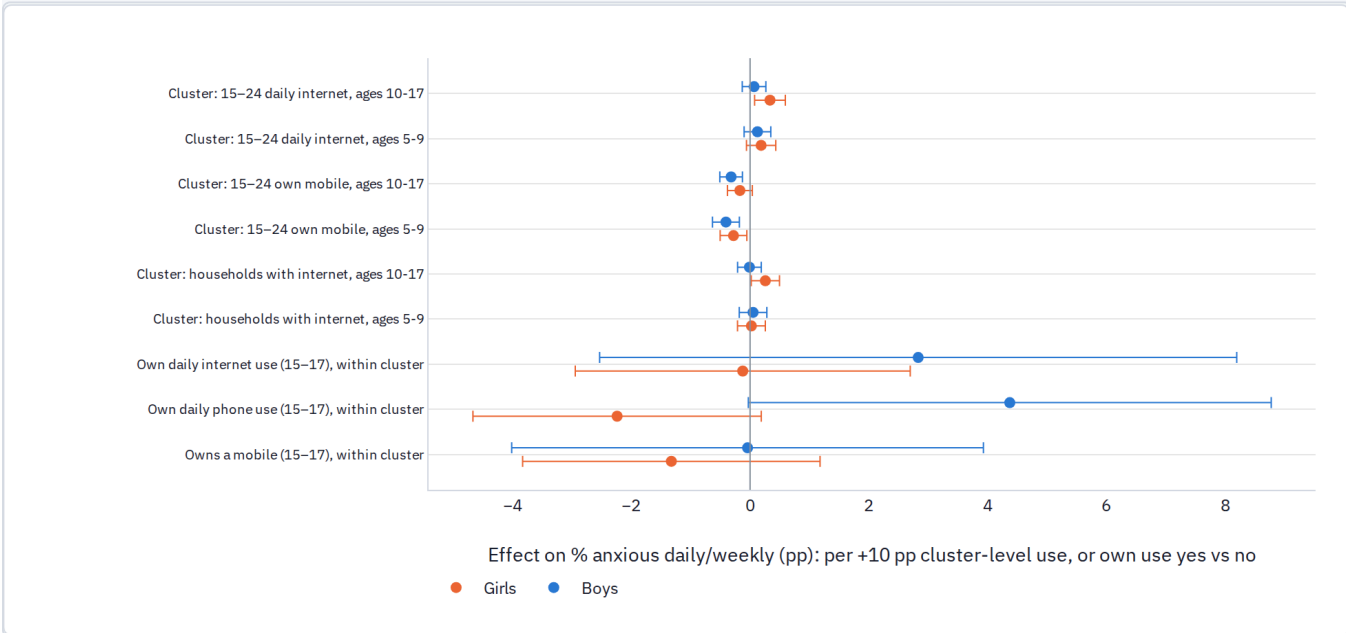
MICS6 and MICS7 (2017 onward) carry the Washington Group / UNICEF Child Functioning Module for every child aged 5–17: the caregiver reports how often the child "seems very anxious, nervous or worried" and how often "very sad or depressed" on a daily / weekly / monthly / a few times a year / never scale, with identical wording in every country. The same surveys ask women 15–49 (and men in about half the surveys) about mobile ownership and how often they used the internet and a mobile phone in the past three months, and the household questionnaire records internet access. All 68 samples on the IPUMS-MICS server with the module were read from the fixed-width files: 507,478 children with valid responses in 67 surveys.

Prevalence across surveys



One point per survey, ages 10–17. Cross-survey variation in caregiver-reported anxiety is enormous (from under 1% in Turkmenistan to 40% in Tunisia) and is dominated by reporting culture, which is why all tests below are within country.

Use and symptoms within countries



Cluster rows: children's anxiety against the share of 15–24-year-olds in their sampling cluster who use the internet almost daily (or own a mobile, or the household-internet share), with survey-by-region fixed effects, wealth quintile, urban and age controls, SE clustered by sampling cluster. Own-use rows: 15–17-year-olds linked to their own record in the women's or men's questionnaire (85% of girls matched, 24% of boys), comparing users to non-users within the same cluster.

Girls aged 10–17 in clusters where more young adults use the internet daily are reported slightly more anxious (+0.34 pp per 10-point rise in the cluster share, SE 0.13). But a similar coefficient

appears for 5–9-year-old girls (+0.19, SE 0.13), who are not the exposed group, so this is a property of connected clusters rather than of adolescent phone use. Mobile ownership in the cluster is *negatively* related to anxiety for every age and sex, household internet is small and positive for girls only (+0.26 per 10 points, SE 0.12, with nothing at 5–9), and depression shows no relation to any measure of connectivity. The within-cluster own-use comparison, the closest MICS comes to the individual exposure Haidt describes, shows no effect or negative for girls: daily internet use –0.1 pp on anxiety (SE 1.4) and –1.8 pp on depression (SE 1.3), daily phone use –2.2 pp (SE 1.2). For boys the own-phone-use coefficients are positive (+4.4 and +5.5 pp, SE ~2) but rest on the quarter of boys with a men's-questionnaire record. The learning and friendship functioning items are flat throughout.

Two-round countries

Country	Ages	Sex	Rounds	Anxious daily/weekly %	Depressed daily/weekly %
KGZ	10-14	boys	2018 → 2023	11.3 → 7.5	7.0 → 6.3
KGZ	10-14	girls	2018 → 2023	13.1 → 8.2	7.5 → 7.4
KGZ	15-17	boys	2018 → 2023	10.7 → 9.3	5.5 → 4.3
KGZ	15-17	girls	2018 → 2023	11.9 → 12.7	6.5 → 6.8
MNG	10-14	boys	2018 → 2023	7.5 → 11.2	4.4 → 7.6
MNG	10-14	girls	2018 → 2023	9.9 → 13.3	6.3 → 9.0
MNG	15-17	boys	2018 → 2023	15.0 → 11.9	7.4 → 7.5
MNG	15-17	girls	2018 → 2023	20.5 → 16.0	13.7 → 10.6
TKM	10-14	boys	2019 → 2024	0.8 → 4.7	0.7 → 2.8
TKM	10-14	girls	2019 → 2024	0.8 → 2.1	0.9 → 1.7
TKM	15-17	boys	2019 → 2024	0.8 → 4.3	0.6 → 1.3
TKM	15-17	girls	2019 → 2024	1.7 → 4.8	0.8 → 3.1
TUN	10-14	boys	2018 → 2023	42.7 → 30.1	15.8 → 13.8
TUN	10-14	girls	2018 → 2023	37.5 → 28.0	12.9 → 14.5
TUN	15-17	boys	2018 → 2023	40.7 → 34.6	15.5 → 17.5
TUN	15-17	girls	2018 → 2023	39.6 → 30.9	16.3 → 20.8

DHS-8 self-reports: Nepal 2022

The DHS-8 mental-health module gives a self-reported counterpart to the MICS caregiver items. Nepal 2022 administered the PHQ-9 and GAD-7 to a subsample of 7,442 women aged 15–49, alongside internet-use frequency and smartphone ownership; Kenya 2022 asked only about

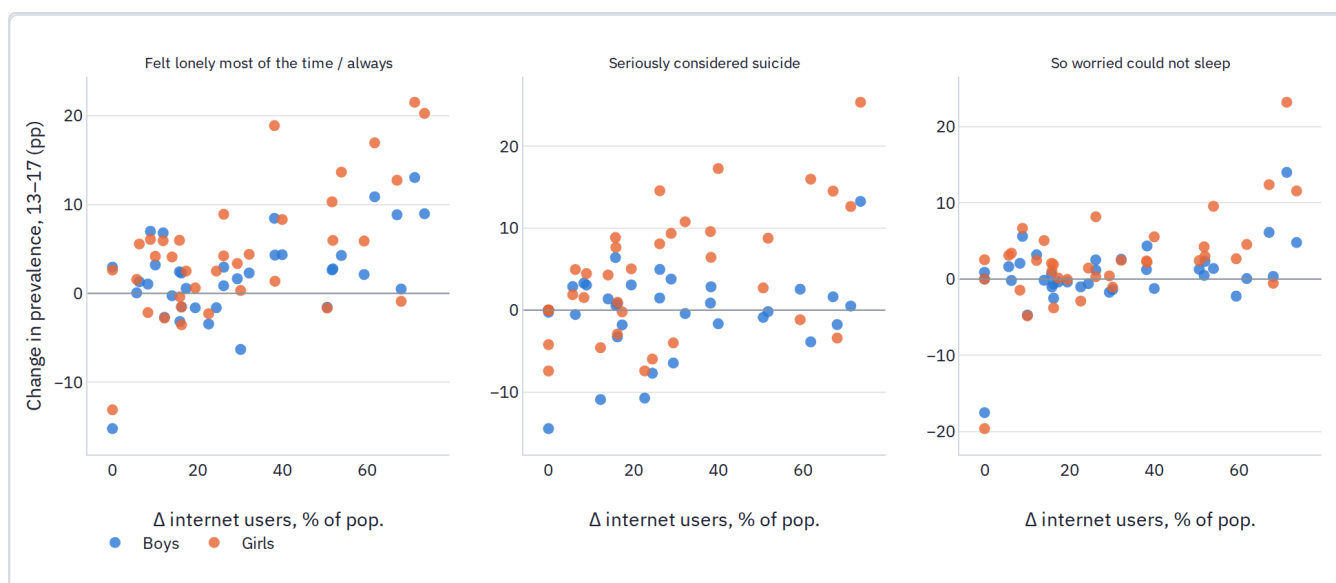
diagnosed depression or anxiety. Among Nepali women aged 15–19 ($n = 1,379$; 4.5% with PHQ-9 ≥ 10 , 42% using the internet almost daily), daily internet users are 2.3 pp more likely to screen positive for depression (SE 2.4) and 3.5 pp for anxiety (SE 2.9) than non-users in the same sampling cluster, with wealth, education and age held fixed; the same coefficients are positive at 20–24 and negative at 25–49. Smartphone ownership itself is unrelated. In Kenya, diagnosed depression among 15–19-year-olds is unrelated to owning a smartphone (-0.1 pp, SE 1.3). These are single-survey cross-sections and too small to be decisive, but they line up with the GSHS hours gradient rather than with the MICS caregiver no effect, which is itself a finding: the association appears when the adolescent reports her own state and her own use.

Countries with both a MICS6 and a MICS7 round on the server (Pakistan excluded: its rounds are different provinces). Turkmenistan's rise is from a very low base; Kyrgyzstan and Tunisia fall; Mongolia is mixed.

What this shows. This is the out-of-sample test the study design asked for, and it does not find the smartphone effect: neither cluster-level connectivity nor a girl's own daily internet or phone use predicts caregiver-reported anxiety or depression once the country and region are held fixed, and the one positive cluster coefficient is matched by the 5–9-year-olds (a placebo comparison, since they are not the exposed group). Qualifications: the outcome is a caregiver report, not self-report; exposure is measured as use rather than as the timing of network coverage, so the design with GSMA coverage maps remains the one that could establish cause; and the before-and-after comparisons will sharpen as more MICS7 rounds are released.

Phone-first countries: GSHS school surveys (149 country-rounds, 2003–2023)

The WHO Global School-based Student Health Survey asks 13–17-year-olds in 100+ countries the same loneliness, anxiety-related sleep loss and suicidality items, with rounds from 2003 onward. All 165 public-use country-rounds are freely downloadable from the WHO microdata repository; 149 were stacked and harmonized, of which 91 rounds in 39 countries have two or more rounds and can be used within-country. Exposure here is the World Bank internet-users share for the survey year; ITU coverage and mobile-broadband exposures follow below.



Each point is a country's change from its first to its last GSHS round (weighted prevalence, ages 13–17), against the change in the national internet-user share over the same years.

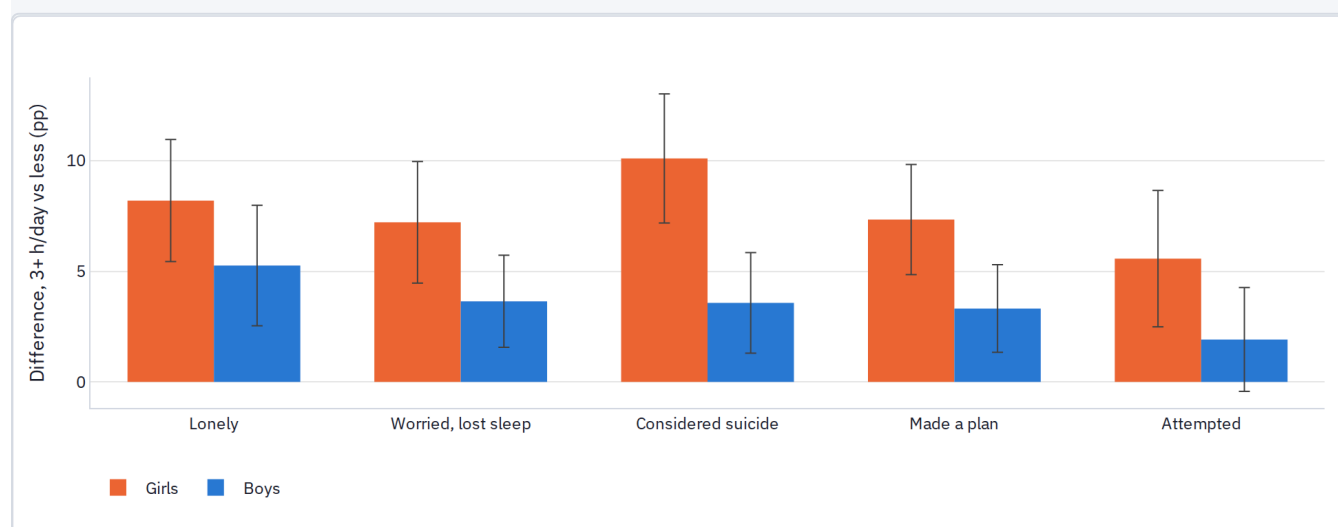
Outcome	Sex	Sample	Effect of +10 pp internet users (SE)
lonely	girls	79 rounds / 34 countries	+1.04 (0.83)
lonely	boys	79 rounds / 34 countries	-0.18 (0.59)
worried sleep	girls	77 rounds / 33 countries	+0.14 (0.87)
worried sleep	boys	77 rounds / 33 countries	-0.78 (0.54)
suicide consider	girls	67 rounds / 30 countries	+0.02 (1.27)
suicide consider	boys	67 rounds / 30 countries	-0.42 (0.99)
suicide plan	girls	70 rounds / 31 countries	+0.86 (0.74)
suicide plan	boys	70 rounds / 31 countries	-0.35 (0.69)
suicide attempt	girls	35 rounds / 19 countries	+0.30 (0.90)
suicide attempt	boys	35 rounds / 19 countries	-0.50 (0.38)
no close friends	girls	79 rounds / 34 countries	+0.48 (0.33)
no close friends	boys	79 rounds / 34 countries	-0.16 (0.36)

Country and year fixed effects, weights = inverse sampling variance, SE clustered by country. Rows with degenerate SEs (too few countries carrying the item) are omitted.

What this shows. With the full stack there is no within-country relation between internet penetration and adolescent distress: for girls, +1.0 pp loneliness per 10-point rise in internet users (SE 0.8) and +0.02 pp for suicidal ideation (SE 1.3); boys' coefficients are slightly negative. A partial stack of 72 rounds earlier in this run had shown +2.5 pp for girls on both items, which did not survive the remaining 77 rounds, a useful reminder of how noisy country-round panels are. Country-level internet penetration is also a coarse exposure: it does not distinguish adolescent from adult access or phone from PC. The two subsections below use, respectively, the 2022–23 rounds that ask adolescents directly about phone hours and the ITU coverage series.

Asking adolescents directly: mobile-phone hours in the 2022–23 rounds

Five GSHS rounds (Ghana 2022, India 2022, Jamaica 2023, Morocco 2023, Wallis and Futuna 2023) added "hours per day you used a mobile phone in the past 7 days". This puts each student's own reported phone use next to the same student's loneliness and suicidality answers.



Country fixed effects and age controls, survey weights, SE clustered by school; 7,708 girls and 7,102 boys, including students who had no phone. Girls using a phone three or more hours a day report loneliness +8 pp (SE 1.4), suicidal ideation +10 pp (1.5) and attempts +6 pp (1.6) compared with everyone else; boys +5, +4 and +2. The gradient rises with hours: relative to under an hour a day, girls at more than 8 hours report loneliness +20 pp, suicidal ideation +17 pp, attempts +14 pp. Students without a phone sit a few points above the under-an-hour group, not below it.

This is the pattern the phone hypothesis predicts: large, concentrated in girls, and rising steadily with hours of use. It is also exactly the kind of one-time self-report association the two camps disagree about, since heavy phone use and distress can share causes or run the other way. It cannot settle the timing question, but it does mark the 2022–23 GSHS rounds (and the MICS7 rounds with adolescent modules) as the place to bring in network-coverage timing, because here the exposure is measured on the adolescent rather than inferred from adults.

Network coverage at country level (ITU)

With the ITU series, the same comparison can use network coverage rather than reported use. Same model: country and year fixed effects, inverse-variance weights, SE clustered by country, countries with two or more rounds.

Exposure (per +10 points)	Countries / rounds	Lonely, girls	Lonely, boys	Considered suicide, girls	...boys	Attempted, girls	...boys
Mobile-broadband subs per 100 (ITU)	22 / 47	+0.21 (0.54)	+0.03 (0.42)	+0.28 (0.43)	+0.35 (0.48)	+0.31 (0.44)	+0.30 (0.50)
3G population coverage (ITU)	17 / 36	-0.72 (0.28)	+0.25 (0.26)	+0.10 (0.59)	+0.28 (0.44)	-0.08 (0.32)	+0.36 (0.36)
4G population coverage (ITU)	15 / 32	-0.40 (0.41)	-0.48 (0.38)	-0.51 (0.63)	-0.20 (0.59)	-0.64 (0.43)	-0.45 (0.69)
Internet users % (ITU)	33 / 78	+0.93 (0.86)	-0.19 (0.60)	-0.04 (1.20)	-0.54 (1.07)	+0.37 (1.02)	-0.22 (0.86)

Nothing here favors H1: 4G coverage, 3G coverage and mobile-broadband subscriptions all give coefficients within about ± 0.5 pp per 10 points (the largest, -0.5 pp for girls' suicidal ideation on 4G coverage, has a standard error of 0.6), with the girls' loneliness coefficient on 3G coverage negative. The 4G rows rest on the 15 countries with a reported 4G series in two or more survey years. Country-level coverage is still a crude measure (within-country changes are large, 20–35 points, but they move with everything else that changed between 2007 and 2019), which leaves local, cluster-level coverage as the remaining test.

Phone-first countries: census microdata (IPUMS International)

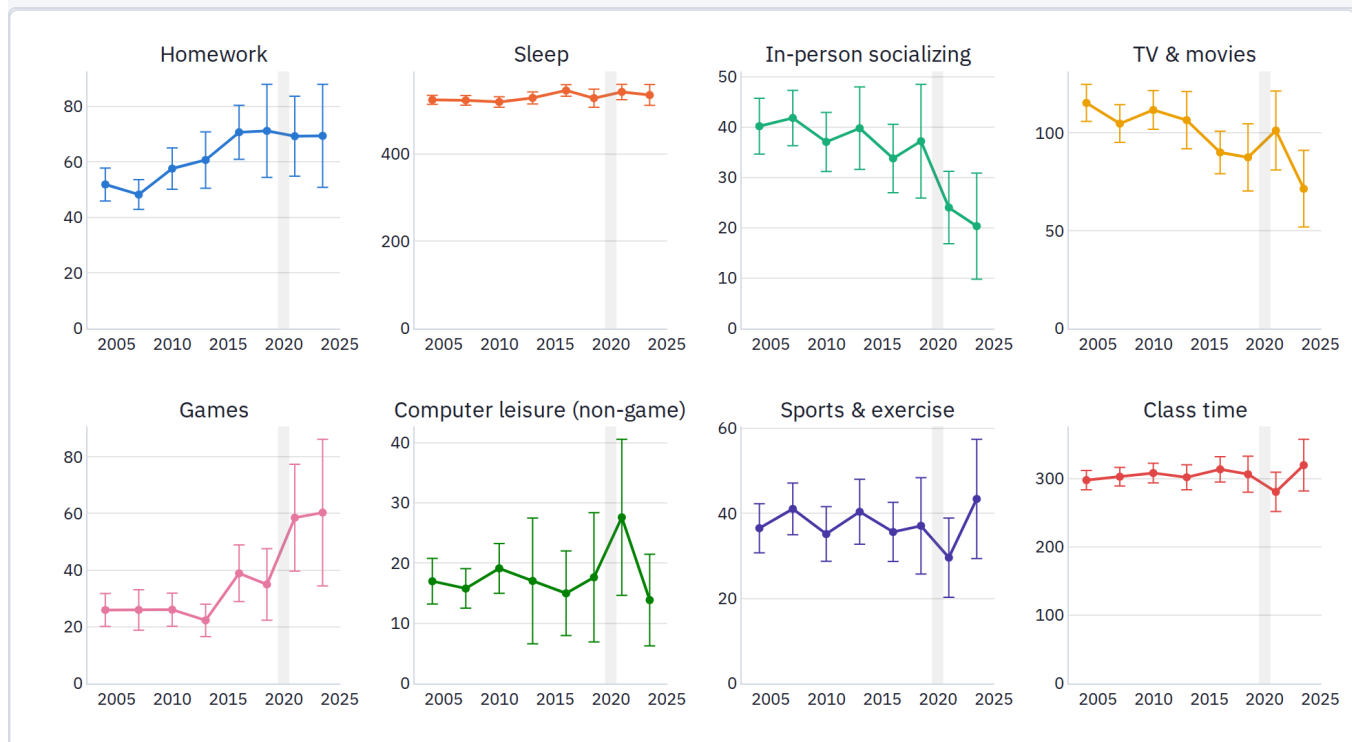
Kenya 2009/2019 and Mexico 2010/2020 were scanned as the best pairs. Both 2020-round censuses carry the Washington Group item on difficulty remembering or concentrating, and Kenya 2019 also asks each person about mobile phone ownership and internet use. But the 2010-round items were yes/no "mental disability" questions, so the two-point change is dominated by wording. Within 2019/2020, prevalence at 12–17 is 0.8–1.8% (any difficulty) and 0.2–0.7% (a lot or cannot), boys exceed girls at every child age, and district-level regressions of the 12–17 rate on teen phone ownership or internet use (Kenya, 76 district-sex observations) or household internet (Mexico, 3,259 municipality-sex observations) give coefficients within ± 0.01 pp per percentage point of exposure level. The census item behaves like a developmental-disability measure and is not sensitive to adolescent distress. This evidence can only corroborate, not test.

Part 3 • School pressure (H2)

The school-pressure hypothesis predicts that distress rose where and when schooling became more demanding: in the high-stakes years, in the school year, in states that adopted accountability testing, and in countries where students report more schoolwork pressure. The tests run from time use, to the cross-national HBSC series, to U.S. state policy.

Did school demands rise? Time use of 15–17-year-olds (ATUS, 2003–2024)

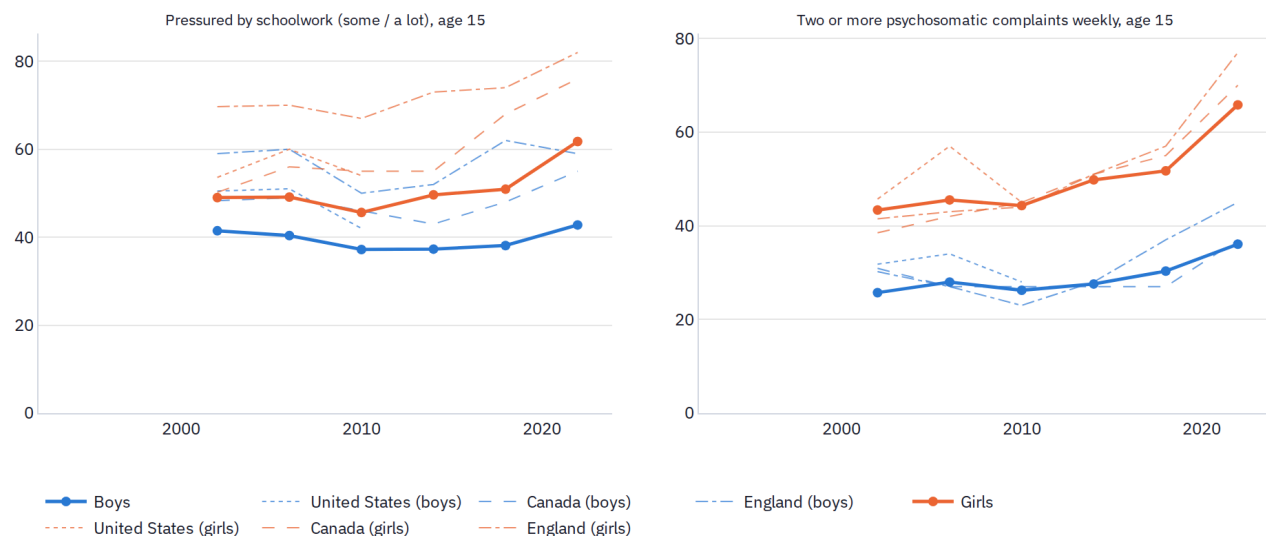
The American Time Use Survey gives a direct look at the school-demand channel. On school-year weekdays, homework time rose from about 50 minutes (2003–08) to about 70 minutes (2015–19) and stayed there; sleep did not fall; in-person socializing drifted down before collapsing after 2020; television fell and games rose. Note the small samples (roughly 200–800 teens per pooled period) and that phone use is mostly recorded as a secondary activity, which ATUS does not capture.



Weighted minutes per day, non-holiday weekdays in September–May, pooled periods; bars are 95% CIs.

Schoolwork pressure across countries (HBSC)

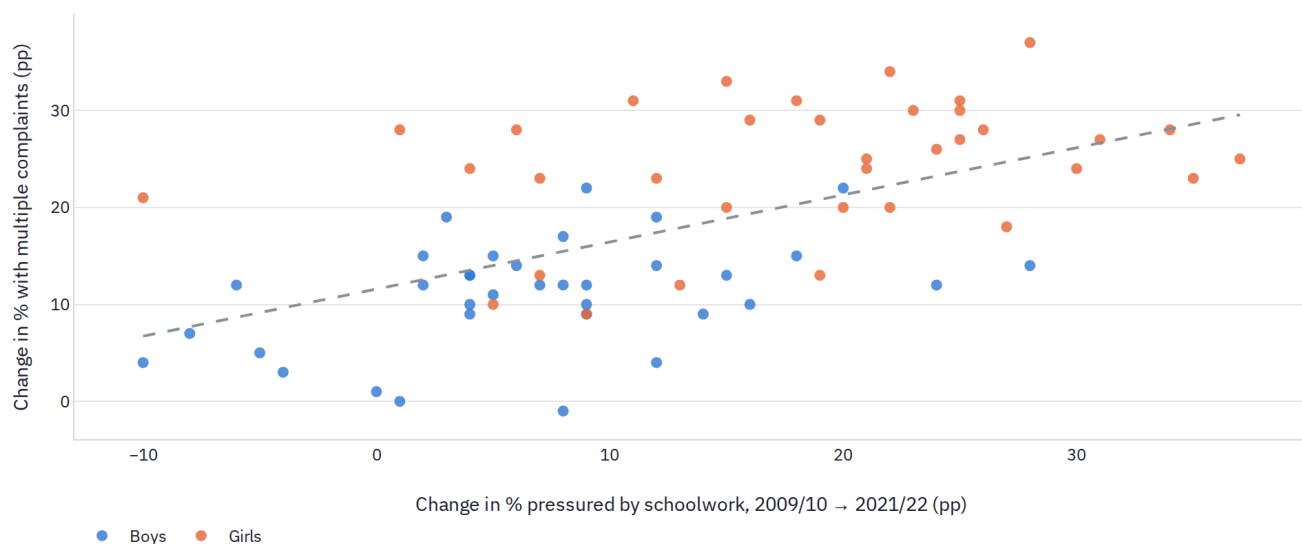
The HBSC surveys ask 11-, 13- and 15-year-olds in ~45 countries the same schoolwork-pressure question and the same psychosomatic-complaints checklist every four years. The published country tables (1997/98 to 2021/22, plus the HBSC data browser for 2014–2022) give a direct public test of Gray's mechanism outside the United States. The U.S. took part only through 2009/10, so the U.S.-specific version of the thesis cannot be tested here after 2010.



Unweighted means across participating countries (thick lines), with the United States, Canada and England drawn thin. Age 15.

Both series were flat or falling through 2009/10, then rose together: for 15-year-old girls the cross-country mean of schoolwork pressure went from 46% (2010) to 50% (2014), 51% (2018) and 62% (2022), while multiple complaints went from 44% to 50%, 52% and 66%. The U.S. was above average on pressure in the 2000s but its pressure fell between 2006 and 2010, the last U.S. wave.

Did complaints rise where pressure rose?



Each point is a country-sex pair at age 15; dashed line is the pooled fit (slope 0.24, SE 0.07, $r = 0.59$). For 2010 → 2018 alone: slope 0.16 (SE 0.07), $r = 0.31$.

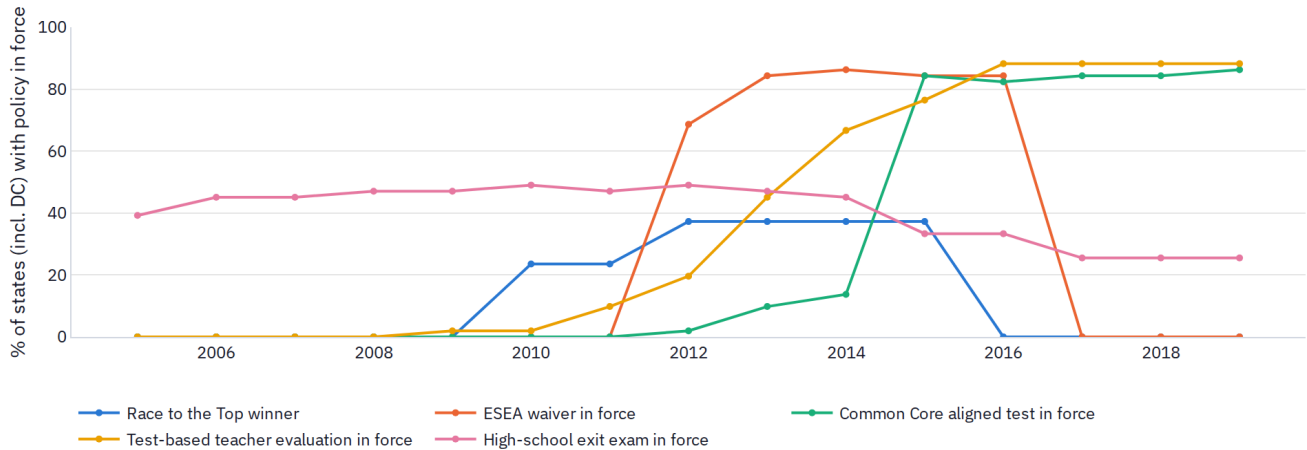
Test (age 11–15 pooled unless noted)	Estimate	Reading
Within-country shared movement, complaints on pressure, country×age×sex and wave×age×sex FE, 1,440 observations	0.14 (SE 0.04); girls 0.16, boys 0.12	a 10-pp rise in pressure goes with a 1.4-pp rise in complaints
Long change 2010 → 2018, age 15, 72 country-sex pairs	mean Δ pressure +5.5 pp, Δ complaints +6.6 pp; slope 0.16	pressure differences explain ~1 pp of the 6.6-pp rise
Long change 2010 → 2022, age 15, 70 pairs	Δ pressure +12.7, Δ complaints +17.7; slope 0.24, r 0.59	stronger, but 2022 includes the pandemic
Placebo period 2002 → 2010	Δ pressure –2.9, Δ complaints –0.1; slope 0.14	same slope when neither was rising
2018/2022 cross-country: complaints on problematic social media use and pressure, wave×age×sex FE, 496 observations	PSMU 0.53 (SE 0.15); pressure 0.13 (SE 0.06)	both independently associated
Same, within-country change 2018 → 2022	PSMU 0.25 (SE 0.18); pressure 0.12 (SE 0.10)	underpowered with one change per country

What this shows. Gray's mechanism has cross-national support in the sense that schoolwork pressure rose after 2010 in most HBSC countries, not only the U.S., and complaints rose more where pressure rose more. But the slope is modest and is the same in the placebo comparison period, so pressure behaves like a stable correlate of complaints rather than the driver of the post-2010 break: most of the rise is common to all countries and is absorbed by the wave effects. The concentration test (are complaints increasingly concentrated among high-pressure students?) needs the HBSC microdata.

U.S. school accountability policy

The central U.S. test of the school-pressure hypothesis: states adopted accountability measures at different times; if school pressure mattered, their teenagers should have turned worse when they did.

The state policy table dates five accountability measures for 51 jurisdictions, with a source for every cell: Race to the Top awards (2010–11), ESEA/NCLB waivers (2012–16), the first operational Common Core aligned test (2012–2019, mostly spring 2015), test-based teacher evaluation (implemented statewide 2009–2017), and high-school exit exams (falling from 25 states to 13). Year is the school year ending that spring.



Share of states with each policy in force. The near-simultaneous 2012–2015 adoption of waivers, teacher evaluation and aligned tests is exactly the timing problem the study design anticipated.

State-year regressions

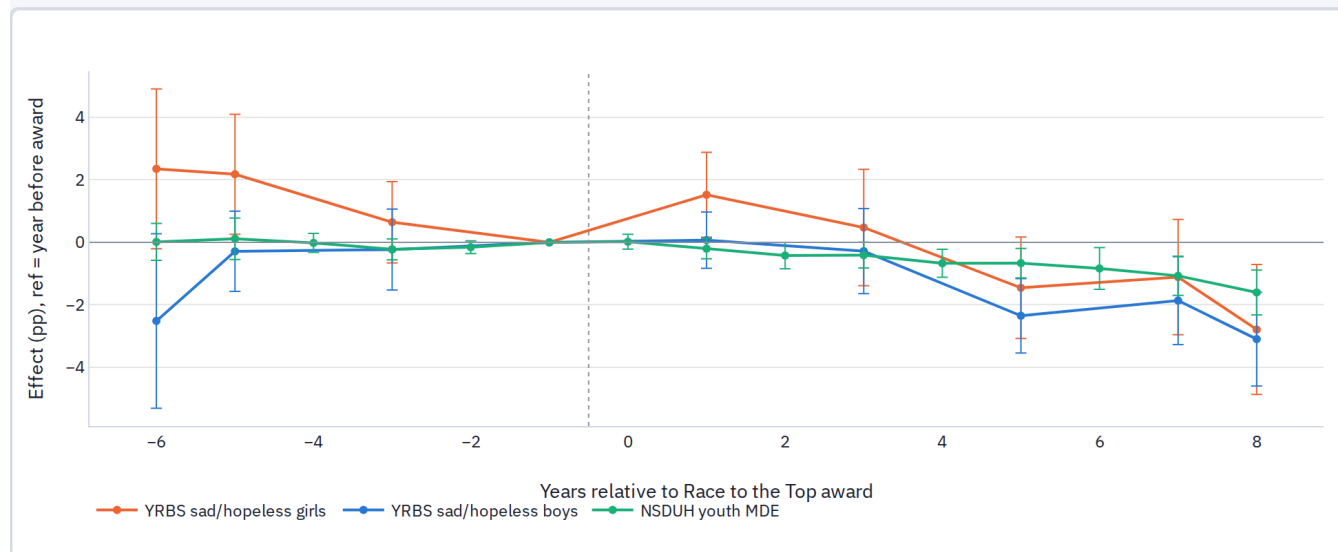
Policy (each estimated alone)	ACS girls	ACS boys	YRBS sad girls	YRBS sad boys	NSDUH youth MDE
Race to the Top winner (award year–2015)	-0.06 (0.07)	-0.00 (0.10)	+1.21 (0.41)	+0.68 (0.34)	+0.16 (0.14)
ESEA waiver in force	-0.01 (0.08)	+0.13 (0.07)	-0.99 (0.58)	-0.16 (0.48)	-0.18 (0.25)
Common Core aligned test in force	-0.00 (0.08)	+0.01 (0.12)	-0.80 (0.79)	-0.80 (0.77)	-0.35 (0.30)
Test-based teacher evaluation in force	-0.03 (0.06)	-0.13 (0.09)	-0.38 (0.60)	-0.28 (0.36)	-0.03 (0.17)
High-school exit exam in force	-0.04 (0.07)	-0.03 (0.12)	-0.47 (0.82)	-0.09 (0.86)	+0.34 (0.27)
Pressure index (sum of the five)	-0.03 (0.03)	-0.01 (0.04)	-0.11 (0.25)	-0.04 (0.20)	+0.02 (0.09)

Coefficients in percentage points, state and year fixed effects, inverse-variance weights, SE clustered by state, 2008–2019 (YRBS biennial, 39 states). NSDUH reports two-year periods; each is matched to the policy averaged over its two survey years. Consecutive periods share half their sample, which the state clustering absorbs; keeping only non-overlapping periods moves every NSDUH coefficient by less than 0.15 pp. Bold = $|t| > 1.96$.

Common Core aligned testing, waivers, teacher evaluation, exit exams and the summed index show no relation to any of the three outcomes; several point estimates are negative. The one coefficient beyond two standard errors is Race to the Top on YRBS girls' sadness (+1.2 pp, SE 0.4; boys +0.7, SE 0.3); the NSDUH coefficient is +0.16 pp (SE 0.14) once each two-year period is

matched to the policy in force across both of its years, and the ACS item is flat. RTTT winners are a selected group (19 jurisdictions with the capacity to write winning applications, including DC, NY, MA, FL). Tracing the same states year by year around the award, in the U.S. follow-ups below, shows this coefficient is an artifact of how the indicator was coded; there is no rise after the award.

Race to the Top, traced year by year



State and year fixed effects, coefficients relative to the year before the award (school year ending 2010 for Phase 1 and 2, 2012 for the December 2011 Phase 3 awards), 2005–2019; both award cohorts are observed in the YRBS wave before their award (2009 and 2011), and years 6 or more before and 8 or more after are pooled into end bins.

The static Race to the Top coefficient from the school-policy section does not survive being traced year by year around the award: afterwards, YRBS sadness and NSDUH youth depression in winning states move flat to downward relative to other states. The static estimate arose from coding the indicator as 1 only for 2010–2015, so the winners' post-2016 years served as controls. There is no evidence that RTTT states' teenagers became more distressed. The U.S. policy-shock version of H2 is therefore not supported; testing H2 further needs school-level pressure items, which no U.S. public survey carries.

Part 4 • The interaction (H3): phones in the same hours as high-pressure school

The interaction hypothesis in its sharpest form says that what harms adolescents is not phones as such but phones *during school hours*, on top of high-stakes academic demands in the same hours: social comparison and notifications running through the day that also carries the tests and the grades. If that is right, three things follow. Pressure and phone use should both have risen after 2010 (they did, in most HBSC countries and in U.S. homework time). Removing phones from the

school day should bring relief out of proportion to the hours removed, and more relief for students under more academic pressure. And out-of-school use, or use in low-pressure settings, should matter less.

What the phone-ban evidence says about this

PISA 2022 allows one direct check: does the ban's association with well-being differ between students under high and low academic anxiety? The only pressure-like measure asked of all students is the mathematics-anxiety index (the schoolwork-anxiety items were not asked in 2022), so "high anxiety" here means above the country median on that index. Same model as before, within country, by sex.

Outcome	Sex	Ban, low-anxiety students (SE)	Ban, high-anxiety students	Difference (SE)	High vs low anxiety, no ban
Lonely at school	girls	-0.27 (0.68)	-1.15	-0.88 (0.93)	+8.2
Lonely at school	boys	+1.12 (0.53)	-0.81	-1.94 (0.98)	+7.4
Nervous without device	girls	-0.75 (0.60)	-0.49	+0.26 (0.78)	+9.3
Nervous without device	boys	+0.13 (0.53)	+0.19	+0.06 (0.81)	+5.8
Belonging index	girls	+2.34 (1.60)	+2.46	+0.11 (1.90)	-22.7
Belonging index	boys	+0.32 (1.43)	+1.91	+1.59 (2.11)	-19.2
Life satisfaction index	girls	+3.09 (3.41)	+2.23	-0.86 (4.53)	-25.3
Life satisfaction index	boys	+2.54 (3.26)	+0.71	-1.82 (4.93)	-25.2

Percentage points for the binary items, hundredths of an SD for the indices; country fixed effects, socio-economic status and grade controls, student weights, SE clustered by school; 441,992 students with the anxiety index.

High-anxiety students are far lonelier (about 8 points more), and the ban's association with less loneliness is larger for them: -1.2 points for high-anxiety girls against -0.3 for low-anxiety girls, and -0.8 against +1.1 for boys. The differences run the way the hypothesis predicts, but they are inside their standard errors, and mathematics anxiety is a rough stand-in for schoolwork pressure. Bans reduce in-school phone use by the same amount for both groups (13 points), so the difference, if real, is in what the removal does rather than in who complies. The rest of the ban evidence below is consistent with the hypothesis without testing it: bans cut in-school use, girls in ban schools report less loneliness and less device anxiety, Norway's middle-school bans reduced girls' psychological consultations, and U.S. sadness moved inversely to school prohibitions. The test that would separate this from the plain phone hypothesis is a ban study with a measure of

school pressure, which the HBSC microdata (schoolwork pressure plus school-level policy in 2021/22) can provide.

Did 4G timing matter more in high-pressure states?

Outcome (ACS 12–17, market panel)	LTE × policy	Interaction (SE)
girls	pressure_index	-0.34 (0.25)
girls	cc_aligned_test_in_force	not identified
girls	waiver_in_force	+0.12 (1.02)
girls	rttt_winner	-0.94 (0.49)
boys	pressure_index	+0.36 (0.34)
boys	cc_aligned_test_in_force	not identified
boys	waiver_in_force	-0.04 (1.63)
boys	rttt_winner	+0.61 (0.75)

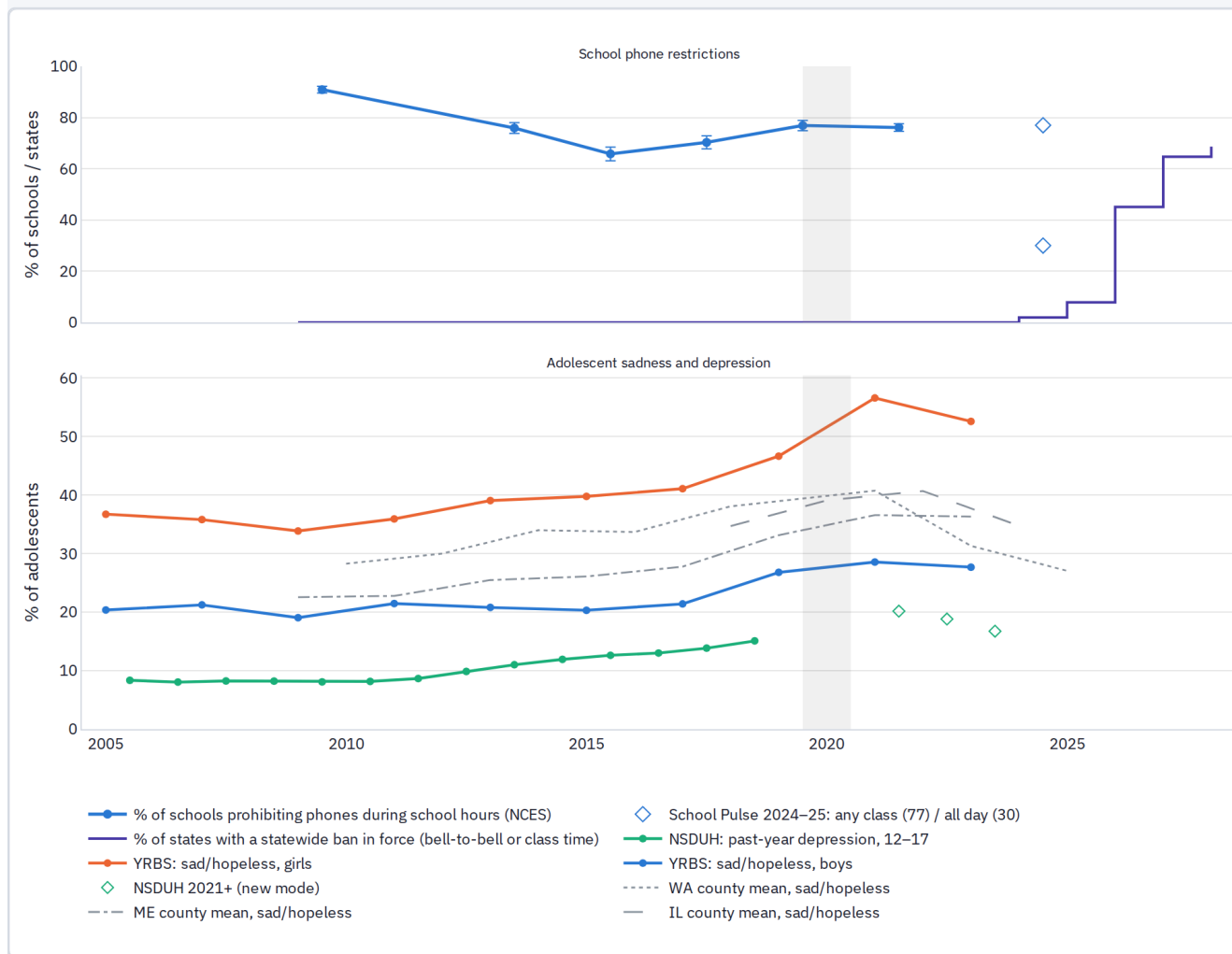
State-by-year fixed effects absorb the policy's own effect; what remains is whether early-LTE metros within a state rose more when the state had the policy. "Not identified": the policy switches on in the same year for nearly every metro in the state, so there is nothing left to compare.

Where identified, the interaction is negative for girls (LTE × pressure index -0.34 , SE 0.25 ; LTE × RTTT -0.94 , SE 0.49): early-LTE markets in high-pressure states did not rise more. This is the opposite of what H3 predicts, though with the same market-level caveats as the main event study.

School phone bans

School phone policy is the one policy both theses agree should matter if phones are the mechanism, and it is the observable from the four-hypothesis table that the first pass had not touched. This section sets what public data allow as of September 2026 beside the causal literature.

Bans and sadness on one time axis



Top: principal-reported share of U.S. public schools prohibiting non-academic phone use during school hours (NCES SSOCS; 2013–14 from a smaller survey; wording changed 2017–18), the 2024–25 School Pulse Panel points (class-time construct), and the share of states with a bell-to-bell or class-time mandate in force, from the policy table. Bottom: NSDUH past-year major depressive episode ages 12–17, YRBS sad/hopeless by sex, and county-mean sadness from the Washington, Maine and Illinois surveys. Grey band = 2020.

Read as a picture, the two panels move in opposite directions twice. Prohibitions fell from 91% to 66% between 2009–10 and 2015–16 while depression and sadness rose from their 2011–13 lows; prohibitions then climbed back through 2019–20 and, after the pandemic peak, sadness fell in every state series as the 2024–25 wave of state mandates arrived. The correlation is far from clean: the sadness series kept rising through 2019–2021 while prohibitions were already recovering, and the post-2021 fall began before most state laws took effect. Two hypotheses fit the shape, and neither is established:

- **Partial re-restriction plus awareness.** The modest recovery in school restrictions from 2016, together with rising attention from parents, schools and adolescents themselves, may have reduced the most acute in-school and late-night exposure even before formal bans. The PISA

cross-section (less in-school use, less device anxiety, less loneliness among girls in ban schools) and Norway's staggered bans are consistent with this, but both are small effects and the U.S. timing is loose.

- **Content rather than access.** If what mattered in 2012–2017 was image-based social comparison (Instagram, Snapchat), the later migration of teen attention toward algorithmic consumption feeds (TikTok, YouTube) could raise total screen time while producing fewer acute internalizing symptoms. This would explain rising hours alongside falling sadness, and it is the mechanism the Facebook-rollout study points to (social comparison), but the pilot has no platform-mix data by year and cannot test it.

Lags and enforcement

Two features of the policy make the simple overlay above too demanding a test, and both argue for reading the chart with a lag of a year or two.

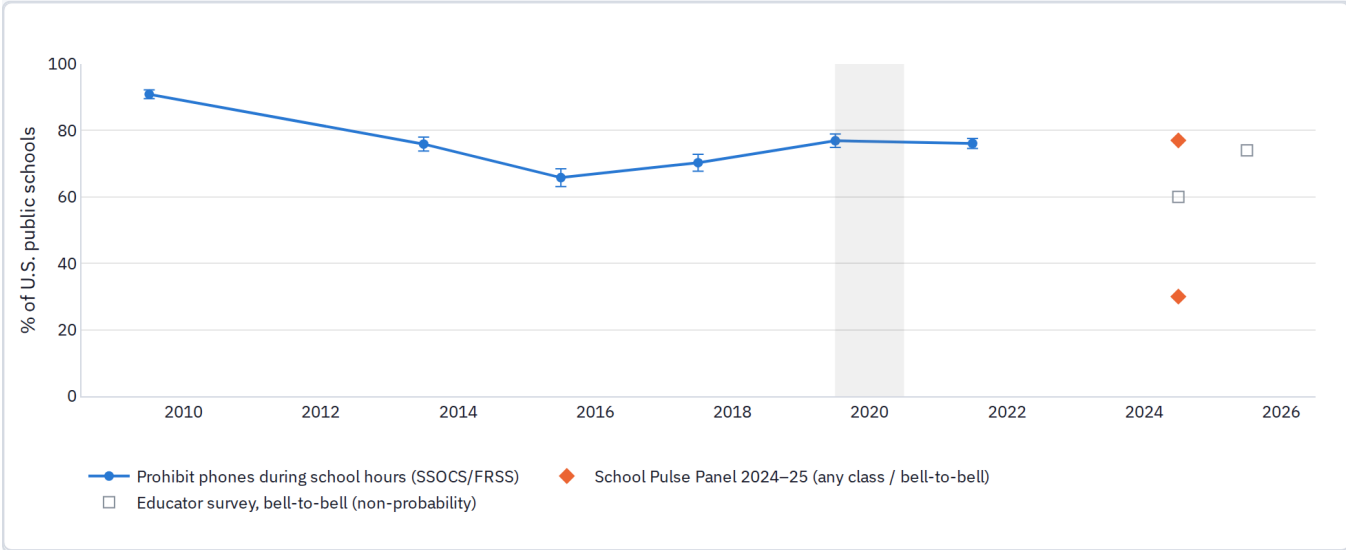
- **Delayed effects are what the causal studies show.** In the 2026 U.S. pouch study, student well-being *fell* by 0.2 SD in the adoption year and only turned positive (+0.16 SD) by the second year; Norway's bullying and mental-health effects grew over three years; the Florida district study found suspensions up 12% in year one, fading in year two. The pattern is resistance first, then adjustment, with administrators refining enforcement over a school year or more. Under that pattern the 2015–16 low point in restrictions would be expected to show up in outcomes around 2017–18, and the 2024–25 mandates around 2026–27, which is roughly what the overlay shows and is also why the 2025 YRBS is only the first, weakest post-ban wave. The U.S. test should therefore follow states for at least two years after a ban, not compare a single before and after.
- **"Policy" is not enforcement.** PISA 2022 measures both. Among schools that report phones are not allowed on the premises, a median 46% of students across 52 systems still used a smartphone at school daily (62% in non-ban schools), and in 24 of the 52 a majority did. The United States is among the weakest: 70% daily use in ban schools against 82% elsewhere, a gap of 12 points, compared with gaps of 30–46 points in Japan, Taiwan and Brunei. Enforcement in the U.S. in 2022 was therefore mostly nominal, which is consistent with the idea that restrictions only began to take hold as administrators gained confidence, and with the SSOCs series measuring stated policy rather than practice. It also means any effect found in the 2025–27 waves will be a lower bound on what full enforcement could do.

Country-by-country daily-use rates in ban and non-ban schools are in [results/tables/pisa_ban_enforcement_by_country.csv](#).

To be clear, this graph raises the question rather than answering it. The 2025 YRBS, the 2025–27 state-survey waves, and the district-level rollouts already in the policy table are what will turn the shape into a test: bans arriving at different times in different states and districts, against outcomes measured before and after, by sex, with the 65+ placebo comparison retained.

In detail: schools *relaxed* phone prohibitions exactly when distress rose: from 91% prohibiting in 2009–10 to 66% in 2015–16, during the years when smartphones replaced feature phones. Prohibitions then climbed back (77% by 2019–20) and the 2024–25 wave of state laws has pushed bell-to-bell policies toward a majority of schools. That is a national series, so it cannot separate school policy from everything else that changed; but it means the 2010s rise happened while access to phones *in school* was expanding, which fits either thesis, and the 2023–25 fall in sadness in the state surveys coincides with the re-tightening.

The national series in detail



Principal-reported share of public schools prohibiting non-academic phone use during school hours (NCES SSOCS; the 2013–14 point is a smaller Fast Response Survey; the item wording changed in 2017–18). The 2024–25 School Pulse Panel asks about class time, not school hours: 77% prohibit phones in any class, 30% in all classes and outside class.

State bans and what can be tested when

Policy type	Effective school year	States
ban during instructional time	2024-25	1 Indiana
ban during instructional time	2025-26	5 Iowa, Nevada, North Carolina, Tennessee, West Virginia
ban during instructional time	2026-27	2 Michigan, Wisconsin
requires districts to adopt a policy	2024-25	1 Minnesota
requires districts to adopt a policy	2025-26	4 Alaska, Arizona, Idaho, New Mexico

Policy type	Effective school year	States
requires districts to adopt a policy	2026-27	5 California, Colorado, Connecticut, Delaware, Wyoming
statewide bell-to-bell ban	2023-24	1 Florida
statewide bell-to-bell ban	2024-25	2 Louisiana, South Carolina
statewide bell-to-bell ban	2025-26	14 Alabama, Arkansas, District of Columbia, Kentucky, Missouri, Nebraska, New Hampshire, New York, North Dakota, Ohio, Oklahoma, Oregon, Texas, Virginia
statewide bell-to-bell ban	2026-27	8 Georgia, Hawaii, Kansas, Maine, New Jersey, Rhode Island, Utah, Vermont
statewide bell-to-bell ban	2027-28	2 Illinois, Maryland

Compiled from Ballotpedia, Education Week, KFF and state statutes; classification disputes between trackers are noted in the data file. Florida's HB 379 (instructional time, 2023–24) is the earliest statewide action.

Almost every statewide ban took effect in fall 2024 or fall 2025. The latest outcome waves in hand are spring 2023 (YRBS, Pennsylvania, Maine), spring 2024 (Illinois) and fall 2025 (Washington, guidance only), so no post-ban statewide outcome exists yet in this project; the 2025 YRBS state files (not released as of this writing) will give the first difference-in-differences: Florida, Indiana, Louisiana and South Carolina (2024–25) against states without a mandate, 2021 → 2023 → 2025, by sex, on sad/hopeless and considered suicide. The county panel then adds Washington 2027, Maine 2025 and Illinois 2026 against their own bans. The model is the same as the LTE one above, with the ban in place of LTE and the 65+ comparison group retained. NCES restricted-use SSOCS files carry state and county identifiers and would allow a school-level version, which is the Phase 2 item.

What the causal literature already shows

Study	Setting	Design	Outcome domain	Headline
Beland & Murphy (2016)	England	DiD across schools with staggered school-level bans (school survey on ban timing + NPD administrative data)	GCSE test scores	Post-ban test scores +6.41% of a SD; gains concentrated among lowest-achieving students (~+14% SD); no effect for high achievers (published: Labour Economics 41, 61-76)

Study	Setting	Design	Outcome domain	Headline
Kessel, Hardardottir & Tyrefors (2020)	Sweden	DiD replication of Beland & Murphy with school survey on ban timing (~75% response) + register data	grade 9 test scores / GPA	No impact on student performance; 'can reject even small-sized gains' (Economics of Education Review 77, 102009)
Beneito & Vicente-Chirivella (2022)	Spain	Synthetic control (Galicia) + DiD (Galicia, Castilla-La Mancha) vs other regions	PISA math/science; reported bullying cases	Galicia PISA math +10.7 and science +12.7 points by 2017 (0.6-0.8 and ~0.7-1 years of learning); CLM positive but insignificant; bullying cases per 10,000 down 15-18% (ages 12-14) and 9.5-18% (15-17), no effect under 12 (Applied Economic Analysis 30(90), 153-175)
Abrahamsson (2024)	Norway	Event study / staggered DiD using a survey of middle schools' smartphone policies linked to administrative health, education and family registers and the national pupil survey	GP and specialist consultations for psychological symptoms; reported bullying; GPA; academic high-school track	Girls: fewer GP visits for psychological symptoms (about -0.2 to -0.3 visits, a sizeable share of the mean) and ~60% fewer specialist mental-health visits; bullying down for both genders (girls exposed 3 years ~46%); girls' GPA +0.08 SD and higher academic-track enrolment; effects larger for low-SES girls; no effects for boys (NHH DP 01/2024; forthcoming Journal of Human Resources per Allcott et al. 2026)
Goodyear, Randhawa, Adab, et al. (SMART Schools) (2025)	England	Cross-sectional comparison of schools with restrictive vs permissive recreational-phone policies (not causal)	WEMWBS mental wellbeing; anxiety; depression; sleep; attainment; phone/social-media time	No difference in wellbeing: adjusted mean WEMWBS difference -0.48 (95% CI -2.05 to 1.06, p=0.62); no differences in anxiety, depression, sleep, attainment or total daily phone/social-media use; less use during school hours only (Lancet Regional Health - Europe 51, 101211)

Study	Setting	Design	Outcome domain	Headline
Figlio & Ozek (2025)	United States (Florida)	DiD comparing schools with high vs low pre-ban student phone activity (Advan building-level GPS) before/after the 2023 statewide ban, within one district	test scores (3x/year), disciplinary incidents, suspensions, absences	Year 1: incidents/suspensions +~25% raw, DiD +12% suspensions, in-school suspensions +~30% for Black male students, fading by year 2; year-2 test scores +0.6 percentiles overall, +1.1 percentiles on spring test (1.3 in middle/high, 1.4 for boys); unexcused absences significantly reduced in both years (NBER WP 34388)
Allcott, Baron, Dee, Duckworth, Gentzkow & Jacob (2026)	United States (national)	Staggered DiD (Callaway-Sant'Anna) of Yondr lockable-pouch adoption using pouch sales records, teacher survey, GPS pings, state test scores and administrative data; pre-registered	phone use; test scores; attendance; discipline; student well-being; classroom attention; perceived online bullying	Phone use sharply down (teacher-reported personal use in class 61% -> 13%); test scores ~0 on average (rules out >0.007 student SD), high-school math +0.024 SD, middle school slightly negative; suspensions +~11% in year 1 then fade; subjective well-being -0.2 SD at adoption recovering to +0.16 SD by year 2; no effect on attendance, attention or online bullying (NBER WP 35132)
Lichand, Moreno-Louzada, da Costa & Gentzkow (2026)	Brazil (Rio de Janeiro)	DiD contrasting middle schools without strict pre-existing rules (treated) vs schools already strict (control) around the Aug 2023 municipal ban (school-wide from Feb 2024); principal survey + admin data	standardized test scores; in-school phone use	Test scores +0.06 SD in treated relative to control schools; in-school phone use fell substantially
Kohnstamm Instituut & Oberon (for Dutch Ministry OCW) (2025)	Netherlands	Descriptive monitoring: school-leader surveys and group interviews; no comparison group	leader-reported concentration, social climate, learning, well-being, enforcement	Secondary: 75% report better concentration, 59% better social climate, 28% better learning; share seeing non-educational phone use as a problem 62% -> 29%; some teachers report more physical bullying/disruption and enforcement workload. Primary: mostly no effect; 23% see better well-being/climate

Study	Setting	Design	Outcome domain	Headline
OECD, PISA 2022 Results Volume II (ch. 5) and 'Managing screen time' (2024) (2023)	OECD/partner countries (cross-national)	Cross-sectional associations using PISA 2022 student/school questionnaires	self-reported distraction by digital devices; in-school phone use; notification habits; performance	About one in three OECD students attend schools banning phones; students in ban schools less likely to report distraction, but 29-30% of them still use a phone several times a day (+21% every day/almost) so bans are often unenforced; in 9 systems bans are concentrated in lower-performing schools; in some systems banned students are less likely to turn off notifications at night. No causal claim.
Bottger & Zierer (2024)	Multi-country (review)	Rapid review + meta-analysis of 5 quantitative studies	social well-being (incl. bullying); academic performance	Pooled d = 0.162 (p<0.05); larger for social well-being than for achievement (Education Sciences 14(8), 906)
Campbell, Edwards, Pennell, et al. (2024)	Multi-country (review)	Scoping review	academic outcomes; mental health and wellbeing; cyberbullying	Mixed: 4 studies report achievement gains (2 restricted to low-achieving/low-SES samples); as many null or negative; no consistent effect on mental health or cyberbullying (J. Psychologists and Counsellors in Schools)

Two readings. On mental health specifically, the strongest evidence is Norway's staggered middle-school bans (fewer psychological consultations for girls, less bullying, no effect for boys), which is the girl-specific pattern H1 predicts. The English one-time comparison shows no effect on well-being, and the 2026 U.S. Yondr-pouch study finds well-being dips at adoption and recovers by year two with no change in online bullying. Achievement effects are small and positive where found, and vanish in Sweden. So bans are a plausible policy with mixed evidence rather than a settled one, and the mental-health effect, where it appears, is on girls.

County social-media use and sadness

State	County item	Outcome	County-grade-years	Across counties (SE)	Within county and grade (SE)
WA	social media daily	sad hopeless	172	+0.22 (0.04)	-0.09 (0.16)
WA	social media daily	considered suicide	172	-0.01 (0.03)	-0.15 (0.10)
WA	screen time 3plus	sad hopeless	172	+0.25 (0.06)	+0.01 (0.08)

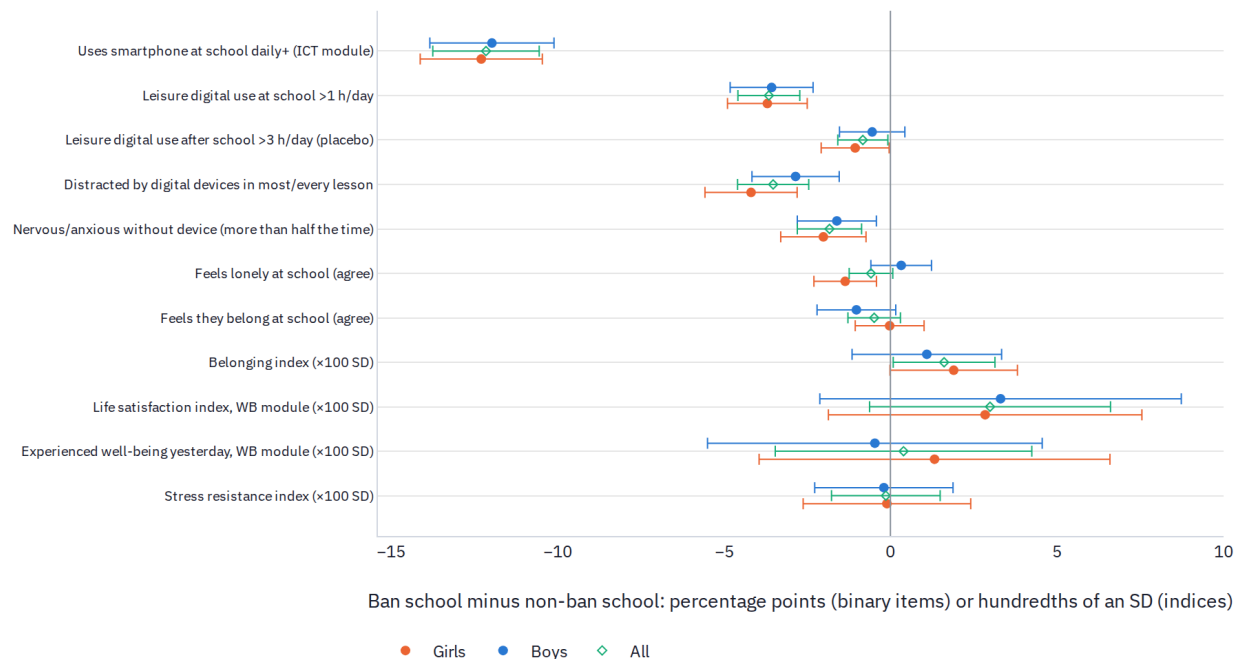
State	County item	Outcome	County-grade-years	Across counties (SE)	Within county and grade (SE)
WA	screen time 3plus	considered suicide	172	+0.03 (0.03)	+0.04 (0.07)
ME	screen time 3plus	sad hopeless	43	+0.07 (0.11)	-0.09 (0.10)
ME	screen time 3plus	considered suicide	43	+0.22 (0.12)	+0.13 (0.13)

Percentage points of outcome per percentage point of the county item, on county-by-grade cells so that item and outcome cover the same grades; year fixed effects (across) or county-by-grade and year fixed effects (within); SE clustered by county. Washington grades 8, 10 and 12 in 2023 and 2025; Maine's all-grade rows 2017–2021.

In Washington, county-grade groups where more students use social media daily report more sadness (0.22 pp per pp, SE 0.04), and the same holds for screen time of three hours or more (0.25, SE 0.06). Both items and sadness fell between 2023 and 2025 (daily social media 75% → 73%, sadness 31% → 27%), but the county-by-county changes do not line up: within the same county and grade, the two-year change in sadness is unrelated to the change in either item (−0.09, SE 0.16, and 0.01, SE 0.08). Maine shows nothing in either direction. So the cross-sectional link is there, as in the individual-level studies, while the only two-wave change available does not reproduce it.

PISA 2022: school phone rules and well-being within countries

PISA 2022 asked every participating school whether "the use of cell phones is not allowed on the school premises" and asked students about loneliness and belonging at school, digital-device habits, and (in 13 systems) life satisfaction. Merging the two files gives 568,374 students in 19,885 schools across 80 systems; 41% of students attend a ban school, from 7% in Finland to 96% in Greece. The comparison below is within country: ban versus non-ban schools with country fixed effects, controls for socio-economic status, sex, grade and school type, student weights, and standard errors clustered by school.



Binary items in percentage points; PISA indices in hundredths of a standard deviation. The ICT module (smartphone use at school) covers 51 systems; the well-being module (life satisfaction, experienced well-being) covers 13–14.

Outcome	All	Girls	Boys	Mean
Uses smartphone at school daily+ (ICT module)	-12.15 (0.82)	-12.29 (0.94)	-11.97 (0.95)	55.3 pp
Leisure digital use at school >1 h/day	-3.65 (0.47)	-3.70 (0.61)	-3.57 (0.64)	33.4 pp
Leisure digital use after school >3 h/day (placebo)	-0.83 (0.38)	-1.06 (0.52)	-0.55 (0.50)	27.7 pp
Distracted by digital devices in most/every lesson	-3.52 (0.55)	-4.19 (0.71)	-2.85 (0.67)	28.1 pp
Nervous/anxious without device (more than half the time)	-1.83 (0.49)	-2.01 (0.65)	-1.61 (0.61)	22.5 pp
Feels lonely at school (agree)	-0.58 (0.33)	-1.36 (0.48)	+0.32 (0.46)	19.7 pp
Feels they belong at school (agree)	-0.49 (0.40)	-0.02 (0.53)	-1.02 (0.60)	77.5 pp
Belonging index (SD units)	+0.02 (0.01)	+0.02 (0.01)	+0.01 (0.01)	-0.2 SD
Life satisfaction index, WB module (SD units)	+0.03 (0.02)	+0.03 (0.02)	+0.03 (0.03)	-0.0 SD
Experienced well-being yesterday, WB module (SD units)	+0.00 (0.02)	+0.01 (0.03)	-0.00 (0.03)	0.1 SD
Stress resistance index (SD units)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.1 SD

Three findings. First, bans change behaviour, partly: daily smartphone use at school is 12 points lower in ban schools (from a 55% base, so 43% of students in ban schools still use one daily), leisure screen use at school falls, and after-school use is unchanged, which is the right placebo

comparison pattern. Second, the well-being differences run in the direction the Norwegian study found and are concentrated in girls: less nervousness without a device, 1.4 pp less loneliness at school for girls (none for boys), and +0.02 to +0.03 SD on belonging and life satisfaction. Third, students in ban schools report less classroom distraction by digital devices (3.5 pp fewer say they are distracted in most or every lesson; 4.2 pp for girls, 2.9 for boys), in line with the lower in-school use. All of this is a comparison across schools at one point in time: schools choose whether to ban, and nothing here says how a ban school would have looked without its ban, so the small favorable girl-specific differences could reflect which schools ban as much as what bans do.

What this shows. The only public, multi-country microdata on school phone rules show a modest, girl-specific association between bans and lower loneliness and device anxiety, alongside clear evidence that bans reduce in-school use and reported classroom distraction. Because the comparison is across schools at one time, it cannot separate the effect of a ban from the kind of school that adopts one. Together with Norway's result, that is enough to make the 2024–25 U.S. state bans the most valuable natural experiment now pending in this project; the 2025 YRBS and the 2025–27 state-survey waves are where it will be tested, using the approach set out above.

How this sits with Haidt's and Twenge's evidence

The [Anxious Generation research pages](#) and the [collaborative reviews](#) present a large body of evidence for a link between social media, smartphones and adolescent distress, much of it from Jean Twenge's work with the Monitoring the Future, YRBS and Millennium Cohort surveys (the model-curve analyses, the "heavy users twice as likely" comparisons, the worldwide loneliness series) and from the natural experiments the book relies on. This section sets that evidence beside the pilot's results. The short version: the pilot reproduced the individual-level association that literature rests on, in new settings; what it did not find is any population-level timing effect, and the studies that establish causation on those pages are not tests of the 2012 smartphone shock either.

What that evidence base contains

Leg	What it is	Typical result
Cross-sectional use–symptom relation (Twenge and colleagues)	Heavy vs light users in national surveys (MTF, YRBS, Millennium Cohort, UK and U.S. panels)	Heavy users about twice as likely to report depression or low well-being; adjusted odds ratios 1.7–3; stronger for girls

Leg	What it is	Typical result
Longitudinal	Use at t predicts symptoms at t+1	Small effects; the contested ground with Orben, Przybylski, Odgers and Ferguson
Experiments	Deactivation and abstinence trials, mostly adults	A few hundredths to about 0.2 SD improvement
Natural experiments	Facebook college rollout 2004–06 ; four broadband instrumental-variable studies (Germany, Italy , Spain, UK) reviewed by McClean, Rausch and Haidt ; and Elaine Guo's British Columbia study (University of Toronto dissertation, 2023; circulated as a job-market paper in 2022), which links twenty years of provincial student records to wireless-coverage maps	Facebook arrival: about a tenth of an SD on a mental-health index, larger for women and the predisposed. Italy: first DSL access 2001–2013, instrumented by distance to old telephone exchanges, raised diagnosed disorders in cohorts exposed before 20 by 0.08 SD. British Columbia: girls in mountain communities covered by high-speed fixed wireless gained about 1 percentage point of severe mental-health designations relative to boys from 2014 on, 90% of their baseline rate; moderate designations and placebo disabilities unchanged
Timing coincidence	The mood-disorders review and the World Happiness Report chapter	Flat through the recession, inflection in 2012, girls first, then the Anglosphere and Western Europe; PISA life satisfaction down in 40 of 47 countries 2015–2018

Where the pilot agrees

- GSHS 2022–23, five countries: girls on a phone 3+ hours a day report loneliness +8 pp and suicidal ideation +10 pp, rising with hours. That is an odds ratio of roughly 1.5 to 1.7, close to the magnitude the reviews call "twice as likely".
- Nepal 2022 self-reported PHQ-9 and GAD-7 point the same way for 15–19-year-olds.
- The U.S. timing facts match: NSDUH turns up in 2011–12, YRBS sadness later, girls dominate in every source, the ACS 65+ placebo comparison falls.
- The NTIA smartphone exposure level predicts YRBS girls' sadness within states (suggestive; it did not replicate at county level).

Where the pilot shows no effect, and why that is not a contradiction

1. **LTE onset, at market and county level.** The natural experiments on the Anxious Generation pages identify off *first* wired-broadband access in the 2000s, or Facebook arriving where there was none. LTE in 2011–2013 was a speed increment on top of 3G smartphones and mobile Facebook that already existed; most large metros got it within 24 months of each other; and, as Haidt himself argues, the harm is a network effect, so a teen in a late-LTE market was exposed through peers and national platforms anyway. Each of those pushes a market-level estimate toward zero. The absence of an LTE effect says the 2010s shock is not identified by U.S. geography, which is different from saying there was no shock. None of the studies on those pages tests the 2012 smartphone transition with a rollout design either; the causal leg of the case is about earlier technologies.

The exception, not on those pages, is Elaine Guo's British Columbia study, the closest design to this one and the one result that reaches into the Instagram years. It compares teenagers in mountain communities that received high-speed wireless internet (25 Mbps, delivered over the cellular network) with teenagers in the mountain communities that did not, before and after visual social media arrived, girls against boys, with the outcome taken from the province's school records: a severe mental-health designation that requires a clinical diagnosis. From 2014 the covered girls gain about one percentage point of severe designations, 90% of their baseline rate, with nothing for boys, nothing for moderate designations, and nothing for disabilities that social media cannot plausibly affect. Three differences explain how it can find what the LTE design cannot. Its exposure is a step from slow or no home internet to broadband in rural towns, not a speed increment on top of 3G in metros. Its outcome is a clinical designation, not a disability screener. And its time variation is calendar time after Instagram rather than rollout order, because the coverage map is a 2019 snapshot and the thesis notes that earlier coverage history was not available; that is also its weakness, since communities covered by 2019 may have differed in when they got service. Read together, the two studies say the 2010s shock shows up where access itself was new and the outcome is clinical, and not where access was a speed upgrade and the outcome is a screener. That is a reason to point the U.S. design at rural broadband gaps and at school-recorded or clinical outcomes, which is where the next-steps section already sends it.

2. **Country-level GSHS with internet penetration, mobile-broadband subscriptions and 3G/4G coverage.** These are the same kind of evidence as Vuorre and Przybylski's 72-country Facebook study, which Haidt's team criticizes for aggregating groups and technologies. That criticism applies to the pilot's country-level no-effect results too; they should not be leaned on.
3. **MICS, 68 surveys.** The one genuinely new negative result. Caregiver-reported anxiety and depression for 10–17-year-olds show no relation to cluster connectivity or to a girl's own daily internet or phone use, and the one positive cluster coefficient is matched by the 5–9-year-old placebo comparison. Two readings are both live: the self-report association is partly reverse causation and shared method (distressed teens use phones more and report both), or caregivers do not see internalizing symptoms (parent–child agreement on internalizing problems is known to be low). The pilot cannot separate these, which is exactly why the GSMA coverage-timing design on the adolescent-reported rounds is the next step.

Two things the pilot adds that those pages do not address

- Schoolwork pressure rose across roughly 45 HBSC countries after 2010, not only in the U.S., and complaints rose more where pressure rose more. It behaves as a stable correlate rather than the cause of the break, but Gray's premise is not a U.S. peculiarity, and the review's dismissal of school pressure rests on one UK exam study.
- The U.S. state surveys show a large reversal after 2021 (Washington's 2025 sadness rate is below its 2010 level) with no comparable fall in phone use. A pure exposure level story has to

explain that; so does a school-pressure story.

In sum. The Anxious Generation evidence establishes a consistent individual association and a plausible causal mechanism from earlier internet rollouts, and Guo's British Columbia study extends that mechanism to 2010s wireless coverage on a clinical outcome. The pilot confirms the association, including in phone-first countries, and shows that the population-level version of the claim is not recoverable from the geography of 4G or adoption in public data, and does not appear in caregiver-reported outcomes. Someone convinced by Haidt should conclude, a little more than before, "the effect, if causal, is not visible in rollout timing and needs adolescent-reported exposure to detect"; a skeptic should conclude "the association replicates everywhere self-report is used." The decisive test remains the one neither side has run: adolescent-reported exposure against local coverage timing.

Are we closer to an answer?

Closer than the books, in three specific ways, and not close on the question the books actually argue about.

Hypothesis	Where it stands after the pilot
H1 · Phones and social media	The individual association is consistent and reappears in places the book's evidence does not reach: girls who use phones heavily report more loneliness and suicidal thoughts in Ghana, India, Jamaica, Morocco and Nepal as in the United States, with the same sex pattern and exposure level gradient. But every test that uses <i>timing</i> or <i>adult-measured</i> exposure shows no effect: 4G rollout across U.S. metros and counties, adoption exposure level at county level, cluster connectivity in 68 MICS surveys, coverage in 39 GSHS countries. Either the effect is real but travels through peers and platforms so evenly that geography cannot see it, or the association is largely reverse causation and shared causes. The public data cannot yet distinguish these; school phone bans (PISA, Norway) lean modestly toward a real effect for girls.
H2 · School pressure	Gray's premise holds up better than Haidt's collaborative reviews allow: schoolwork pressure rose after 2010 in most HBSC countries, not only the U.S., and U.S. homework time rose about 40%. But pressure behaves as a stable correlate of complaints rather than the cause of the break, U.S. accountability shocks show no effect, and the age pattern lacks the high-stakes-years signature. The real limitation is measurement: "pressure" is one survey item, while phone use can at least be counted in hours, so H2 is harder to test than H1 rather than refuted.

Hypothesis	Where it stands after the pilot
H3 · Interaction	Not supported through state policy crossed with 4G timing, but the sharper version, phones during school hours on top of high-stakes academics, fits the ban evidence best: bans cut in-school use, girls in ban schools report less loneliness and device anxiety, the association is larger for students under high schoolwork anxiety (not significantly), Norway's bans reduced girls' psychological consultations, and U.S. sadness moved inversely to school prohibitions. It is the hypothesis the pending U.S. ban studies and the HBSC microdata can test directly, and the one that would reconcile the two books.
H0 · Measurement	Rejected as the main story: the rise is present in self-report, caregiver-report and diagnostic measures, the 65+ placebo comparison moves the other way, and it is concentrated in girls everywhere. Reporting and diagnosis-seeking plausibly add to it (the rise is steepest among affluent teens), but do not create it.

Three findings that neither book contains. First, the rise was broad and nearly simultaneous across places with very different connectivity and schooling histories, which is why coarse designs cannot separate the theses and why both camps can find support in national trends. Second, the post-2021 reversal in the U.S. state surveys is as large as the 2013–2021 rise, without a comparable fall in phone use, and it coincides with schools re-tightening phone rules; a pure exposure level story and a pure school-pressure story both have to explain it. Third, the association between phone use and distress appears only when the adolescent reports both her own use and her own state; caregiver-reported symptoms in 68 countries show nothing. That is the fact the next phase has to resolve.

What would settle it, and what this project should do next.

- Follow the 2024–26 U.S. state and district phone bans through the 2025–27 YRBS and state-survey waves, by sex, for at least two years after each ban, with the 65+ comparison group retained. The policy table and the model are ready; the first data wave is the 2025 YRBS.
- License GSMA coverage polygons and apply for the MICS and DHS cluster coordinates, then test the adolescent-reported phone-hours items (GSHS 2022–23, MICS7) against local coverage timing. This is the one design that measures exposure on the adolescent herself while using something other than her own choices to explain it.
- Apply for the HBSC microdata to run the concentration test (are complaints increasingly concentrated among high-pressure students?) and the student-level social-media × pressure interaction, which is the only available test of H3.
- Stop extending the ACS market event study and the census cognitive item; they have done what they can. Defer the restricted-data route (the Census Bureau research data centers) until a state-survey or ban result justifies it.

Reproducibility: scripts in `scripts/`, tables in `results/tables/` of the project repository. Data: IPUMS USA (ACS 1-year 2005–2024), IPUMS ATUS 2003–2024, IPUMS International (ke2009a, ke2019a, mx2010a, mx2020a), IPUMS CPS; public NSDUH, NHANES, YRBS, HBSC, GSHS, NTIA files as noted in each section.