The Effects of School Smartphone Restrictions in the United States

Hunt Allcott[*] Jason Baron[†] Thomas Dee[‡] Angela Duckworth[§]
Matthew Gentzkow[¶] Brian Jacob[‖]

March 24, 2026

Abstract

In response to concerns that smartphones may disrupt learning, attention, in-person interaction, and student well-being, schools across the U.S. have increasingly restricted their use during the school day. We evaluate one type of restriction – lockable phone pouches – using nationwide data on U.S. middle and high schools drawn from teacher, parent, and student surveys, smartphone GPS pings, standardized test scores, and school administrative records, combined with sales records from the largest lockable phone pouch provider. Survey evidence indicates that teachers strongly prefer restrictive policies and perceive these policies to substantially reduce in-school smartphone use. Using a staggered difference-in-differences design, we find that pouch adoption significantly reduces smartphone activity as measured by GPS pings but does not measurably improve test scores, attendance, or student-reported measures of classroom attention and online bullying. In contrast, adoption is associated with increased disciplinary incidents and dynamic effects on student-reported well-being, with short-run declines followed by later improvements.

[*]Stanford University and NBER. Email: allcott@stanford.edu
[†]Duke University and NBER. Email: jason.baron@duke.edu
[‡]Stanford University and NBER. Email: tdee@stanford.edu
[§]University of Pennsylvania. Email: duckwort@wharton.upenn.edu
[¶]Stanford University and NBER. Email: gentzkow@stanford.edu
[‖]University of Michigan and NBER. Email: bajacob@umich.edu

1 Introduction

The impacts of widespread smartphone use on young people are hotly debated. An important part of this debate concerns smartphone use in schools. Many argue that phone access during the school day disrupts attention and learning, weakens in-person relationships, and contributes to rising anxiety and depression among students, leading to calls for phone-free schools (Haidt, 2023). Policymakers have responded by sharply restricting smartphone use in schools: as of 2026, roughly two-thirds of U.S. states and more than half of countries worldwide have enacted legislation limiting student phone access (Prothero, Langreo and Klein, 2026; D'Addio, 2025). At the same time, others emphasize variation in implementation, the potential educational benefits of smartphones, the possibility that enforcing restrictive rules may itself introduce classroom disruptions, and the limited causal evidence on the effects of smartphone restrictions (Rahali, Kidron and Livingstone, 2024; Hess, 2026). There is therefore considerable debate about whether restricting smartphone use in schools improves academic performance, student behavior, and well-being.

We provide the first nationwide, quasi-experimental analysis of the effects of smartphone restrictions in U.S. schools, focusing on one well-defined and rapidly expanding intervention: lockable phone pouches designed to limit students' access to smartphones during the school day. Under this approach, students place their phones in magnetically sealed pouches upon arrival at school and are not allowed to access them until dismissal. The intervention has been widely adopted, with nearly 5000 schools using pouches as of 2026.

By introducing a physical-access constraint rather than relying solely on classroom rules or teacher discretion, this policy is intended to create a more consistently implemented restriction on in-school phone use. The distinction between rule-based and physically binding restrictions has become central in the policy debate, with some observers arguing that physically binding, bell-to-bell restrictions—such as those enabled by lockable pouches—are the most effective approach for limiting in-school phone use (Haidt, 2023). Our focus on lockable pouches therefore allows us to study one of the most stringent forms of smartphone restriction currently implemented in U.S. schools. At the same time, implementation may vary across schools, and reports from some districts indicate that students sometimes circumvent the pouches, suggesting that even physically binding restrictions may not fully eliminate in-school phone access (Kircher and Holtermann, 2026).

We draw on multiple nationwide data sources for our analysis. To measure school-specific policies, we fielded a nationwide teacher survey (henceforth, "NTS") on schools' current and past phone policies. To date, approximately 68,000 teachers have completed the survey. In addition, we partnered with Gallup to field nationally representative survey questions to

roughly 1800 parents of middle and high school students and to their children, allowing us to characterize attitudes and expectations regarding school phone bans. We complement this with confidential sales data from Yondr, the leading provider of lockable smartphone pouches. We use smartphone GPS data from two different providers to measure the impact of school restrictions on smartphone use. We combine all of this with school-by-year outcome data for middle and high schools, including test scores (middle schools only), administrative data on attendance and disciplinary incidents, and student surveys administered by Panorama Education. Across these outcome domains, we publicly pre-registered primary outcome definitions prior to estimating treatment effects.

We begin by presenting descriptive evidence on school smartphone policies in the U.S., drawing on data from the NTS and Gallup surveys of parents and their children, fielded between April 2025 and February 2026. Although many states have recently adopted legislation requiring restrictions, implementation decisions are typically delegated to districts or individual schools, resulting in substantial heterogeneity in practice. Teachers report the policies in place at their school at the time of the survey and, for schools that experienced a recent change, retrospectively describe prior conditions. Policies range from no formal restrictions to moderate approaches such as "no-show" rules, in which students may carry their phones but must keep them out of sight, to stringent restrictions including lockable pouches, lockers-only policies, centralized collection, or complete bans on campus. Across this spectrum, teacher satisfaction increases and in-class phone use declines with policy strictness. Among schools that adopted lockable phone pouches, retrospective reports indicate particularly large increases in satisfaction and declines in in-class phone use following adoption. At the same time, parents expect lockable pouches to increase test scores, improve student relationships, and they express support for such policies, whereas students tend to oppose them and expect little change in test scores or peer relationships, though both groups anticipate improvements in mental health.

We then turn to estimating the causal effects of lockable phone pouches on objectively measured and student-reported outcomes, which presents several empirical challenges. Most importantly, adoption is not randomized: schools that adopt differ systematically from non-adopting schools along observable dimensions such as size, urbanicity, and student demographics. Our research design compares changes in outcomes over time in schools that adopt lockable pouches to changes in schools that are similar along observable baseline characteristics but do not adopt. A key strength of our setting is the availability of a large pool of never-treated schools, which allows us to construct a well-matched comparison group using baseline school characteristics. We examine pre-adoption trends and show that treated and comparison schools exhibit similar outcome trajectories prior to adoption.

We implement this design using the difference-in-differences framework developed by Callaway and Sant'Anna (2021), which is well suited to settings with staggered treatment adoption and heterogeneous treatment effects. The estimator identifies group-time average treatment effects by comparing adopting schools to appropriate comparison schools while conditioning on baseline covariates. These covariates, measured prior to treatment, include enrollment, racial composition, charter status, neighborhood income-to-poverty quartiles, urbanicity, and grade levels. We use the doubly robust implementation of this estimator, which combines inverse probability weighting and outcome regression. We also show that our results are robust to conventional two-way fixed effects specifications that incorporate the same baseline controls interacted with year fixed effects. Because our data are at the school level, we standardize outcomes using the standard deviation of school-level means (what we call "school standard deviations"), which differ from individual-level standard deviations commonly reported in the education literature.

Our goal is to estimate the effect of Yondr adoption relative to schools' prior policies and to schools that did not adopt Yondr during the study period. The NTS helps clarify what these treatment and control conditions represent in practice for the subset of Yondr schools that overlap with the survey. Most adopting schools previously operated under no formal policy or relatively weakly enforced rules such as "no-show" policies. While control schools were not policy-inactive, only a small share adopted similarly stringent restrictions during the period when Yondr adoption accelerated. Our estimates therefore capture the effect of adopting Yondr relative to the range of policies typically observed in schools. In addition, the NTS data allow us to explore alternative counterfactuals; for example, by restricting the control group to schools reporting no recent phone policy changes.

We begin by examining the "first-stage" effect of Yondr adoption on in-school smartphone use. Using GPS-based measures of device activity, we find that adoption leads to large and persistent reductions in smartphone use during school hours. In the year of adoption, net smartphone activity declines by approximately 19 percent relative to pre-adoption levels, with the magnitude of the decline increasing to roughly 32 percent by the second year after adoption. These magnitudes should be interpreted with caution, however, as the GPS-based measures are imperfect proxies for phone activity—for example, they do not distinguish between student and staff devices and capture location pings rather than direct measures of screen time. We discuss these measurement considerations in more detail below.

Having established that lockable phone pouches substantially reduce in-school smartphone use, we next examine whether these changes translate into differences in student outcomes. We begin with academic achievement. In this draft, we focus on middle school students, where standardized assessments are consistently administered across states and

smartphone use is widespread.[1] Averaging across post-adoption periods, the estimated treatment effect on test scores is -0.025 (SE = 0.012) school-level standard deviations. Using standard assumptions about the relative size of school- and student-level standard deviations for test scores (What Works Clearinghouse, 2020), this corresponds to roughly -0.012 student-level test-score standard deviations. For context, this implies a decline of less than one-half of a percentile point in the test-score distribution (von Hippel, 2025), roughly one-tenth the achievement gains associated with having a higher value-added teacher (Chetty, Friedman and Rockoff, 2014), and about 4 percent of the typical annual learning gains observed among students at this age (Hill et al., 2008).

Event-study estimates indicate that this average effect reflects near-zero impacts in the year of adoption followed by modestly negative point estimates in the subsequent two years. Cohort-specific event studies suggest that the negative test score effects are largest for earlier adoption cohorts and smaller for later cohorts, with the 2025 adoption cohort exhibiting a positive estimate in its year of adoption, though this estimate is not statistically significant. It is important to note, however, that the relatively short panel limits our ability to completely separate dynamic effects from cross-cohort heterogeneity.

We next examine impacts on school attendance. Averaging across post-adoption periods, the estimated effect on the attendance rate is -0.052 (SE = 0.096) percentage points and is statistically indistinguishable from zero. The 95 percent confidence interval rules out improvements larger than 0.136 percentage points. Relative to a mean attendance rate of 92 percent, this corresponds to a change of less than 0.1 percent. While event-study estimates indicate that effects become modestly more positive over time, they remain small in magnitude and statistically indistinguishable from zero in all post-adoption years. We find similarly precise and small effects for chronic absenteeism.

In contrast to test scores and attendance, Yondr adoption is associated with a statistically significant and sharp increase in disciplinary incidents. Averaging across post-adoption periods, the estimated effect on a standardized discipline index is 0.067 (SE = 0.023) school-level standard deviations. Event-study estimates show a clear and statistically significant increase in discipline in the year of adoption. However, this effect is short-lived: estimates decline substantially in subsequent years and are close to zero by the first and second years after adoption, though less precisely estimated.

We also examine student self-reported outcomes at the school level using survey data from Panorama. We focus on three primary outcomes: a classroom attention index, a measure of students' perceived likelihood of online bullying, and a subjective well-being

[1] In ongoing work, we plan to extend the analysis to high school test scores by assembling comparable data from state education agency websites.

(SWB) index. These measures capture three central channels through which smartphone restrictions are hypothesized to operate—students' in-class focus, their emotional status, and peer interactions that may extend into digital settings. Because Panorama surveys are administered at the discretion of schools, survey data are available for approximately 8 percent of school-year observations; these estimates are therefore based on a smaller and potentially selected subset of schools and should be interpreted in that context.

We find no evidence that Yondr adoption measurably improves self-reported classroom attention. Averaging across post-adoption periods, the estimated effect on the classroom attention index is -0.11 (SE = 0.074) school-level standard deviations. Confidence intervals rule out even modest positive effects. Event-study estimates are negative and statistically significant in some post-adoption years and near zero in others, with no consistent pattern over time and no evidence of sustained improvements.

We similarly find no detectable effect on students' perceived likelihood of online bullying. Averaging across post-adoption periods, the estimated effect on the standardized online bullying measure is -0.044 (SE = 0.096) school-level standard deviations. The index is reverse-coded so that higher values correspond to lower perceived bullying risk. Event-study estimates show no evidence of differential pre-trends and no statistically significant changes following adoption, with coefficients that fluctuate around zero but display no systematic pattern over time.

Finally, we examine impacts on SWB. Averaging across post-adoption periods, the estimated effect on the SWB index is -0.365 (SE = 0.213) school-level standard deviations. However, this average masks substantial heterogeneity over time. In the year of adoption, the SWB index declines by approximately 0.67 (SE = 0.243) school-level standard deviations, before becoming positive by the second post-adoption year, on the order of 0.41 (SE = 0.178) school-level standard deviations. Using standard assumptions about the relative size of school- and student-level standard deviations for behavioral outcomes (What Works Clearinghouse, 2020), this translates to roughly a 0.22 student-level standard deviation decline in the year of adoption and an increase of 0.13 student-level standard deviations by the second post-adoption year. For comparison, Allcott et al. (2020) find that deactivating Facebook for four weeks increased subjective well-being by 0.09 standard deviations.

Taken together, our findings indicate that restricting smartphone use may not generate measurable improvements in academic outcomes such as standardized test scores, attendance, or self-reported classroom attention. For test scores in particular, the estimated effects are small in magnitude and generally negative. Instead, restrictions appear to meaningfully alter aspects of the school environment, for example by increasing disciplinary incidents and possibly shifting students' reported well-being.

Several mechanisms could plausibly explain these patterns. For example, the increase in disciplinary incidents may partly reflect enforcement of new phone rules (Figlio and Ozek, 2025). The fact that this increase is concentrated in the year of adoption is consistent with a short-run implementation effect, as schools introduce and enforce new restrictions. The small negative test score effects we observe could arise if enforcement diverts teacher time and attention away from instruction (Hess, 2026). In addition, restricting smartphone access may reallocate student attention toward in-person interaction; if smartphones previously absorbed attention that would otherwise have manifested in peer interaction (Allcott et al., 2020), their removal may increase socialization and conflict during the school day (Kamenetz, 2025; Jacob and Lefgren, 2003), particularly during the transition to a new policy environment. These mechanisms may also help explain the dynamic pattern in subjective well-being. The introduction of a binding restriction represents a discrete change to students' daily routines and social environment, which may impose short-run adjustment costs. The subsequent rebound is consistent with adaptation over time and potential longer-run gains from increased in-person engagement. Of course, these interpretations are necessarily speculative, and additional data would be needed to distinguish among them.

Our analysis has several important limitations. First, we do not have a randomized experiment, and policy adoption could be correlated with confounding trends. However, we find no evidence of differential pre-adoption trends, no systematic changes in student composition, and similar results across alternative specifications. Second, our outcome measures capture only certain dimensions of student performance and school climate. While standardized test scores and validated survey instruments are widely used in the education literature, they may not reflect all potential effects of smartphone restrictions. Third, because smartphone restrictions have expanded only recently, we observe outcomes for at most three years beyond adoption for most cohorts; longer-run effects remain an open question. Fourth, while we focus on a specific and well-defined form of smartphone restriction, schools also use other approaches, including centralized collection, locker storage requirements, or complete campus-wide bans, which may differ in both implementation and effects.

Related Literature. Our paper's primary contribution is to provide the first large-scale, nationwide quasi-experimental evidence on the effects of school smartphone restrictions in the United States. We build on a small and mixed literature on school smartphone bans, most of which comes from outside the U.S. Beland and Murphy (2015) find that school-level bans in England increased test scores, particularly among lower-achieving students, whereas Kessel, Hardardottir and Tyrefors (2020) find no academic effects of similar bans in Sweden. Abrahamsson (2024) finds that bans in Norwegian middle schools improved girls' GPA, reduced mental health consultations, and lowered bullying. In contrast, Shi and Villarroel

(2026) find that classroom-level restrictions in Chile modestly reduced in-class phone use but did not measurably improve test scores. Finally, Sungu, Choudhury and Bjerre-Nielsen (2025) find that mandatory in-class phone collection in Indian higher education settings improved grades, with no significant changes in overall student well-being.

A recent study of Florida's statewide smartphone ban is perhaps most relevant for our context. Using administrative student records and smartphone GPS data, Figlio and Ozek (2025) evaluate a district-wide "no-show" restriction requiring phones to be silenced and kept out of sight throughout the school day in a large urban district in Florida. For identification, they compare schools within the district with higher versus lower pre-ban smartphone activity. They document short-run increases in disciplinary incidents followed by improvements in test scores and attendance in the second year after the ban.

We complement this important work by studying a particularly stringent and physically binding restriction across a nationwide sample of schools. The national scope provides substantial statistical power to examine treatment effect heterogeneity across school contexts and student populations, while strengthening the external validity of our findings. The presence of a large pool of never-treated schools also enables the construction of well-matched comparison groups. Beyond administrative outcomes and independently measured smartphone activity, we leverage original survey data to enrich the analysis. The NTS allows us to characterize the policy environment in both treated and control schools, document enforcement practices, and measure teacher satisfaction. Separately, student survey data from Panorama allow us to examine classroom attention, subjective well-being, and perceived online bullying—outcomes central to the debate over smartphone restrictions. Together, this approach provides a comprehensive assessment of how restrictive smartphone policies affect academic performance, behavior, and students' in-school experiences.

Our paper is also related to research on the effects of personal digital devices and social media. Experimental and quasi-experimental studies show that reducing access to social media platforms can improve subjective well-being (Allcott et al., 2020), while the staggered rollout of Facebook across U.S. colleges worsened student mental health and academic performance (Braghieri, Levy and Makarin, 2022). Other work documents that intensive mobile app usage negatively affects academic performance and early labor market outcomes, with substantial peer spillovers (Barwick et al., 2026). This literature primarily studies adults or college students. We examine adolescents in K–12 schools—a group at an earlier developmental stage for whom self-regulation and peer influences may operate differently.

Sections 2 through 7 present, respectively, a description of Yondr pouches and the conceptual framework; the data; descriptive facts on school smartphone policies; the difference-in-differences empirical strategy; the empirical estimates; and the conclusion.

2 Background

2.1 Yondr Lockable Pouches

Yondr pouches are fabric cases equipped with magnetic locking mechanisms that schools use to restrict students' access to their smartphones during school hours.[2] Upon arrival at school, students place their phones inside a pouch, which is then magnetically sealed. The pouch remains locked throughout the school day, preventing access to the device while allowing students to retain physical possession of it. At the end of the school day—or during emergencies—the pouches can be unlocked using magnetic bases stationed around the school.

Although schools can restrict student smartphone use in many ways—ranging from "no-show" policies (under which students may carry phones but are prohibited from displaying or using them) to classroom phone caddies—the use of Yondr lockable pouches represents a particularly well-defined and high-powered intervention. First, our focus on Yondr allows us to leverage detailed administrative data from the company, yielding a treatment measure with high implementation fidelity. This stands in contrast to policy-based bans that may be unevenly enforced or weakly binding in practice, especially in the absence of a physical enforcement mechanism. Second, the design of Yondr pouches facilitates a uniquely sharp treatment contrast. By physically preventing access to devices once sealed, the pouches substantially reduce students' ability to use smartphones during the school day and lower the enforcement burden on teachers and administrators. As a result, schools adopting Yondr are likely to experience larger reductions in in-school phone use than schools relying on less restrictive approaches. Estimates based on Yondr adoption therefore provide a strong test of the effects of binding smartphone restrictions. Third, the national scope of the Yondr data affords statistical power and enhances the external validity of our findings. Finally, the breadth of the sample enables us to examine treatment effect heterogeneity across a wide range of school contexts, including variation in student demographics.

2.2 Conceptual Framework and Hypotheses

School smartphone restrictions are often justified on the grounds that constant access to digital devices undermines sustained attention, fragments learning, weakens in-person relationships, and contributes to rising anxiety and depression among students (Haidt, 2023). Our goal is to evaluate whether restricting smartphone access in schools achieves these intended objectives. To do so, we examine outcomes that map onto these underlying constructs.

[2]When the company was founded in 2014, it initially focused on providing pouches for concerts and music venues.

We study standardized test scores and attendance as measures of academic performance and school engagement. While standardized assessments capture only certain dimensions of learning and may not reflect broader skills or forms of learning that smartphone restrictions could plausibly influence (Duckworth, Quinn and Tsukayama, 2012), we use them for several reasons. First, standardized test scores can be consistently measured across states and over time. Second, they are widely used in the economics of education as summary measures of achievement. Finally, prior research shows that teachers' impacts on student test scores are correlated with students' longer-run outcomes such as teenage pregnancy, college attendance, and earnings (Chetty, Friedman and Rockoff, 2014). We also study student-reported classroom attention as a measure of in-class focus and engagement.

We study disciplinary incidents and students' perceived likelihood of online bullying as indicators of student behavior and peer interactions, and student-reported subjective well-being as a measure of students' emotional experience. Together, these outcomes allow us to assess whether restricting smartphone access impacts academic performance and engagement, alters behavioral dynamics, or affects students' psychological well-being.

The expected effects of restricting smartphone access on academic performance and engagement are ambiguous ex ante. On the one hand, reducing access to smartphones during the school day may decrease digital distraction and increase time on task. If students reallocate attention toward instructional activities, this could improve learning, increase engagement, and raise test scores, attendance, and reported classroom focus. On the other hand, the academic consequences depend critically on what students substitute toward when phone access is removed. For example, if attention shifts toward distracting in-class interactions with peers, improvements in academic performance and attention may be limited. Attention could also shift to digital distractions that are not blocked, such as accessing video or social media sites on laptops. In addition, implementing and enforcing a new restriction may require monitoring and compliance efforts that divert teacher time and fragment classroom instruction. Stricter enforcement could also affect student–teacher relationships, potentially diminishing trust or perceived autonomy, with uncertain implications for engagement and learning. In settings where smartphones had been used for instructional or administrative tasks (such as polling students or accessing learning platforms), removing access may require alternative workflows that consume instructional time.

The expected effects on disciplinary incidents and school climate are also ambiguous. On the one hand, limiting smartphone access may reduce certain types of off-task behavior related to devices and decrease disruptions directly tied to phone use. On the other hand, introducing a new rule creates scope for new disciplinary actions when that rule is violated (Figlio and Ozek, 2025). Increased monitoring may raise the number of recorded

disciplinary incidents even if underlying behavior changes little. Moreover, if smartphones previously absorbed attention that might otherwise have been directed toward peers (Allcott et al., 2020), their removal may increase in-person interaction during the school day. Greater peer engagement could strengthen social relationships, but it may also increase opportunities for conflict or classroom disruption (Jacob and Lefgren, 2003), potentially contributing to higher disciplinary incidents. For students' perceived likelihood of online bullying, if smartphones facilitate online harassment among peers during the school day, restricting access may reduce such opportunities; however, because online bullying can occur outside school hours or through other devices, the overall effect of in-school restrictions is ambiguous.

The expected effects on subjective well-being are similarly ambiguous. On the one hand, reducing access to smartphones may reduce digital distraction and social comparison during the school day, potentially lowering anxiety and improving students' emotional experience (Allcott et al., Forthcoming). On the other hand, restricting smartphone access represents a discrete change in students' daily routines and social environment. Such a change may generate short-run adjustment costs—through frustration, reduced perceived autonomy, altered peer dynamics, or increased monitoring—that could temporarily reduce well-being. Over time, increased in-person interaction and reduced dependence on constant connectivity could strengthen peer relationships and lessen pressures associated with online engagement, potentially improving students' emotional well-being. At the same time, if students value digital communication or experience the restriction as persistently limiting autonomy, well-being could remain lower even in the longer run (Gajdics and Jagodics, 2022). For these reasons, the direction and magnitude of the effects are ultimately empirical questions.

3 Data

A key challenge for national research on school smartphone policies is the absence of standardized records documenting policy adoption and implementation. Although most school districts have some form of smartphone policy, official records are not systematically maintained, and stated policies may differ substantially from their effective implementation.

A related challenge is that many of the outcomes most central to the public debate—such as students' emotional well-being, attention, and in-class experiences—are not typically captured in administrative data. Studying these outcomes at scale requires nationwide student surveys. Even for more traditional outcomes such as standardized test scores, attendance, and discipline, data are reported through distinct state-level systems that differ in coverage, definitions, and availability across states and years, complicating efforts to construct consistent national measures.

To address these challenges, our analysis draws on multiple data sources that together allow us to identify the timing of school-level adoption of lockable phone pouches, characterize changes in school smartphone policies and student phone use, and measure an array of student outcomes across states and years.

We combine administrative adoption data from Yondr with smartphone location data, a national teacher survey, nationwide student survey data, and harmonized school-level administrative datasets. These sources allow us to (i) precisely date Yondr adoption at the school level, (ii) measure changes in student phone use following adoption, (iii) characterize the broader policy environment in both Yondr and non-Yondr schools, and (iv) construct consistent measures of student outcomes—including both administrative outcomes and student-reported measures of well-being and classroom experience. All datasets are merged using standardized National Center for Education Statistics (NCES) school identifiers.

This section describes the data sources used in the analysis and outlines the construction of the main analytic samples. Appendix C provides detailed documentation of sample construction, including the sequence of sample restrictions applied for each outcome and the resulting sample sizes.

For each outcome domain—test scores, attendance, discipline, and student surveys—we pre-registered the primary outcome definitions, including how each primary outcome would be measured and aggregated, as well as planned heterogeneity analyses, prior to estimating treatment effects. The pre-registration document, which details these design choices, is publicly available here.

3.1 Data Sources

Yondr Adoption Data. Information on school adoption of lockable phone pouches comes from administrative records provided by Yondr. We established a data use agreement with Yondr, which supplied school- and district-level records covering 4607 U.S. public schools that purchased pouches between 2014 and the end of our study period. These records include school identifiers and implementation timing. We match Yondr accounts to NCES school identifiers using provided IDs and manual matching based on school names, addresses, and district information, allowing us to construct a unified school-level adoption panel.

Our primary treatment variable is the first academic year in which a school is recorded as implementing Yondr pouches. Because accurate treatment timing is central to our difference-in-differences design, we exclude a small number of schools for which implementation dates cannot be reliably established based on the administrative records. In a limited number of cases, shipments were recorded at the district level and could not be reliably attributed to

individual schools; we exclude these districts from the analysis (including from the control group) because school-level treatment status cannot be determined.

Appendix C.1 provides additional details on the construction of the adoption panel and timing classifications, including summary counts of schools in each category and the number excluded from the main analysis.

Smartphone Location Data. To measure the "first stage" effect of school-based smartphone restrictions on phone use, we use smartphone GPS data from two different commercial providers: Advan Research and Quadrant. These data are based on time-stamped location "pings" generated when mobile devices interact with partner apps or websites.

Our primary results use data from Advan Research, which provides pre-processed school-level aggregates at an hourly resolution. Advan tracks around 35 million cell phones in the United States. Our dataset consists of hourly data for a sample of 40,542 schools, from January 2019 to January 2026. Device counts are provided as scaled "visits," which are estimates of the actual number of devices present at a given location and time, extrapolated by Advan based on their panel size and characteristics, and demographics of the population at that location. As discussed in detail below, we use this measure as a proxy for phone usage because it is based on pings recorded with actual phone activity. We construct year-level outcomes in several steps. First, we aggregate the hourly data to monthly averages of "net visits," defined as average weekday school-hours visits (using alternative hour windows in robustness checks) minus the corresponding weekend activity during the same hours within the same month. To further reduce background noise, we subtract the weekday–weekend difference in nighttime activity (00:00–06:00). We then form an academic-year panel by averaging these monthly measures over September, October, November, February, March, April, and May.

While we use Advan as our primary data source because it provides greater time coverage, we also use ping-level Quadrant data as a robustness exercise to verify that our findings are not driven by Advan-specific aggregation and scaling. We map about 2 trillion raw phone coordinates—observed in February and October of 2022-2024, as well as February 2025—to school polygons (i.e., building geographical perimeters) and construct outcomes directly from the underlying ping data, in which each observation represents an "event" when a smartphone is observed. We build a school-year panel by computing the average of these two months for each academic year.

Because pings are generated mostly when phones are actively in use, a phone that is powered off or locked away—for example, inside a lockable pouch—should generally generate

fewer location signals.[3] We discuss this as well as other data processing steps in more detail in Appendix C.2. Still, the GPS data from both sources are inherently noisy proxies for phone use: location events depend on app adoption, operating-system settings, permission choices, and the set of observable devices changes over time as users install/uninstall apps or opt in/out. Accordingly, we interpret these measures as proxies for relative smartphone use across schools and periods, not as direct measures of overall usage levels.

Despite this limitation, in Appendix C.2 we present several validation exercises that suggest the data contain meaningful signal. Phone activity during school hours is higher in middle and high schools than in elementary schools, scales positively with school enrollment, and is substantially lower at night and on weekends. In addition, school-level phone activity is correlated with teacher-reported phone use and policy strictness from the NTS, described below, and we find detectable changes in phone activity following Yondr adoption.

Nationwide Teacher Survey. Administrative data from Yondr provide reliable information on the timing and location of lockable pouch adoption, but they do not capture schools' prior phone policies, policies in non-Yondr schools, or how phone policies are implemented in practice. To fill this gap, we use the NTS, a national survey developed by members of the research team to systematically document school cell phone policies and student phone use.

We collaborated with educators to design a brief (approximately five-minute) survey that efficiently collects information on school phone policies, enforcement, and student behavior. The survey was fielded nationally through partnerships with state education leaders and direct outreach to educators, including email recruitment. Since its national release on April 8, 2025, the survey has received nearly 68,000 responses from educators in all 50 states. Throughout the paper, we restrict attention to responses from middle and high schools, where smartphone use is most prevalent and where Yondr adoption is concentrated.

The NTS collects information on both the content and implementation of school cell phone policies. Teachers report when personal phone use is restricted during the school day and where students are permitted to keep their phones (e.g., on their person or in lockers). To assess enforcement and student behavior, teachers report how consistently the policy is enforced, the share of students using phones between classes, and the share using phones for personal reasons during class. Teachers also evaluate their school's policy, including their level of satisfaction and whether they believe the policy should be more or less restrictive. Appendix C.3 reproduces the exact wording of the survey questions used in the analysis.

[3]Quadrant's data is provided as raw ping counts; Advan's data is provided as scaled estimates based on the number of pings. These pings are recorded only when devices in the provider panel request location information through partner apps or websites, usually when the user is actively using the app (e.g., a weather or entertainment app that requests the user's location).

The survey further documents policy changes over time. Teachers report whether their school has experienced a major change in its smartphone policy in the past five years and, if so, when the change occurred, what the prior policy was, and how that policy was enforced. To assess potential confounding factors, teachers are also asked whether other major school-level changes—such as leadership turnover or the adoption of new programs—occurred around the same time as the phone policy change.

Table A1 compares baseline characteristics of schools in NTS to the full NCES population of middle and high schools. Relative to the NCES population, schools represented in the NTS are larger on average and more likely to serve high school grades. NTS schools are also somewhat more suburban, less likely to be charter schools, and serve student populations that are from higher-income neighborhoods.

Common Core of Data. We also use the Common Core of Data (CCD), maintained by the NCES, which provides annual, school-level administrative data for all public schools in the United States. The CCD serves two primary purposes in our analysis.

First, we use CCD data to construct baseline school-level covariates for our difference-in-differences analyses. These include total school enrollment, enrollment by race/ethnicity (used to construct racial composition measures), state identifiers, urbanicity, and charter school status. We also use grade-level enrollment counts to construct indicators for whether a school serves grades below 6 or above 9. These indicators allow us to distinguish, for example, K–8 schools from schools that serve only middle school grades, as well as middle schools from combined middle–high schools, which is important given our focus on test scores in grades 6 through 8.

Urbanicity is defined using NCES locale codes, which we aggregate into four categories: urban, suburban, town, and rural. Charter status is defined using the CCD charter school indicator. For the small share of observations with missing covariate values, we impute covariates using mean values within cells defined by state, urbanicity, and charter status, so that the imputation relies on schools with similar observable characteristics.

To measure school-level poverty, we rely on the EDGE School Neighborhood Poverty Estimates, which combine American Community Survey (ACS) data with NCES school locations to estimate neighborhood income-to-poverty ratios around school buildings. We use this measure in place of free or reduced-price lunch (FRPL) eligibility, which may reflect supply-side policy choices in addition to underlying economic conditions. Using baseline values, we assign schools to state-specific poverty quartiles, which enter the analysis as indicator variables.

Second, we use CCD data to conduct compositional checks that assess whether Yondr

adoption is associated with changes in school enrollment or student demographic composition. Specifically, we examine changes in total enrollment, poverty rate, and racial composition following adoption.

Standardized Test Score Data. A central challenge for national research on student achievement is the lack of a single assessment administered uniformly across states. States administer distinct exams with different scales, and publicly available data vary in whether they report average scale scores or only proficiency rates. To construct comparable school-level measures of achievement across the United States, we combine state assessment data compiled by the Education Data Center (EDC) with harmonized test score estimates from the Stanford Education Data Archive (SEDA).

Grades 3 through 8 are the most consistently tested across states under federal and state accountability systems. All U.S. states administer standardized assessments in these grades, allowing us to obtain comparable achievement data across states and over time. In contrast, high school testing practices vary widely across states. Some states administer end-of-course exams in selected subjects, others rely on a single exit exam, and many states use college entrance exams such as the ACT or SAT to satisfy accountability requirements. These differences make it substantially more difficult to construct consistent, school-level measures of high school achievement that are comparable across states and years. For this reason, and because student smartphone use is much more widespread in middle school than elementary school grades, we focus on middle school grades (6 through 8).

We begin with data from EDC's State Assessment Data Repository, or "Zelma," which aggregates publicly available state assessment data from all 50 states and the District of Columbia and integrates standardized NCES school identifiers. For the 22 states that report school-level average scale scores, we obtain English Language Arts and Mathematics results for grades 6 through 8.[4] These data are available for selected years between the 2017–18 and 2024–25 school years, excluding years in which statewide testing was disrupted by the COVID-19 pandemic (2019-20 and 2020-21). Hereafter, we refer to academic years by the spring year (e.g., the 2019–20 school year is denoted as 2020).

For each state-by-year-by-grade-by-subject cell, we standardize school-level average scale scores to have mean zero and unit variance, and then construct school-level achievement measures by taking enrollment-weighted averages across grades. That is, all test score outcomes are measured in *school-level standard deviations*, reflecting variation in school mean achievement rather than individual student outcomes.

Not all states report average scale scores. For states that report only proficiency rates,

[4]See Table A2 for details on which source was used for each state.

we rely on achievement estimates from SEDA, which use state proficiency data linked to the National Assessment of Educational Progress (NAEP) to place test scores on a common scale that is comparable across states, grades, and years. SEDA reports school-by-year-by-grade achievement estimates in student standard deviations, which we aggregate to the school-year level using the same procedures applied to the EDC data.[5]

Our main analysis combines these two sources to maximize national coverage. As a robustness check, we also present results using only the subset of states for which average scale scores are available from EDC, excluding SEDA-based measures. For each school and year, we construct an average achievement measure by combining standardized Math and ELA scores. When one subject is missing (in approximately 3% of school–year observations), we construct the average using an imputation procedure described in Appendix D.

School-Level Attendance Data. To examine the effects of Yondr adoption on student attendance, we assemble school-level attendance data from state education agencies. Unlike test scores, there is no comprehensive national database that reports comparable school-level attendance measures across states and years. We construct this dataset by collecting and harmonizing attendance data on a state-by-state basis, focusing on states with substantial Yondr adoption and with reporting practices that allow for meaningful comparison over time.

We focus on two related attendance measures: attendance rates and chronic absenteeism. Attendance rates are typically defined as the percentage of enrolled student-days on which students are recorded as present, including both excused and unexcused absences. Chronic absenteeism is defined as the share of students who miss at least a state-defined threshold of instructional days during the academic year, most commonly 10 percent or more. Because attendance rates are continuous and have been used in prior research on school phone bans (e.g., in Figlio and Ozek (2025)), we committed ex-ante to treat attendance rates as our primary outcome and present chronic absenteeism as a secondary measure.

In all states included in our analysis, out-of-school suspensions are coded as absences, implying that attendance rates may reflect both instructional attendance and disciplinary exclusions. This is relevant for interpreting estimated effects, as policies that affect student behavior could plausibly influence both engagement and discipline. California is an exception in that it reports absences disaggregated by reason, including excused absences, unexcused absences, and absences related to disciplinary actions. Therefore, for California, we addi-

[5]We prioritize EDC data where available and supplement with SEDA estimates when necessary for two reasons. First, SEDA does not include achievement estimates for several large states during our sample period, including New York, which accounts for a substantial share of Yondr adoption. Second, the most recent publicly available SEDA data extend only through the 2024 school year, whereas EDC assessment data are available through 2025; given that Yondr adoption accelerates over time, incorporating the most recent achievement data is important for capturing schools' outcomes during periods of rapid policy uptake.

tionally construct an alternative attendance measure that excludes disciplinary absences and use this measure in robustness checks.

Data availability varies across states and years. For attendance rates, we assemble data for 27 states between the 2018 and 2025 academic years; for chronic absenteeism, data are available for 36 states. We construct the attendance dataset by beginning with states that have the highest levels of Yondr adoption and sequentially expanding coverage, retaining only states that report a comparable attendance rate measure consistently across years. Appendix C.5 and Table A2 provide additional details on the construction of the dataset, including the states included for each outcome. As with our other outcomes, we restrict the sample to schools serving middle and high school grades.

School-Level Disciplinary Outcomes. To examine the effects of Yondr adoption on student discipline, we assemble a school-level disciplinary dataset from state education agencies. As with attendance, there is no comprehensive national database that reports comparable school-level disciplinary measures across states and years. We therefore construct this dataset by collecting, harmonizing, and standardizing discipline data on a state-by-state basis, focusing on states with substantial Yondr adoption and with reporting practices that permit consistent measurement over time.

The underlying state data report a variety of discipline-related outcomes, including in-school suspensions, out-of-school suspensions, and broader measures of disciplinary actions that may also include expulsions or any other removal from regular classroom activities as a result of negative behavior. Accordingly, we construct a unified measure of disciplinary intensity that uses the most comprehensive measure available for each state–year observation. Specifically, we prioritize (i) the share of students disciplined, when available; (ii) the share of students receiving any suspension; (iii) the combined share receiving in-school or out-of-school suspension; and, if necessary, (iv) suspension rates based on in-school or out-of-school suspensions alone.[6]

We focus on disciplinary outcomes expressed as rates, using state-reported enrollment denominators. We then standardize the resulting disciplinary measure within state–year cells. Our primary outcome is this standardized discipline index, which captures within-state–year variation in disciplinary intensity and ensures that differences in discipline definitions or reporting practices across states do not mechanically drive the estimates. Appendix C.5 and Table A2 detail the composition of the discipline index, including the share of observations derived from each underlying disciplinary measure. Data availability varies across states and

[6]Measure (iii) will "double-count" students who received both in-school and out-of-school suspensions. However, we standardize all of our disciplinary measures within state-year cells, which allows us to make valid comparisons between treatment and control schools.

years. The final analytic sample includes 22 states with discipline data between the 2018 and 2025 academic years. As with our other outcomes, we restrict the sample to schools serving middle and high school grades.

Panorama Student Survey Data. Standard administrative data do not directly capture many of the mechanisms through which smartphone restrictions are hypothesized to affect students, such as changes in attention, engagement, or emotional well-being during the school day. To examine these dimensions, we use student survey data provided by Panorama Education, an education technology company that administers surveys to K–12 schools measuring students' academic experiences, classroom engagement, and socio-emotional well-being.

Panorama surveys are administered at the discretion of districts and schools rather than through a nationally representative sampling design, so coverage varies substantially across states, schools, and years. To protect client confidentiality, Panorama does not disclose school identities. Instead, Panorama provided school-by-year aggregates through a secure matching process. For each matched school–year, Panorama reports the mean response to each survey item across all responding students. The resulting dataset includes only anonymized school identifiers that are consistent over time.

The Panorama data span academic years 2019 through 2025. In total, the Panorama file includes 61,570 school–year observations with non-missing survey data during this period, representing 17,429 unique schools across 46 states. Across all school–year observations in the NCES panel during this period, 8.7 percent have at least one Panorama survey measure observed in a given year.

Because Panorama's platform includes a large number of distinct survey items and because survey content varies across schools and years, we adopted a careful, pre-specified process to identify a small number of primary outcomes prior to estimating treatment effects. This process, which we describe in more detail in Appendix C.6 was designed to reduce concerns about specification searching and selective reporting.

We focus on three pre-specified primary outcomes derived from Panorama surveys. The first captures students' subjective well-being (SWB), combining items that measure positive feelings (e.g., happiness, excitement, feeling safe) and challenging feelings (e.g., sadness, worry, frustration). The second captures classroom attention, combining items related to students' ability to focus, pay attention, and exert effort during class. The third captures students' perceived likelihood of online bullying, based on a survey item asking how likely it is that someone from their school would bully them online. These outcomes reflect central channels through which smartphone restrictions are hypothesized to affect students: emotional experience, in-class focus, and peer behavior in online settings.

For each outcome, we first code individual survey items so that higher values correspond to more desirable outcomes and then standardize them within survey year to have mean zero and unit variance. For SWB and classroom attention, we compute indices as the unweighted average of all available standardized items within the relevant construct and re-standardize the resulting index by year. The online bullying outcome is constructed from a single standardized survey item. As a result, all Panorama outcomes are expressed in school-level standard deviation units, facilitating comparison across outcomes.

Table A1 compares baseline characteristics of schools with non-missing Panorama data for at least one of the three primary indices to the full NCES population, showing that, relative to NCES schools overall, Panorama participation is more common among larger schools, schools located in urban areas, schools serving more racially diverse student populations, and schools in somewhat higher-income neighborhoods.

Appendix C.6 provides more specific details about the Panorama data, including additional documentation of the outcome selection process, index construction, and analysis sample construction.

3.2 Outcome-Specific Analysis Samples

All analyses are conducted at the school–year level using administrative and survey data drawn from multiple state and federal sources. Because outcomes are obtained from distinct data systems with different coverage across states, years, and schools, we analyze each outcome using its own outcome-specific sample. For each outcome, we use the largest possible set of schools and years for which that outcome is consistently observed, subject to a common set of sample restrictions.

Across all outcomes, we restrict the sample to schools serving at least one middle or high school grade (6–8 and 9–12, respectively), where smartphone use is more prevalent and Yondr adoption is concentrated. We focus on academic years 2018 through 2025 and exclude the pandemic-affected years 2020 and 2021. We exclude schools that adopted Yondr prior to the 2023 academic year, schools in districts identified as potentially contaminated control districts—specifically, districts that may have adopted Yondr but for which school-level adoption timing cannot be reliably determined—and adopting schools for which implementation timing is classified as lowest confidence in the Yondr administrative data. We further restrict the sample to states with at least one Yondr adoption between the 2023 and 2025 academic years, ensuring that identification relies on within-state variation in adoption timing.

We then apply outcome-specific restrictions. For test score analyses, we require non-missing standardized Math or English Language Arts scores. For attendance analyses, we

require a non-missing and comparable attendance rate measure. For discipline analyses, we require at least one disciplinary measure available from which a standardized discipline index can be constructed. Finally, for the Panorama analyses, we require sufficient survey data to construct the SWB, classroom attention, or online bullying index.

The resulting analytic samples differ across outcomes due to differences in data availability and reporting regimes. Appendix C provides detailed documentation of the construction of each outcome-specific analysis sample.

4 Descriptive Facts on School Smartphone Restrictions

This section presents descriptive facts that characterize the landscape of school smartphone restrictions in the U.S. and motivate our causal analysis in later sections. We draw on data from the NTS, launched in April 2025, as well as nationally representative Gallup surveys of 1,801 dyads of parents and their children, fielded in February 2026. The NTS sample includes 13,684 unique middle and high schools and 55,848 educators across all 50 states, reflecting responses received between April 2025 and February 2026 (the time of this writing).

4.1 Schools Have a Wide Variety of Specific Smartphone Policies

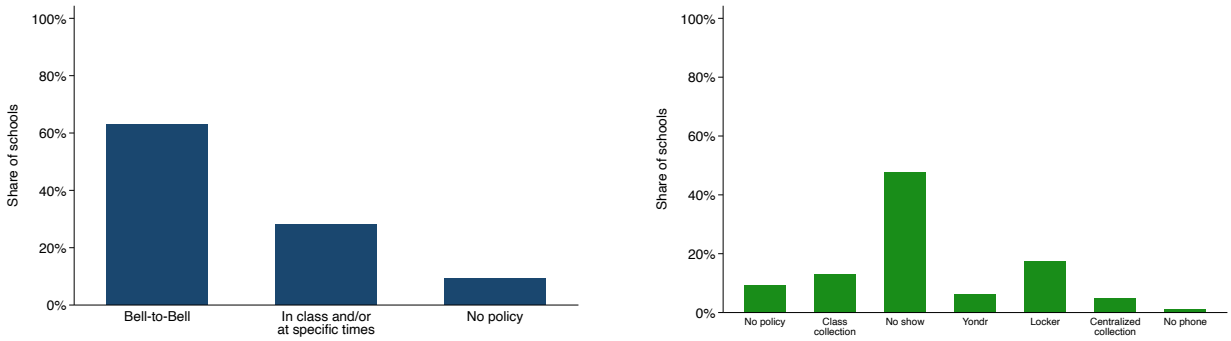
Schools currently regulate student smartphone use in a wide variety of ways, both in terms of when phone use is restricted and where students are permitted to keep their devices during the school day. Figure 1 illustrates the resulting heterogeneity in school smartphone policies. The figure is constructed at the school level: when multiple teachers respond from the same school, policies are defined using the modal teacher report.

Panel A shows the distribution of policies governing *when* personal phone use is restricted. 63 percent of schools report bell-to-bell restrictions, while 28 percent allow phone use during non-instructional periods such as passing time or lunch. A nontrivial share of schools—9 percent—report having no formal restrictions on personal phone use.

Panel B shows comparable heterogeneity in *where* students are permitted to keep their phones during the school day. The most common arrangement is a "no show" policy, reported by 48 percent of schools, under which students may keep phones in backpacks or pockets but are required to keep them out of sight. 17 percent of schools require students to store phones in lockers and 6 percent of schools report using lockable pouches such as Yondr.

Figure 1: Types of Smartphone Policies

(a) When is personal phone use restricted? **(b)** Where can students keep their phones?



Notes: This figure summarizes school smartphone policies using data from the NTS. Panel A reports when phones are restricted during the school day (Q2), with categories corresponding to away-for-the-day (bell-to-bell) restrictions, schedule-based restrictions (e.g., use permitted during lunch or between classes), and no school-wide restriction. Panel B reports where students are allowed to keep their phones (Q3), with categories indicating whether phones are prohibited on campus, centrally collected, secured in lockable pouches such as Yondr, collected in classrooms, stored in lockers, kept out of sight ("no show"), or governed by no school-wide policy. See Appendix C.3 for the exact wording of each question. When multiple teachers respond from the same school, policies are defined at the school level using the modal teacher report. The sample is restricted to middle and high schools.

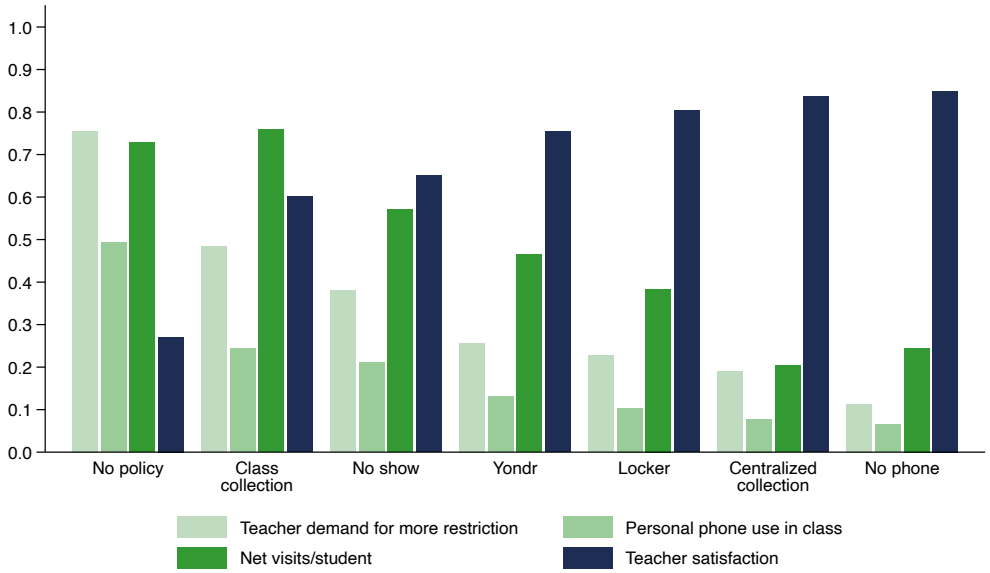
4.2 Teachers Prefer More Restrictive Policies

Figure 2 presents descriptive evidence on teacher-reported outcomes and smartphone use across schools with different types of smartphone policies. The figure combines data from the NTS with smartphone location data and reports outcomes at the school level, grouped by the modal policy governing where students are permitted to keep their phones during the school day.

Schools that limit students' physical access to phones—such as requiring phones to be stored in hallway lockers, left at home, centrally collected, or secured in lockable pouches like Yondr—exhibit lower reported levels of personal phone use during class and higher teacher satisfaction with their school's phone policy. These patterns are mirrored in the GPS-based measures: schools with more restrictive policies show fewer net smartphone pings per student during the school day, suggesting lower overall phone activity on campus.

By contrast, less restrictive approaches—such as "no show" policies or classroom-based collection—are associated with higher reported phone use and lower satisfaction. The figure points to a consistent association between the restrictiveness of smartphone policies, reduced student phone use, and higher teacher satisfaction.

Figure 2: Outcomes Associated with Alternative Smartphone Policies

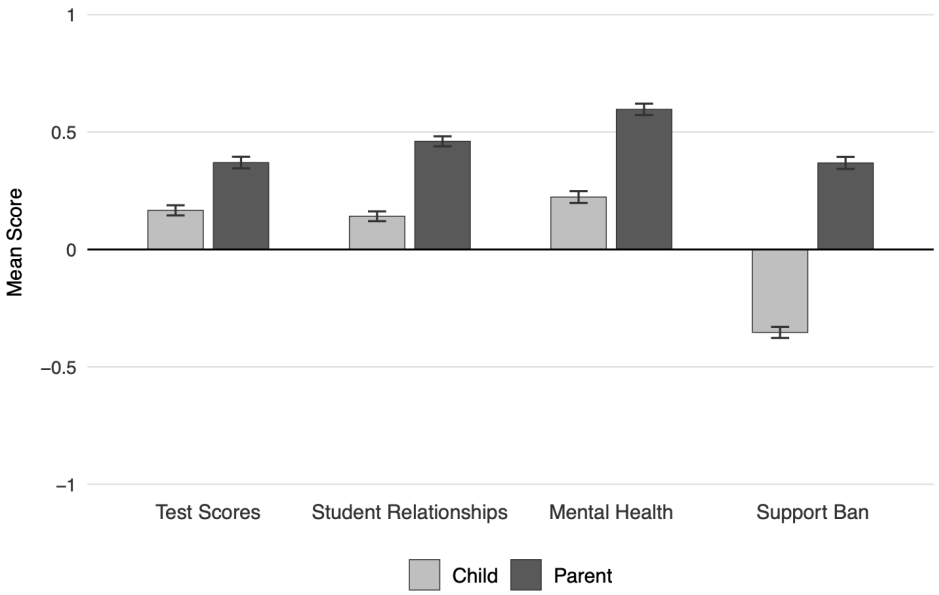


Notes: The figure reports school-level averages constructed from the NTS and smartphone location data. Teacher satisfaction with the current policy is based on responses to Q8, measured on a 0–100 scale. Desire for more restriction is based on responses to question Q9, coded on a five-point scale ranging from "much less restrictive" to "much more restrictive." Personal phone use in class reflects teachers' estimates of the share of students using phones for personal reasons during class (Q6), reported in 10 percentage point increments from 0 to 100 percent. Survey outcomes are averaged across teachers within a school, and schools are classified by their modal reported smartphone policy. See Appendix C.3 for the exact wording of each question. Net visits are derived from Advan data measured in November 2025, and are defined as average weekday school-hours (10h-11h) minus the corresponding weekend activity in the same month, further subtracting the weekday–weekend difference in nighttime activity (00:00–06:00). The sample is restricted to middle and high schools.

4.3 Parents Support Phone Bans While Students Oppose Them

Parents were asked to imagine a policy in which schools require students to keep phones in lockable pouches throughout the school day. On the recoded scale ranging from −1 to 1 (-1 = get worse / oppose; 0 = won't change / neutral; 1 = improve / support), parents on average report that such a policy would increase standardized test scores (0.370), improve student relationships (0.461), and improve mental health (0.597), and they also express net support for such bans (0.369). Students are notably less favorable: they expect only modest improvements in test scores (0.167), student relationships (0.141), and mental health (0.223), and on average tend to oppose lockable pouch bans (−0.353). Thus, while both groups are most optimistic about mental health effects, parents are substantially more positive than students across all outcomes and differ especially sharply on support for the ban itself.

Figure 3: Parent and Child Attitudes Towards Smartphone Policies



Notes: The figure reports mean scores from the Gallup survey based on the following recoding: -1 = get worse / oppose; 0 = won't change / neutral; 1 = improve / support. Error bars show ± 1 standard error. Means are weighted separately for parent responses and child responses adjusting to make the sample representative of U.S. youth and young adults aged 12-29 and the parents of youth aged 12-18. N = 1,801 parent-child pairs. See Appendix C.4 for the exact wording of each question and details on the data. Table A4 shows unweighted sample demographics.

5 Difference-in-Differences Empirical Strategy

Estimating the causal impact of restricting student smartphone use in our context presents several empirical challenges. Most importantly, as we show below, the adoption of Yondr

lockable pouches is not randomized: schools that adopt Yondr differ systematically from non-adopting schools along observable dimensions such as size, urbanicity, and student demographics.

Our research design compares changes in outcomes over time in schools that adopt Yondr to changes in outcomes in observationally similar schools that do not adopt. A key strength of our setting is the availability of a large pool of never-treated schools, which allows us to construct a well-matched comparison group using baseline covariates. We assess the validity of this design by examining pre-adoption trends and show that treated and comparison schools exhibit similar outcome trajectories prior to adoption.

We begin by describing the standard two-way fixed effects (TWFE) event-study specification that motivates our analysis. Formally, the TWFE analogue of our event-study specification can be written as:

$$Y_{it} = \sum_{k \neq -1} \beta_k \cdot \mathbb{1}[\text{event_time}_{it} = k] + X_i \times \delta_t + \alpha_i + \delta_t + \varepsilon_{it}, \tag{1}$$

where Y_{it} is an outcome for school i in year t, event_time_{it} indexes years relative to the first year of Yondr adoption ($k = -1$ is the omitted reference period), α_i are school fixed effects, and δ_t are year fixed effects. The coefficients β_k trace the dynamic effects of Yondr adoption before and after implementation. The coefficients in the pre-adoption period provide a direct test of the identifying assumption that, absent Yondr adoption, outcomes in treated and comparison schools would have followed parallel trends.

Here, X_i denotes a vector of baseline school characteristics, including total enrollment, racial composition (shares White, Black, and Hispanic), indicators for serving high school or elementary grades, charter status, poverty quartiles, and group indicators defined by state $\times$ urbanicity. These covariates enter the model interacted with year fixed effects, allowing their associations with outcomes to vary flexibly over time while remaining anchored to pre-treatment values.

While the TWFE event-study specification is intuitive and widely used, recent work shows that it can yield biased estimates in settings with staggered treatment adoption and heterogeneous treatment effects. For this reason, our primary empirical approach uses the difference-in-differences framework of Callaway and Sant'Anna (2021), which is explicitly designed for such settings. This framework estimates group–time average treatment effects by comparing outcomes for treated schools to outcomes for an appropriate comparison group. Importantly, the estimator allows for the inclusion of covariates, so identification relies on parallel trends conditional on observables, which is important in our setting where adopters differ substantially from non-adopters.

We use the improved doubly robust implementation of this estimator, which combines inverse-probability tilting to estimate the propensity score and weighted least squares for the outcome regression, and remains consistent if either the treatment model or the outcome model is correctly specified (Sant'Anna and Zhao, 2020). Treatment is defined as the first year of Yondr adoption at the school level, and never-treated schools serve as the comparison group.

More specifically, the estimands of interest are group-time average treatment effects, denoted $\text{ATT}(g, t)$, which capture the average effect of Yondr adoption in period t for the group of schools first treated in period g. In our setting, g corresponds to the first year of Yondr adoption (e.g., 2023, 2024, or 2025). Once the $\text{ATT}(g, t)$ are estimated for different values of g and t, we aggregate them to summarize treatment effects in three main ways.

First, we construct event-study estimates that trace the evolution of outcomes relative to the year immediately preceding adoption, allowing treatment effects to vary flexibly over time. Second, we estimate cohort-specific average treatment effects for schools adopting Yondr in different years, allowing us to explore heterogeneity by adoption timing. Third, we report an overall ATT, defined as $\tau = \mathbb{E}[Y_{i1} - Y_{i0} \mid T_i = 1]$, which averages treatment effects across all post-adoption years and cohorts.

As a complement to the Callaway–Sant'Anna estimates, we also report results from TWFE event-study specifications. In these specifications, we mirror the primary analysis by applying inverse-probability weights based on baseline covariates and including covariates interacted with year fixed effects. These TWFE estimates serve as a useful benchmark but are secondary, as conventional TWFE estimators may be biased in the presence of staggered adoption and heterogeneous treatment effects. In practice, however, the two approaches yield very similar point estimates, consistent with the large share of never-treated schools and the limited number of adoption cohorts in our setting.

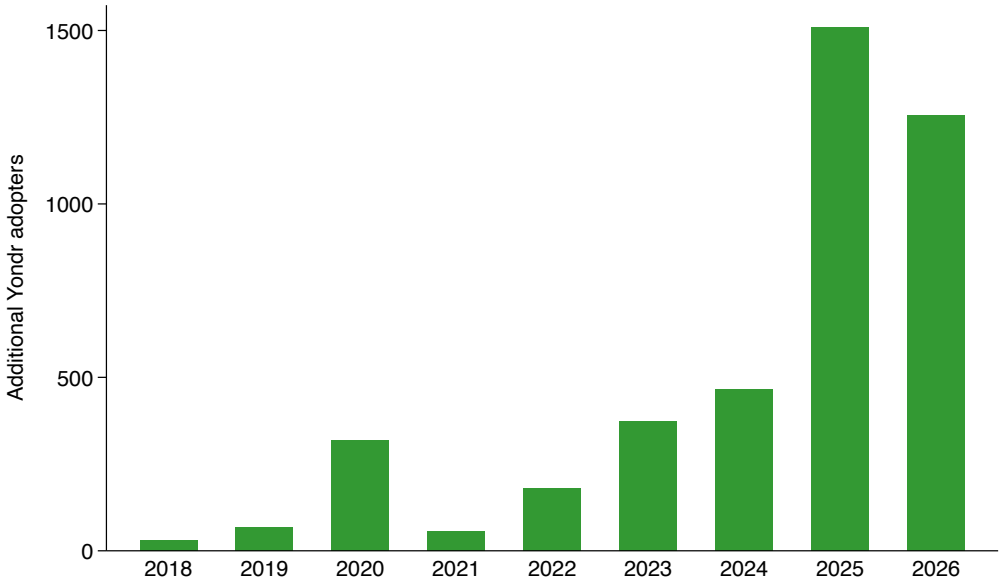
5.1 Characterizing Yondr Treatment and Control

Treatment. Figure 4 illustrates the timing of Yondr adoption across schools. While a small number of schools adopted Yondr prior to the COVID-19 pandemic, adoption increased sharply beginning in the 2023 school year, with the largest number of schools adopting in 2025 and 2026. To focus on the period in which adoption accelerates and to allow for longer pre-treatment trend analysis, our main analyses restrict attention to schools that adopt Yondr in 2023, 2024, or 2025. Schools that adopted Yondr between 2018 and 2022 are excluded from the analysis.

Schools scheduled to adopt Yondr in 2026 are grouped with never-treated schools and

serve as part of the control group throughout the analysis. In robustness checks, we show that our results are not sensitive to how these schools are handled, including specifications that exclude 2026 adopters entirely or use them as the sole comparison group.

Figure 4: Yondr Adoption Over Time



Notes: The figure shows the number of schools adopting Yondr lockable pouches by adoption year, based on administrative data from Yondr. Adoption year is defined as the spring of the academic year in which a school is first observed implementing Yondr (e.g., adoption during the 2024–25 school year is recorded as 2025). Each school appears once, in its initial year of adoption.

Table 1 reports summary statistics for schools in the test score sample, comparing Yondr adopters to schools that never adopt Yondr during the study period. Columns (1) and (2) show means in the raw sample. Several clear differences emerge. Relative to never adopters, Yondr-adopting schools are larger, more likely to be located in urban areas, and serve student populations that are disproportionately Black and Hispanic. In addition, Yondr adopters have substantially lower baseline (i.e., 2018) test scores in both Math and ELA and are located in neighborhoods with lower income-to-poverty ratios. These patterns underscore the non-random selection into Yondr adoption and motivate the need for careful adjustment.

Columns (4) and (5) show the same summary statistics after reweighting the sample using inverse probability weights constructed from baseline school characteristics. We compute these weights using logit, with treatment defined as eventual Yondr adoption (in 2023, 2024, or 2025) and covariates given by the characteristics listed in the table. All covariates are measured in the first year a school appears in the sample and therefore precede treatment for all adopting schools. Income-to-poverty ratios enter the weighting procedure as state-specific quartiles. Importantly, baseline test scores are not included in the treatment model.

The reweighted sample is substantially more balanced along a wide range of observable dimensions. Differences between adopters and non-adopters in enrollment size, racial composition, urbanicity, grade span, charter status, and neighborhood income are sharply attenuated and, in many cases, effectively eliminated. Although baseline test scores do not enter the weighting procedure, they are also much closer across the two groups in the reweighted sample.

Although identification in our difference-in-differences framework rests on parallel trends rather than balance in levels, the degree of balance achieved on baseline characteristics—including outcomes not explicitly included in the treatment model—is reassuring. This feature of the data motivates our use of the Callaway–Sant'Anna estimator, which incorporates these covariates in a doubly robust way, and our complementary weighted TWFE specifications, which apply the same inverse probability weights and allow covariate effects to vary flexibly over time. Together, these approaches allow us to compare outcomes in treated schools before and after Yondr adoption to outcomes in control schools that were observationally similar prior to adoption.

Table 2 summarizes the phone policy environments that preceded Yondr adoption, based on teacher reports from the NTS. The subset of Yondr-adopting schools that appear in NTS is not fully representative of all adopters: relative to Yondr schools overall, NTS adopters are somewhat larger, serve a higher share of White students, are less likely to be urban, and are located in higher-income neighborhoods (see Table A3). Accordingly, we interpret these responses as descriptive evidence on how Yondr policies are implemented and perceived by teachers, rather than as a representative account of all adopting schools.

With this caveat in mind, the table shows that adoption most commonly represents a shift away from weakly enforced or nonexistent phone policies. Roughly 61 percent of adopters previously operated under a "no show" regime, in which students were permitted to keep their phones in backpacks or pockets but were required to keep them out of sight during the school day. An additional 24 percent report having had no formal phone policy at all. Transitions from more restrictive alternatives—such as centralized collection, lockers-only policies, or complete bans on phones on campus—are comparatively rare.

Table 1: Summary Statistics by Yondr Adoption Status

	Raw sample			Reweighted sample		
	(1) Never adopters	(2) Yondr adopters	(3) Diff. p-value	(4) Never adopters	(5) Yondr adopters	(6) Diff. p-value
Total enrollment	521	622	0.000	751	622	0.355
Share White	0.52	0.30	0.000	0.30	0.30	0.870
Share Black	0.15	0.30	0.000	0.31	0.30	0.176
Share Hispanic	0.25	0.32	0.000	0.31	0.32	0.290
Urban	0.27	0.58	0.000	0.58	0.58	0.957
Suburban	0.30	0.21	0.000	0.21	0.21	0.755
Rural/Town	0.43	0.21	0.000	0.20	0.21	0.657
Charter school	0.14	0.13	0.100	0.16	0.13	0.062
Schools with grades above 9	0.16	0.20	0.000	0.21	0.20	0.508
Schools with grades below 6	0.53	0.41	0.000	0.41	0.41	0.790
Income-to-poverty ratio	302	272	0.000	266	272	0.256
Avg. test score index	0.01	-0.46	0.000	-0.41	-0.46	0.166
Avg. Math score	0.01	-0.44	0.000	-0.40	-0.44	0.183
Avg. English score	0.01	-0.45	0.000	-0.39	-0.45	0.059
Number of schools	36,056	1385		32,585	1385	

Notes: This table reports summary statistics for schools in the test score analysis sample, comparing our treated group (Yondr-adopting schools during the 2023 through 2025 school years) to the control group (non-Yondr adopters). Columns (1) and (2) report means in the raw sample. Column (3) reports p values on the differences between (1) and (2), using robust standard errors. Columns (4) and (5) report means after reweighting the sample using inverse probability weights constructed from baseline school characteristics. Column (6) reports p values on the differences between (4) and (5). Weights are estimated using a logit model in which treatment is defined as eventual adoption of Yondr lockable pouches during the 2023–2025 adoption window. Covariates include total enrollment, racial composition, urbanicity, charter status, grade span indicators, neighborhood income-to-poverty quartiles, and state-by-urbanicity indicators. All covariates are measured in the first year each school appears in the sample and therefore precede treatment for all adopting schools. Baseline test scores are reported in the table but are not included in the weighting model.

These patterns indicate that Yondr adoption typically reflects a substantial tightening of phone restrictions rather than a marginal adjustment to an already strict regime. Moreover, only about 24 percent of NTS Yondr adopters report that the policy change coincided with other major school-level changes, suggesting that adoption is usually implemented as a standalone reform rather than as part of a broader set of contemporaneous interventions. This distinction is relevant for our difference-in-differences design: if adoption systematically coincided with other large reforms, estimated effects could reflect those changes rather than Yondr itself. The teacher reports therefore mitigate—though do not eliminate—concerns about confounding from simultaneous school-level interventions.

Table 2: Prior Phone Policies Among Yondr Adopters

Prior policy	Percent
No show	60.96
No policy	23.97
Class collection	7.19
Locker	5.99
Centralized collection	1.20
No phone	0.68
Total	100.00

Notes: This table summarizes the phone policy environments that preceded adoption of Yondr lockable pouches, based on educator responses to the NTS. The sample is restricted to middle and high schools that (i) appear in Yondr's internal administrative sales records, (ii) report in NTS that their school experienced a major change in its official cell phone policy in the past five years (Q10), and (iii) report Yondr or a similar lockable pouch as the school's current policy (Q3). Prior policy categories are based on educators' retrospective reports of where students were officially allowed to keep their phones before the policy change (Q14). When multiple educators respond from the same school, policies are defined using the modal educator report. See Appendix C.3 for the exact wording of each question.

Control. Figure A1 uses NTS data to characterize phone policy changes among schools that do not adopt Yondr. We focus on non-Yondr schools in the NTS sample and calculate the share that report a major policy change in any given year. We classify a change as *strict* if the school switches to lockers-only policies, centralized collection, or a full phone ban—policies that Figure 2 associates with relatively low reported phone use and high teacher satisfaction.

Policy activity among control schools rises modestly over time but remains small in absolute terms. In 2023, roughly 1 percent of control schools adopt a strict alternative policy; the share remains below 2 percent in 2024 and below 5 percent in 2025. Thus, even as Yondr adoption expands rapidly, only a small fraction of comparison schools implement similarly restrictive alternatives. Our main estimates capture the effect of Yondr adoption relative to the mix of phone policies observed among non-Yondr schools. In robustness checks, we restrict the control group to schools reporting no strict phone policy changes and,

in a separate specification, to schools reporting no phone policy changes of any kind during the study period.

6 Difference-in-Differences Estimates

6.1 "First Stage" Effects on Smartphone Use

We begin with descriptive survey evidence on how school environments change when schools adopt Yondr. The NTS asks teachers whether their school experienced a major change in its official cell phone policy in the past five years. Teachers reporting a policy change are asked parallel questions about current conditions and about conditions prior to the policy change, allowing us to construct retrospective before-and-after comparisons.

We classify schools based on the modal teacher report of where students are allowed to keep their phones, defining Yondr schools as those reporting lockable pouches and all others as non-Yondr. For each school, we compute average pre- and post-policy outcomes by first averaging teacher responses within school and then averaging across schools within Yondr and non-Yondr groups. For schools whose modal response indicates no policy change, the survey does not collect retrospective "before" measures. For these schools, we set pre- and post-policy outcomes equal to their current reported values. The resulting comparison therefore mirrors our research design: Yondr adopters are compared to non-Yondr schools, some of which had policy changes and others of which do not.

Figure 5 presents these patterns. Panel A reports mean levels before and after the reported policy change, and Panel B reports the corresponding changes relative to the pre-policy baseline. Transitions to Yondr are associated with large shifts in teacher-reported outcomes. Teacher satisfaction increases from approximately 27 percent to 76 percent (49 percentage points). Reported in-class personal phone use falls from 60 percent to 12 percent (48 percentage points), and reported phone use between classes declines by 53 percentage points. In contrast, among non-Yondr schools, satisfaction increases by 23 percentage points, while reported in-class phone use declines by 23 percentage points and between-class use declines by 18 percentage points.

We emphasize that these comparisons are purely descriptive and rely on retrospective teacher reports. In addition, the Yondr schools observed in the NTS data represent only the subset of Yondr adopters that overlap with both our survey sample and the administrative adoption records; as discussed above, they may not be fully representative of all Yondr schools nationwide. With these caveats in mind, two patterns stand out. First, teachers report large increases in satisfaction and reductions in student phone use following the adoption of Yondr.

Second, these reported changes are substantially larger than those in non-Yondr schools. We next turn to objective GPS-based measures of smartphone activity to assess whether similar patterns emerge using our primary research design.

Figure 6 presents event-study estimates of the ATT of Yondr adoption on in-school smartphone activity using GPS location data from Advan. Estimates are based on the Callaway and Sant'Anna (2021) doubly robust difference-in-differences framework and are normalized relative to the year immediately preceding adoption. The outcome is the natural log of net smartphone visits per student during instructional hours. We focus on visits occurring within school building polygons between 10:00 and 11:00, a period when students in most schools are in class rather than arriving or at lunch.[7]

The event-study coefficients show no evidence of differential pre-trends: estimates in the years prior to adoption are small and statistically indistinguishable from zero. Following adoption, smartphone activity declines substantially. In the year of implementation, net visits per student fall by approximately 0.17 log points. The decline grows in subsequent years, reaching 0.24 log points in the first post-adoption year and 0.28 log points by the second year after adoption. While magnitudes should be interpreted carefully as we cannot discern staff devices from student devices, these patterns indicate a large and persistent reduction in in-school smartphone activity following Yondr adoption.

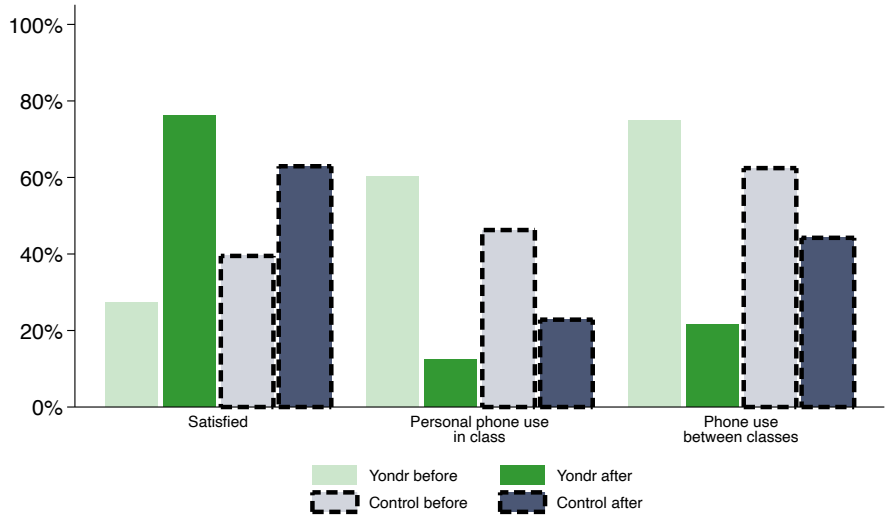
Table A5 reports estimates of the ATT averaged across post-adoption periods and cohorts. Across specifications, Yondr adoption reduces net smartphone visits by roughly 0.15 to 0.20 log points. The estimates are robust to alternative definitions of instructional hours, including broader windows such as 9:00–13:00 and 8:00–15:00. Appendix C.2 presents additional robustness checks using alternative samples and estimates from a different GPS data provider (Quadrant). Across measures and data sources, the results consistently indicate a meaningful and sustained decline in smartphone activity during the school day following Yondr adoption.

6.2 Effects on Test Scores

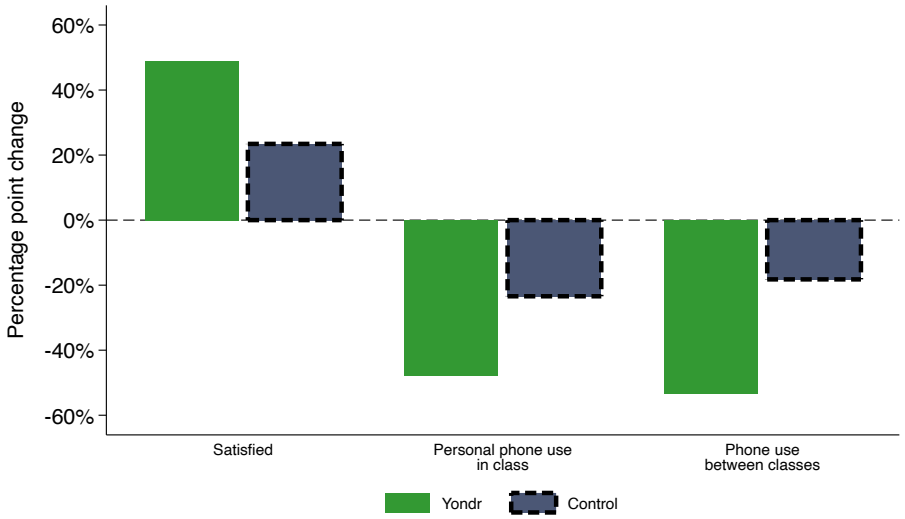
Having established that Yondr adoption leads to large and sustained reductions in student smartphone use, we next examine whether restricting phone access affects student academic achievement. Figure 7 presents event-study estimates of the ATT of Yondr adoption on average standardized test scores in grades 6 through 8, measured in school-level standard

[7]The regression outcome is $\log\Big(\big(P_{\text{weekday,day}} - P_{\text{weekend,day}}\big) - \big(P_{\text{weekday,night}} - P_{\text{weekend,night}}\big)\Big)$, where $P_{\text{weekday,day}}$ denotes the number of pings/visits during school hours on weekdays, $P_{\text{weekend,day}}$ denotes the number of pings/visits during the same hours on weekends, and $P_{\text{weekday,night}}$ and $P_{\text{weekend,night}}$ denote the corresponding nighttime counts.

Figure 5: Teacher-Reported Outcomes Before and After Reported Phone Policy Changes



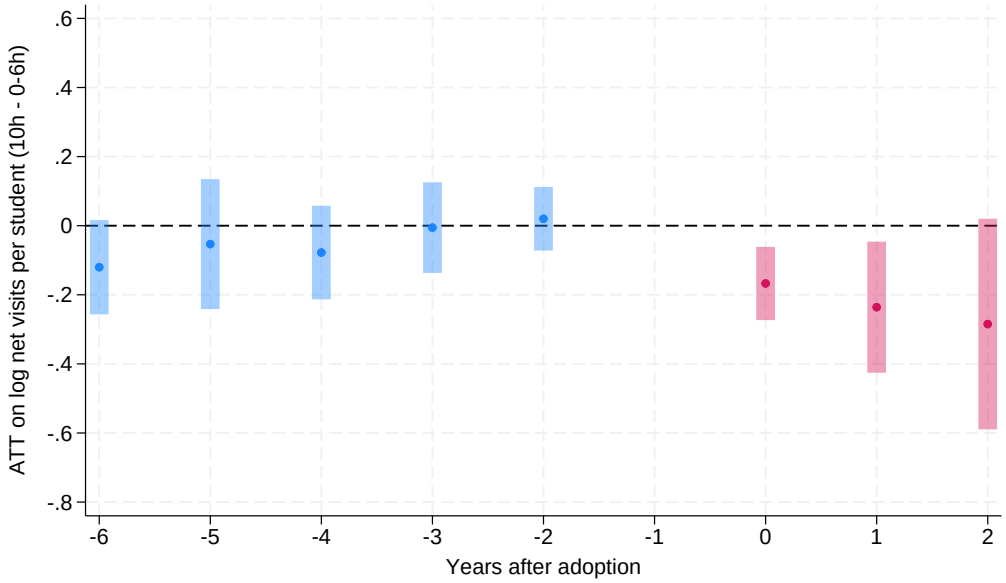
(a) Levels Before and After Policy Change



(b) Change Relative to Pre-Policy Baseline

Notes: This figure uses data from the NTS to compare teacher-reported outcomes before and after reported phone policy changes. Schools are classified using the modal response to Q3 ("Where are students allowed to keep their phones?"), with Yondr schools defined as those reporting lockable pouches and all others classified as control. Policy change status is defined using the modal response to Q10 ("Have there been any major changes in your school's official cell phone policy in the last 5 years?"). For schools reporting a policy change, teachers provide both current outcomes and retrospective outcomes prior to the change. For schools whose modal response indicates no change, retrospective measures are not collected; for these schools, pre- and post-policy outcomes are set equal. Individual teacher responses are first averaged within school and then averaged across schools within Yondr and non-Yondr groups. Outcomes include overall satisfaction with the school's phone policy (Q8) and the percentage of students using phones during class (Q6) and between classes (Q5), measured on 0–100 scales. Panel A reports mean levels before and after the reported change. Panel B reports changes relative to the pre-policy baseline. See Appendix C.3 for the exact wording of each question.

Figure 6: Event Study of the Effects of Phone Pouches on Smartphone GPS Pings



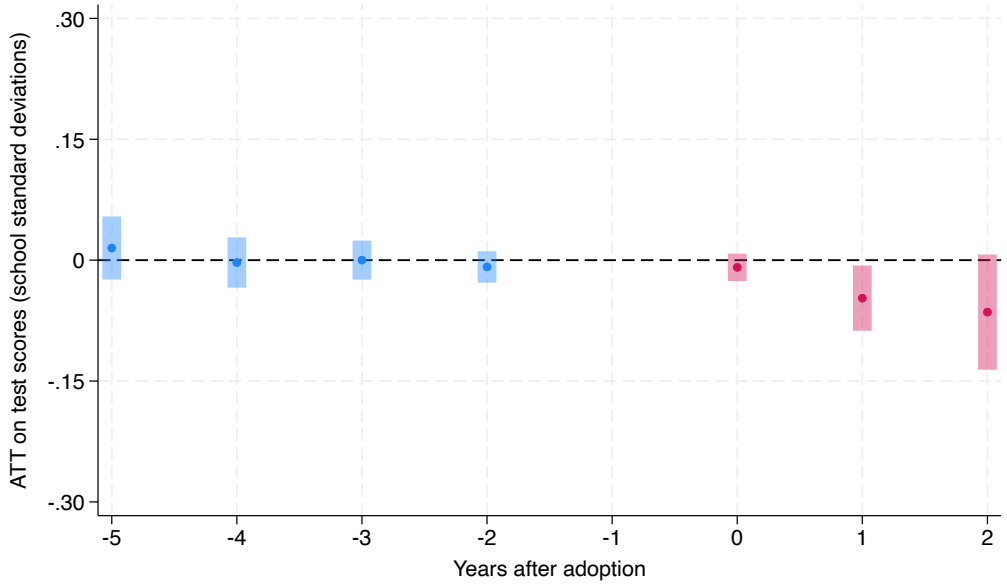
Notes: This figure reports event-study estimates of the effect of Yondr adoption on smartphone use, measured using GPS location data from Advan. Estimates are based on the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator. The plotted coefficients correspond to event-time averages of group-time average treatment effects, $\text{ATT}(g,t)$, aggregated across adoption cohorts and normalized relative to the year immediately preceding adoption ($t = -1$). The outcome is the natural logarithm of net smartphone visits per student during school hours (10:00-11:00) within school buildings. Net visits are defined as average weekday school-hours visits minus the corresponding weekend activity in the same month, further subtracting the weekday–weekend difference in nighttime activity (00:00–06:00). The sample is restricted to middle and high schools, and never-treated schools serve as the comparison group. Covariates and weighting follow the main empirical specification.

deviations. Figures A2 and A3 report analogous event-study estimates for Math and ELA.

Because standardized test scores are unavailable in most states during the 2020 and 2021 academic years due to the COVID-19 pandemic, we exclude these two years from the analysis and relabel the 2018 and 2019 school years as 2020 and 2021, respectively, so that event times $t-5, \ldots, t-1$ correspond to successive observed pre-adoption years. This relabeling is purely expositional and affects only the visual interpretation of pre-treatment periods; it does not affect the estimated post-adoption treatment effects.[8]

The pre-adoption coefficients are small and statistically indistinguishable from zero, providing no evidence of differential pre-trends in achievement between treated and control schools. The estimate for the year of adoption ($t = 0$) is also close to zero and statistically insignificant. In the first post-adoption year ($t+1$), the point estimate is negative and statistically significant, at approximately -0.047 school-level standard deviations. By the second post-adoption year ($t+2$), the estimated decline increases in magnitude to 0.064 school-level standard deviations, though this estimate is not statistically significant at the 5% level.

Figure 7: Event Study Estimates of Effects of Phone Pouches on Middle School Test Scores



Notes: This figure reports event-study estimates of the effect of Yondr adoption on average standardized test scores in grades 6–8, measured in school-level standard deviations. Estimates are based on the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator. The plotted coefficients correspond to event-time averages of group-time average treatment effects, $\text{ATT}(g,t)$, aggregated across adoption cohorts and normalized relative to the year immediately preceding adoption ($t = -1$). Never-treated schools serve as the comparison group. Covariates and weighting follow the main empirical specification.

[8]Figure A4 presents the same event-study estimates without this relabeling, showing additional pre-treatment years. The pre-trend estimates remain close to zero, and the post-adoption ATTs are identical to those in the main specification.

Table 3 reports estimates of the ATT of Yondr adoption on standardized test scores, averaged across post-adoption periods and adoption cohorts. The table reports effects on the combined test score index, Math scores, and ELA scores, respectively. All outcomes are measured in school-level standard deviations. Column (1) shows that Yondr adoption is associated with a negative effect on the combined test score index, on the order of 0.025 school-level standard deviations, statistically significant at the 5% level. Columns (2) and (3) indicate that this pattern is driven primarily by Math scores, while the estimate for English Language Arts is smaller in magnitude and not statistically distinguishable from zero.

Because these estimates are expressed in school-level standard deviations, they overstate the magnitude of effects relative to the student-level standard deviations typically reported in the education literature. A student-level effect size can be approximated by multiplying the school-level estimate by the square root of the relevant intraclass correlation coefficient. Using the default intraclass correlation of 0.2 for test scores (What Works Clearinghouse, 2020), this implies that student-level effect sizes are roughly half the reported school-level magnitudes. For example, the overall ATT of -0.025 school-level standard deviations corresponds to a student-level effect of approximately -0.012 standard deviations.

Effects of this magnitude are small relative to many other educational interventions. For comparison, Chetty, Friedman and Rockoff (2014) estimate that a one standard deviation increase in teacher value added raises student test scores by about 0.14 standard deviations in math and 0.10 in English. Put differently, our estimated effect corresponds to a decline of less than one-half of a percentile point in the test score distribution (von Hippel, 2025), roughly one-tenth the gain associated with having a higher value-added teacher (Chetty, Friedman and Rockoff, 2014), and about 4 percent of typical annual learning gains at this age (Hill et al., 2008).

Table 3: Difference-in-Differences Estimates of Effects of Phone Pouches on Middle School Test Scores

	Combined score (1)	Math score (2)	ELA score (3)
ATT	-0.025^{**}	-0.029^{**}	-0.017
	(0.012)	(0.014)	(0.013)
Observations	166,471	163,109	163,428

Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption on standardized test scores in grades 6–8. Column (1) reports estimates for the combined Math and English Language Arts test score using the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator. Columns (2) and (3) report results separately for Math and ELA. All specifications use an imbalanced panel. Standard errors are clustered at the school level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A6 reports a series of additional robustness checks. Across a wide range of alternative specifications, the estimated effects remain small and negative. We first assess sensitivity to the estimation approach. Replacing the doubly robust Callaway–Sant'Anna estimator with a two-way fixed effects specification that implements inverse probability weighting and includes the same baseline covariates interacted with year fixed effects yields nearly identical results. Using regression adjustment only, rather than the doubly robust estimator, produces similar estimates.

We next modify the composition of the control group. Using information from the NTS, we restrict the control group to schools reporting no strict phone policy changes and, alternatively, to schools reporting no phone policy changes of any kind in the past five years. We also present estimates that exclude future adopters (i.e., the 2026 adoption cohort) from the control group and another specification that relies solely on within-adopter timing variation (i.e., where the 2026 adoption cohort serves as the only group not treated during our analysis period). Additional specifications limit the treated sample to schools with high-confidence implementation dates, exclude schools with canceled accounts, restrict the sample to a balanced panel, control for the number of students tested, and omit imputed test scores. Appendix E provides further details on these specifications.

Table A7 provides a complementary check by examining how the estimated ATT changes as we incrementally expand the set of baseline covariates included in the specification, culminating in the baseline model used in the main analysis (Column (4)). Column (1) reports estimates from a minimally controlled model that includes only state-by-urbanicity fixed effects. Column (2) adds baseline demographic characteristics, including enrollment, racial composition, and neighborhood poverty. Column (3) further adds grade span indicators, and Column (4) adds charter status. Across specifications, the estimated effect remains small and negative. Importantly, the estimate does not move toward zero as we approach the baseline specification. Because our identification strategy relies on parallel trends conditional on these baseline characteristics, the stability of the estimates across specifications suggests that the results are not driven by differential trends associated with observable school characteristics.

Taken together, these robustness checks reinforce the conclusion from the baseline analysis: over the period we observe, Yondr adoption does not generate measurable positive effects on standardized test scores.

Estimates by cohort

We also examine event-study estimates separately by adoption cohort. Our main event study averages treatment effects across schools adopting in 2023, 2024, and 2025. Figure A5 reports cohort-specific event-study estimates. Two patterns stand out.

First, for each cohort, the pre-adoption coefficients are statistically indistinguishable from zero. As in the pooled specification, there is no evidence of differential pre-trends in test scores prior to Yondr adoption. Second, the post-adoption estimates suggest some attenuation in the negative point estimates across cohorts. For the 2023 cohort, test scores decline by roughly 0.07 school-level standard deviations in the three years after adoption. The 2024 cohort shows negative estimates in both the first and second post-adoption years as well, but these are smaller in magnitude than those for the 2023 cohort and there does not appear to be a downward trend across implementation years. The point estimates in the year of adoption and first year post adoption are both small (-0.027 and -0.017, respectively) and are not statistically significant at the 5% level. For the 2025 cohort—for which we only observe the adoption year—the point estimate is small and positive (0.021), though again not statistically significant.

Student compositional changes

Our analysis relies on school-level outcome data and therefore cannot directly track changes for individual students over time. As a result, one potential concern is that Yondr adoption could induce changes in student composition—for example, if some students leave or enter the school following adoption—leading to changes in average test scores even in the absence of effects on incumbent students.

To assess this possibility, Table A8 examines whether Yondr adoption is associated with changes in school enrollment or student demographics. We examine multiple dimensions of potential compositional change, including total school enrollment, the shares of students who are White, Black, and Hispanic, and the share eligible for free or reduced-price lunch (FRPL). Sample sizes vary across outcomes due to missingness in the NCES data.

Across these measures, we find little evidence of economically meaningful compositional change following Yondr adoption. The estimated change in total enrollment is small (approximately -4.5 students relative to a mean of 540) and statistically indistinguishable from zero. Estimated changes in racial and ethnic composition are also small in magnitude. While the share of White students increases by 0.3 percentage points and the share of Hispanic students increases by 0.7 percentage points, these effects are small relative to their baseline means (53 and 27 percent, respectively). The estimated change in the Black student share is close to zero and not statistically significant. For FRPL, we observe a decline of approximately 0.7 percentage points relative to a baseline mean of 56 percent.

Although some of these estimates are statistically significant, their magnitudes are small. Still, to further assess whether compositional shifts could account for the test score results, Figure A6 reports event-study estimates using as the outcome the fitted value from a school-

level regression of average test scores on total enrollment and the shares of White, Black, Hispanic, and FRPL students. The event-study estimates for the impacts of Yondr adoption on predicted test scores are close to zero throughout the event window. All post-adoption coefficients are small in magnitude and statistically insignificant. The overall ATT is 0.008 with a standard error of 0.007.

Subgroup Analysis

Before exploring any subgroup analysis, we publicly pre-registered four dimensions of heterogeneity that we thought were most important to test. This section presents a subgroup analysis along those four dimensions.

Table A9 reports estimates for two types of heterogeneity. Columns (1) and (2) report effects at the school level, comparing high- and low-poverty schools, defined as schools above and below the within-state median of the neighborhood income-to-poverty ratio. Columns (3) through (9) report effects for within-school student subgroups, including gender, race and ethnicity, and economic disadvantage. The number of observations varies across columns because subgroup-specific test scores are not reported for some school–years with small cell sizes.

Across all pre-registered subgroups, the estimates are negative and generally small in magnitude. Notably, while the point estimates are modest, they tend to be more negative for relatively advantaged students and schools. The estimates are more negative for low-poverty schools than for high-poverty schools, and for students who are not economically disadvantaged than for economically disadvantaged students. Among racial and ethnic groups, the estimates are most negative for White students, while estimates for Black students are closer to zero. In contrast, the point estimates are very similar across gender.

Taken together, these patterns suggest that the absence of positive average effects is not masking large benefits for any particular group. If anything, the results indicate that any small negative effects of phone pouch adoption are more concentrated among relatively advantaged students.

6.3 Effects on School Attendance, Discipline, and Student Survey Outcomes

While standardized test scores capture an important dimension of student outcomes, school smartphone restrictions may also affect other aspects of school climate, student behavior, and students' subjective experiences during the school day. We therefore examine the effects of Yondr adoption on student attendance, disciplinary outcomes, and three student-reported

outcomes from the Panorama surveys: subjective well-being (SWB), classroom attention, and students' perceived likelihood of online bullying. Together, these outcomes provide a broader picture of how smartphone restrictions may impact the school environment and student well-being.

Attendance and Discipline. We begin by presenting event-study estimates for the school-level attendance rate and the standardized disciplinary index in Panels A and B of Figure 8.

For attendance, the event-study coefficients are generally close to zero in the pre-adoption period, with no consistent pattern of differential trends. In the year of adoption, the point estimate is very close to zero. In subsequent years, the estimates become slightly positive and increase gradually, but they remain small in magnitude and are not statistically significant at the 5% level in any post-adoption period.[9]

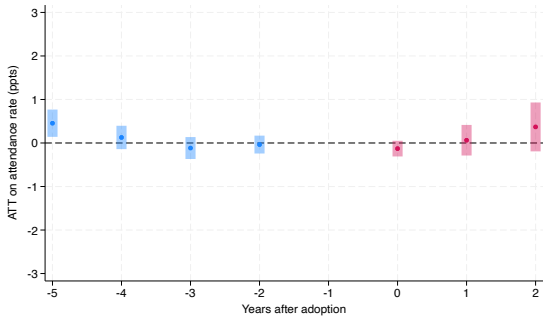
In contrast, the discipline event study shows a clear and immediate response at the time of adoption. Pre-adoption coefficients are tightly centered around zero, indicating no evidence of differential trends prior to treatment. In the year of implementation, disciplinary incidents rise sharply. Importantly, the initial spike in disciplinary incidents is consistent across all three adoption cohorts. However, this effect is short-lived: in the first and second post-adoption years, the estimates are substantially smaller and close to zero, though less precisely estimated. Overall, the pattern indicates a discrete increase in disciplinary incidents at the time of adoption, followed by a return toward baseline levels in subsequent years.

Table 4 reports the corresponding ATTs averaged across cohorts and post-adoption periods. Column (1) shows no statistically significant effect on attendance. The estimated effect is -0.052 percentage points, with a 95% confidence interval ranging from -0.240 to 0.136 percentage points. Relative to a mean attendance rate of 92 percent, this interval rules out economically meaningful changes in either direction. Column (2) reports a statistically significant increase in the standardized discipline index of 0.067 school-level standard deviations. Because the index is standardized within state–year cells, this estimate reflects an increase in disciplinary intensity relative to other schools in the same state and year.

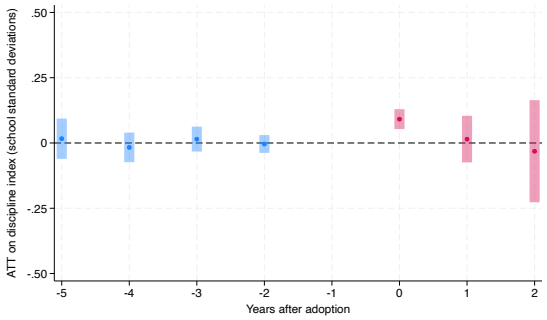
To further probe potential attendance responses, Table A10 examines alternative attendance-related outcomes. We find no statistically or economically meaningful effects on chronic absenteeism. In California, the only state in our data that reports attendance disaggregated by reason, we also find no effect on attendance net of disciplinary suspensions. These results indicate that the increase in disciplinary actions does not mechanically translate into changes in measured attendance, alleviating concerns that attendance effects are being attenuated by increases in out-of-school suspensions and suggesting that the impacts on the disciplinary

[9]There are no clear differential patterns across adoption cohorts.

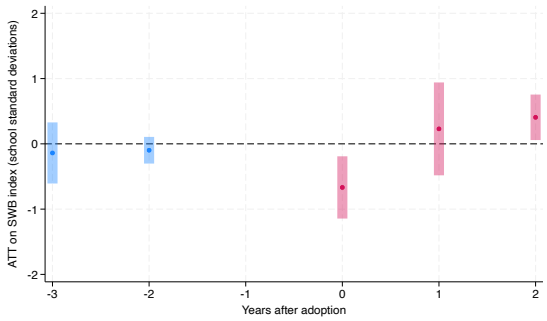
Figure 8: Event Study Estimates of Effects of Phone Pouches on Middle and High School Attendance, Discipline, and Student Survey Outcomes



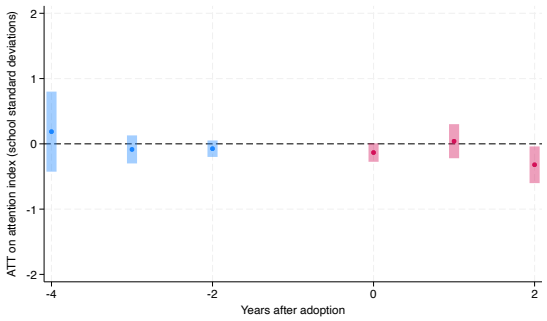
(a) Attendance



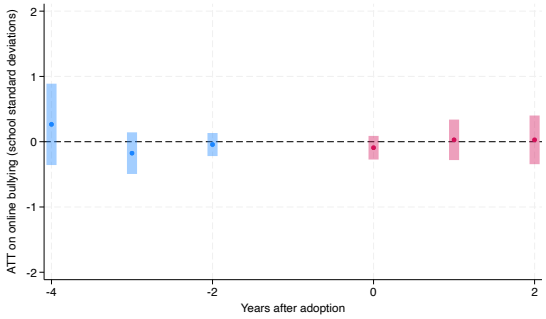
(b) Discipline Index



(c) SWB Index



(d) Classroom Attention Index



(e) Online Bullying

Notes: This figure reports event-study estimates of the effects of Yondr phone pouch adoption on attendance, discipline, subjective wellbeing, attention, and online bullying. Estimates are based on the Callaway and Sant'Anna (2021) difference-in-differences framework and plot event-time averages of group-time average treatment effects, $\text{ATT}(g,t)$, aggregated across adoption cohorts and normalized relative to the year immediately preceding adoption ($t = -1$). All specifications include the same baseline covariates and weighting scheme as in the primary test score analysis. The shorter pre-treatment windows for the SWB, classroom attention, and online bullying outcomes reflect the later start of Panorama survey coverage.

Table 4: Difference-in-Differences Estimates of Effects of Phone Pouches on Middle and High School Attendance, Discipline, and Student Survey Outcomes

	Attendance rate (1)	Discipline index (2)	SWB index (3)	Attention index (4)	Online bullying (5)
ATT	−0.052	0.067***	−0.365*	−0.113	−0.044
	(0.096)	(0.023)	(0.213)	(0.074)	(0.096)
Mean of dependent variable	91.85	–	–	–	–
Observations	196,647	162,658	3546	9388	5364

Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption using the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator. Column (1) reports effects on the attendance rate. Column (2) reports effects on a standardized school-level discipline index constructed from suspension and discipline measures and standardized within state-by-year cells. Column (3) reports effects on a school-level subjective wellbeing index constructed from Panorama survey questions. Column (4) reports effects on a school-level classroom attention index constructed from Panorama survey questions. Column (5) reports effects on a standardized school-level online bullying measure constructed from a panorama survey question. The estimates in Column (1) are in percentage points, and the estimates in columns 2-5 are in school-level standard deviations. All specifications use an imbalanced panel and include state-by-urbanicity fixed effects and baseline covariates. Standard errors are clustered at the school level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

index may operate primarily through in-school suspensions or other disciplinary incidents such as detentions.

Student Survey Outcomes. We next present event-study estimates for student-reported SWB, classroom attention, and online bullying in Panels C through E of Figure 8.[10]

The event-study estimates for SWB show no evidence of differential pre-trends: coefficients in the pre-adoption periods are statistically indistinguishable from zero. In the year of implementation, subjective well-being declines sharply by approximately 0.668 school-level standard deviations, and the estimate is statistically significant. In subsequent years, the estimates recover steadily. By the second post-adoption year, the point estimate becomes positive and statistically significant, on the order of 0.407 school-level standard deviations. This pattern indicates an initial decline in reported well-being followed by a rebound in later years.

Because these estimates are expressed in school-level standard deviations, they overstate the magnitude of effects relative to the student-level standard deviations. A student-level effect size can be approximated by multiplying the school-level estimate by the square root of the relevant intraclass correlation coefficient. Using the default intraclass correlation of 0.1 for behavioral and survey-based outcomes (What Works Clearinghouse, 2020), this implies

[10]Because Panorama survey outcomes are available for a substantially smaller set of school–year observations, we estimate these effects using the regression-adjustment implementation of the Callaway–Sant'Anna estimator rather than the doubly robust version used in our primary specifications. With the smaller sample, the doubly robust estimator fails to compute some cohort-by-year ATTs, whereas the outcome-regression-only implementation is well behaved. In practice, as shown above for test scores, the two approaches yield nearly identical estimates in settings where both can be implemented.

that student-level effect sizes are roughly one-third of the reported school-level magnitudes. Under this adjustment, the estimated decline in the year of adoption corresponds to approximately 0.214 student-level standard deviations, while the second-year increase corresponds to roughly 0.13 student-level standard deviations. For comparison, Allcott et al. (2020) find that deactivating Facebook for four weeks increased subjective well-being by 0.09 standard deviations.

In contrast, the event-study estimates for classroom attention show no evidence of improvement following adoption. None of the pre-adoption coefficients are statistically significant, although some are imprecisely estimated. In the year of adoption, the point estimate is modestly negative and not statistically significant at the 5% level. In the first post-adoption year, the estimate is slightly positive and not statistically significant. By the second post-adoption year, the estimate turns negative and is statistically significant. Across periods, the estimates do not exhibit a consistent pattern, and none provide evidence of measurable improvements in classroom attention.

The event-study estimates for students' perceived likelihood of online bullying show a similar absence of systematic effects. There is no evidence of differential pre-trends, and none of the post-adoption coefficients are statistically significant. Across event time, the estimates fluctuate around zero without exhibiting a consistent pattern, providing no evidence that Yondr adoption measurably changes students' reported likelihood of being bullied online.

Table 4 reports the corresponding ATTs averaged across cohorts and post-adoption periods. Column (3) shows that the estimated average effect on SWB is -0.365 school-level standard deviations. Although negative in sign, the estimate is not statistically significant at the 5% level. The confidence interval rules out even modest positive effects on average subjective well-being over the post-adoption period. However, as the event-study estimates demonstrate, this average masks substantial dynamics: a sharp initial decline followed by a subsequent rebound. Column (4) reports an estimated effect on classroom attention of -0.113 school-level standard deviations. This estimate is also negative and not statistically significant at the 5% level. The confidence interval rules out even small improvements in classroom attention following Yondr adoption. Column (5) reports an estimated effect on students' perceived likelihood of online bullying of -0.044 school-level standard deviations (SE = 0.096). Although relatively more imprecise, this estimate is small in magnitude and not statistically significant, providing no evidence that Yondr adoption measurably changes students' reported likelihood of being bullied online.

Finally, because our test score analyses focus exclusively on middle school grades (6–8), we also report ATT estimates for attendance, discipline, and survey outcomes restricting the sample to middle schools only. Table A11 presents these results. The estimates are

qualitatively similar to those in Table 4. We continue to find no statistically significant effects on attendance and no evidence of measurable improvements in SWB, classroom attention, or students' perceived likelihood of online bullying in the post-adoption period. For disciplinary incidents, the point estimates remain positive and similar in magnitude to the baseline results, but are less precisely estimated and no longer statistically significant, consistent with the significantly smaller sample size in this specification.

Interpretation. Taken together, these findings suggest that lockable phone pouches substantially reduce in-school smartphone use and meaningfully alter aspects of the school environment. Adoption is associated with a sharp initial increase in disciplinary incidents and dynamic changes in students' reported well-being. Yet we find no evidence of measurable improvements in standardized test scores, attendance, self-reported classroom attention, or students' perceived likelihood of online bullying. In other words, while phone pouch policies tighten access to smartphones and shift aspects of school climate, these changes do not translate into detectable gains in academic focus or achievement over the period we observe.

Several mechanisms could plausibly explain this set of results. One possibility is that implementing a binding restriction on smartphone use generates short-term disruption. Enforcing new rules may require additional monitoring and disciplinary interactions, particularly in the initial period after adoption. Such disruptions could fragment instructional time or alter classroom dynamics in ways that offset any potential gains from reduced phone access. The subsequent recovery in well-being is consistent with adjustment and adaptation over time as students and staff acclimate to the new policy environment.

Another possibility is that restricting smartphone access changes how students allocate attention during the school day. If smartphones previously absorbed attention that might otherwise have been directed toward peers, their removal could increase in-person interaction. Greater peer interaction may raise opportunities for social engagement but also for conflict and rule enforcement, particularly during the transition to a new policy environment, potentially contributing to higher disciplinary incidents without translating into measurable improvements in classroom attention or academic performance. While we do not directly observe changes in peer interaction, this interpretation is consistent with the observed short-run increase in disciplinary incidents. Moreover, a peer interaction story could also be consistent with the dynamic pattern in subjective well-being if the benefits of increased in-person interaction emerge only after an initial period of adjustment.

These interpretations are necessarily speculative and not exhaustive. Our data do not directly measure classroom interactions or enforcement practices, and additional qualitative or observational evidence would be needed to distinguish among possible mechanisms.

7 Conclusion

Smartphone restrictions in schools have become increasingly common in the United States and around the world, motivated by concerns that constant access to digital devices may undermine learning, attention, and student well-being. We study a particularly well-defined and rapidly expanding form of restriction—lockable phone pouches that physically prevent students from accessing their devices during the school day.

Using nationwide data and a staggered difference-in-differences design, we find that the adoption of lockable pouches substantially reduces in-school smartphone activity as measured by independent GPS data. The first-stage effects are large and persistent, indicating that lockable pouches meaningfully tighten phone access relative to prevailing alternatives.

Despite these sizable reductions in smartphone use, we find no evidence that Yondr adoption measurably improves standardized test scores, attendance, or student-reported measures of classroom attention or online bullying. Estimated effects on test scores are small in magnitude and generally negative. In contrast, adoption is associated with a clear initial increase in disciplinary incidents and dynamic changes in student-reported subjective well-being: an initial decline in the year of implementation followed by improvements in subsequent years.

Our study leaves several important questions for future research. We observe outcomes for at most three years beyond the year of adoption for most cohorts, and longer-run effects remain an open question. In addition, we focus on one specific and well-defined implementation design. Other forms of smartphone restrictions—particularly those that rely more heavily on classroom-level discretion or that limit access only during instructional time—may operate differently. Evaluating the longer-run impacts of phone restrictions and comparing alternative policy designs are important priorities as schools continue to experiment with approaches to managing digital access.

References

Abrahamsson, Sara. 2024. "Smartphone Bans, Student Outcomes and Mental Health." NHH Department of Economics Discussion Paper 01.

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. "The Welfare Effects of Social Media." *American Economic Review*, 110(3): 629–76.

Allcott, Hunt, Matthew Gentzkow, Benjamin Wittenbrink, Juan Carlos Cisneros, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Sandra González-Bailón, Andrew M. Guess, Young Mie Kim, David Lazer, Neil Malhotra, Devra Moehler, Sameer Nair-Desai, Brendan Nyhan, Jennifer Pan, Jaime Settle, Emily Thorson, Rebekah Tromble, Carlos Velasco Rivera, Arjun Wilkins, Magdalena Wojcieszak, Annie Franco, Chad Kiewiet de Jonge, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker. Forthcoming. "The Effect of Deactivating Facebook and Instagram on Users' Emotional State." *American Economic Journal: Economic Policy*.

Barwick, Panle Jia, Siyu Chen, Chao Fu, and Teng Li. 2026. "Digital distractions with peer influence: The impact of mobile app usage on academic and labor market outcomes." *The Quarterly Journal of Economics*, 141(1): 1–49.

Beland, L. P., and R. Murphy. 2015. "Technology, distraction & student performance." Centre for Economic Performance, London.

Braghieri, Luca, Ro'ee Levy, and Alexey Makarin. 2022. "Social Media and Mental Health." *American Economic Review*, 112(11): 3660–93.

Callaway, B., and P. H. C. Sant'Anna. 2021. "Difference-in-Differences with multiple time periods." *Journal of Econometrics*, 225(2): 200–230.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9): 2593–2632.

Cook, Cody, Aboudy Kreidieh, Shoshana Vasserman, Hunt Allcott, Neha Arora, Freek van Sambeek, Andrew Tomkins, and Eray Turkel. 2025. "The Short-Run Effects of Congestion Pricing in New York City." National Bureau of Economic Research.

D'Addio, Anna Cristina. 2025. "The "quiet" revolution in schools: More and more countries are locking up phones – Part 1." *UNESCO Global Education Monitoring Report*.

Duckworth, Angela L, Patrick D Quinn, and Eli Tsukayama. 2012. "What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades." *Journal of educational psychology*, 104(2): 439.

Figlio, David N., and Umut Ozek. 2025. "The Impact of Cellphone Bans in Schools on Student Outcomes: Evidence from Florida." National Bureau of Economic Research Working Paper 34388.

Gajdics, Janka, and Balázs Jagodics. 2022. "Mobile phones in schools: With or without you? Comparison of students' anxiety level and class engagement after regular and mobile-free school days." *Technology, Knowledge and Learning*, 27(4): 1095–1113.

Goldhaber, Dan, Nick Huntington-Klein, Nate Brown, Scott Imberman, and Katharine O Strunk. 2024. "The Call Is Coming from inside the School! How Well Does Cell Phone Data Predict Whether K12 School Buildings Were Open during the Pandemic? Working Paper No. 309-1124." *National Center for Analysis of Longitudinal Data in Education Research (CALDER)*.

Haidt, Jon. 2023. "The Case for Phone-Free Schools." After Babel. Accessed: 2026-02-15.

Hess, Rick. 2026. "What's the Right Way to Limit Phones in School?" *Education Week*.

Higa, Minoru. 2024. "Traffic congestion and labor supply: Evidence from smartphone data." *Working Paper*.

Hill, Carolyn J, Howard S Bloom, Alison Rebeck Black, and Mark W Lipsey. 2008. "Empirical benchmarks for interpreting effect sizes in research." *Child development perspectives*, 2(3): 172–177.

Jacob, Brian A., and Lars Lefgren. 2003. "Are Idle Hands the Devil's Workshop? Incapacitation, Concentration, and Juvenile Crime." *American Economic Review*, 93(5): 1560–1577.

Kamenetz, Anya. 2025. "How the Phone Ban Saved High School." *New York Magazine*.

Kessel, D., H. L. Hardardottir, and B. Tyrefors. 2020. "The impact of banning mobile phones in Swedish secondary schools." *Economics of Education Review*, 77: 102009.

Kircher, Madison Malone, and Callie Holtermann. 2026. "A Blow to the PhoneFree Classroom." *The New York Times*. Accessed March 3, 2026.

Prothero, A, L Langreo, and A Klein. 2026. "Which States Ban or Restrict Cellphones in Schools? A look at statewide laws and policies on cellphones in schools." *EducationWeek*.

Rahali, Miriam, Beeban Kidron, and Sonia Livingstone. 2024. "Smartphone Policies in Schools: What Does the Evidence Say?"

Sant'Anna, Pedro HC, and Jun Zhao. 2020. "Doubly robust difference-in-differences estimators." *Journal of econometrics*, 219(1): 101–122.

Shi, Ying, and Francisco Villarroel. 2026. "The Consequences of Cellphone Restrictions in Classrooms." *Working Paper*.

Sungu, Alp, Pradeep Kumar Choudhury, and Andreas Bjerre-Nielsen. 2025. "Removing Phones from Classrooms Improves Academic Performance." SSRN Working Paper 5370727.

von Hippel, Paul T. 2025. "Multiply by 37 (or divide by 0.027): A surprisingly accurate rule of thumb for converting effect sizes from standard deviations to percentile points." *Educational evaluation and policy analysis*, 47(3): 960–969.

What Works Clearinghouse. 2020. "What Works Clearinghouse Procedures Handbook." National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Version 4.1.

Zheng, Yunhan, David R Keith, Shenhao Wang, Mi Diao, and Jinhua Zhao. 2024. "Effects of electric vehicle charging stations on the economic vitality of local businesses." *Nature communications*, 15(1): 7437.

Appendix

Table of Contents

A Additional Figures **49**

B Additional Tables **59**

C Data Appendix **75**

C.1 Additional Details on Yondr Data 75

C.2 Additional Details on GPS Data 78

C.3 Wording of NTS Questions Used in the Analysis 82

C.4 Wording of Gallup Questions Used in the Analysis and Weighting Scheme . . 84

C.5 Construction of Test Score, Attendance, and Discipline Samples 85

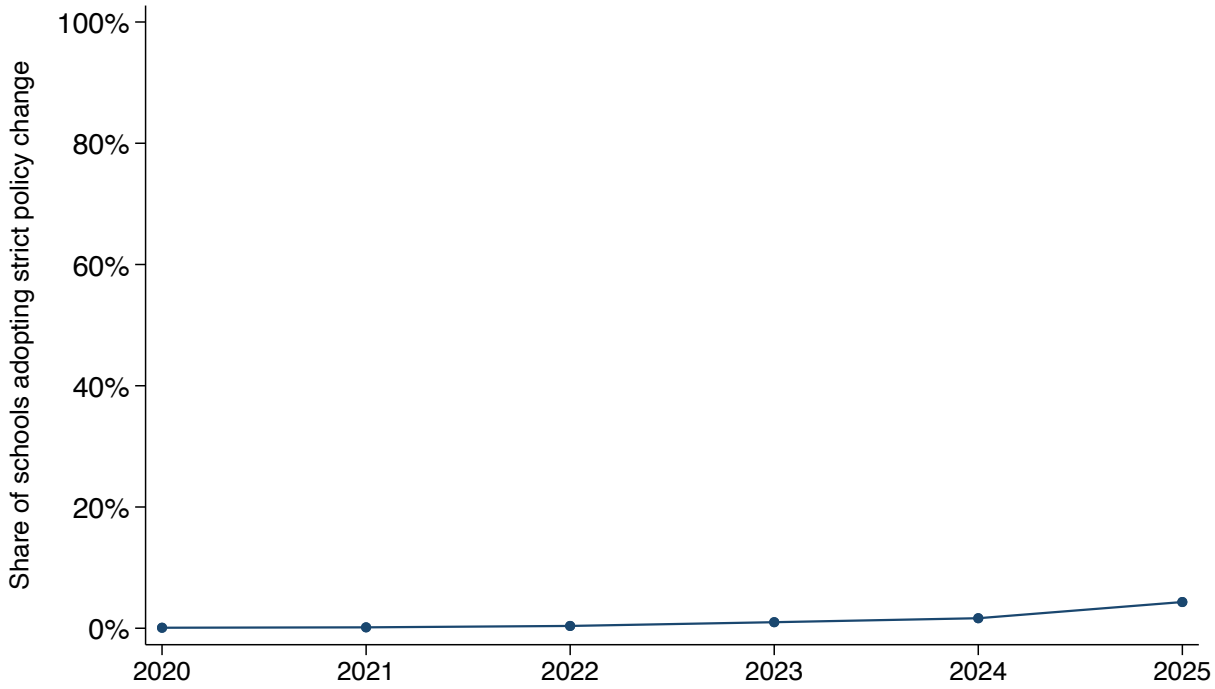
C.6 Additional Details on Panorama Data and Analysis Sample 88

D Imputation of Missing Test Scores **94**

E Additional Robustness Checks for Test Score Estimates **95**

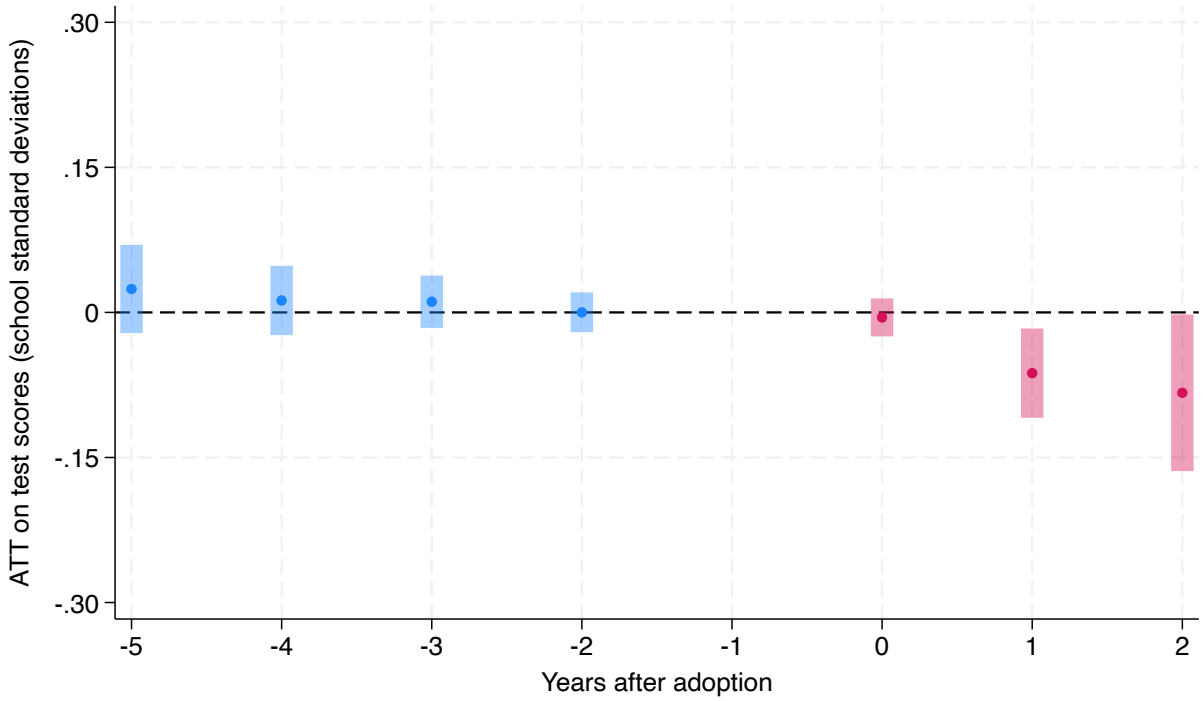
A Additional Figures

Figure A1: Adoption of Strict Non-Yondr Phone Policies Among Control Schools in NTS



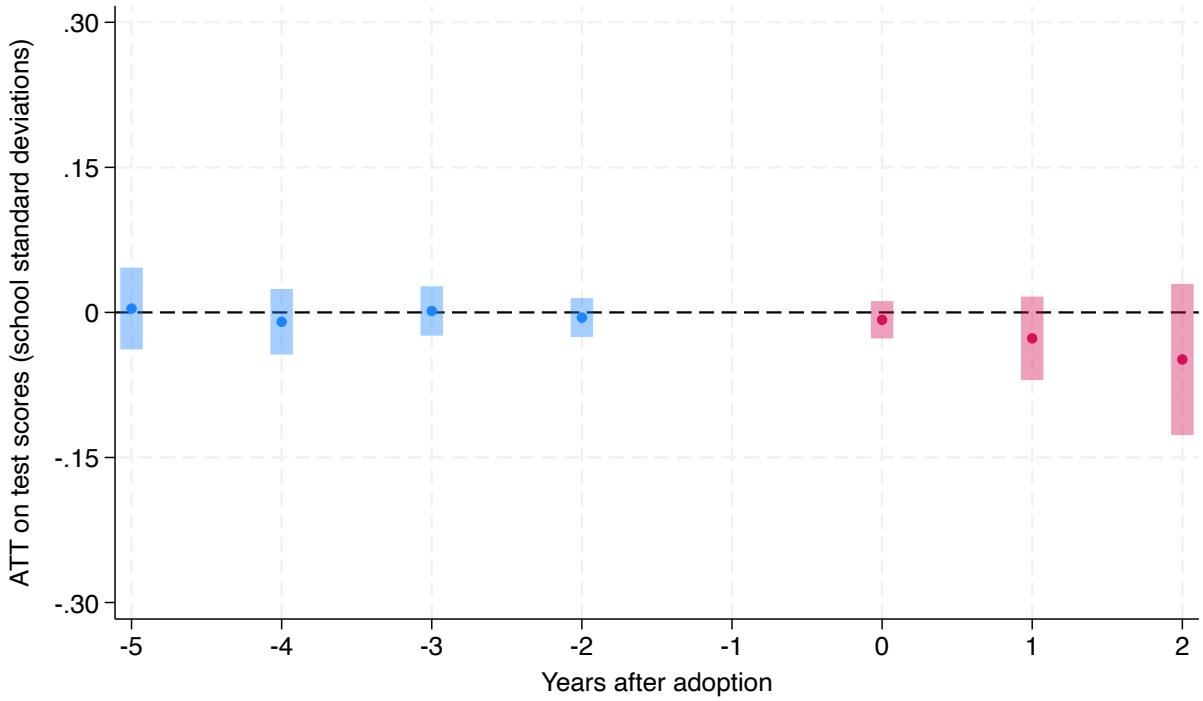
Notes: This figure uses data from the NTS to show phone policy changes among schools that do not adopt Yondr. We combine responses to Q10–Q11 (whether the school experienced a major phone policy change in the past five years and the timing of that change) with the school's current policy reported in Q3 ("Where are students allowed to keep their phones?"). See Appendix C.3 for the exact wording of each question. We restrict attention to schools whose modal Q3 response does not indicate Yondr adoption. For these non-Yondr schools, we classify a policy change as *strict* if the modal current policy is lockers only, centralized collection, or no phones on campus. For schools with multiple teacher responses, policy status and timing are defined using the modal school-level response. The bars report the percentage of non-Yondr schools adopting a strict non-Yondr policy in each year. Because the survey was launched in 2025 and asks about policy changes within the previous five years, years prior to 2020 are not shown.

Figure A2: Event Study Estimates of Effects of Phone Pouches on Middle School Math Test Scores



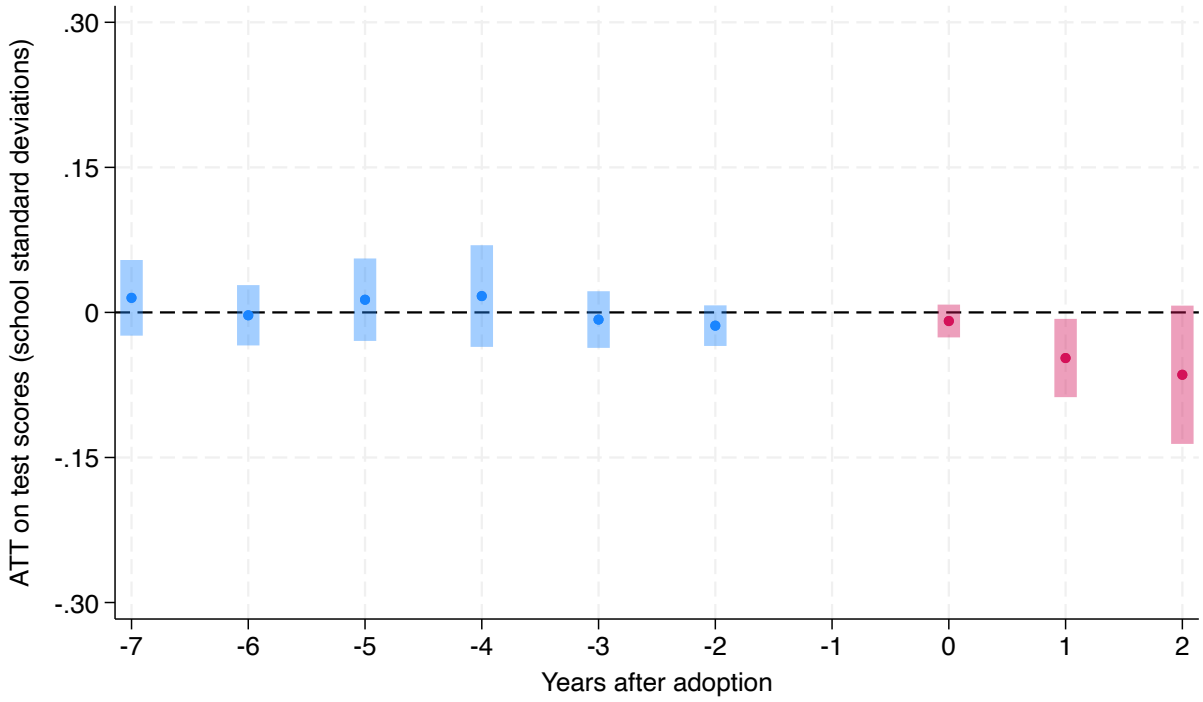
Notes: This figure reports event-study estimates of the effect of Yondr adoption on average Math test scores in grades 6–8, measured in school-level standard deviations. Estimates are based on the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator. The plotted coefficients correspond to event-time averages of group-time average treatment effects, $\text{ATT}(g,t)$, aggregated across adoption cohorts and normalized relative to the year immediately preceding adoption ($t = -1$). Never-treated schools serve as the comparison group. Covariates and weighting follow the main empirical specification.

Figure A3: Event Study Estimates of Effects of Phone Pouches on Middle School ELA Test Scores



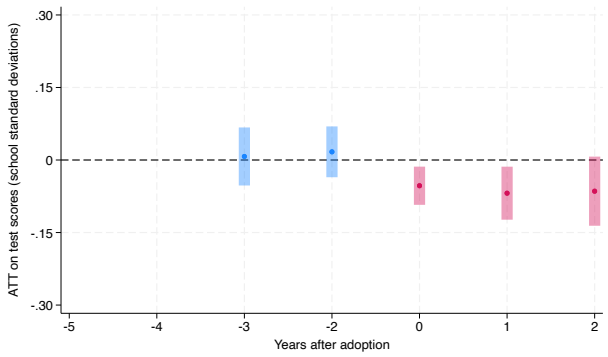
Notes: This figure reports event-study estimates of the effect of Yondr adoption on average English Language Arts (ELA) test scores in grades 6–8, measured in school-level standard deviations. Estimates are based on the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator. The plotted coefficients correspond to event-time averages of group-time average treatment effects, ATT(g, t), aggregated across adoption cohorts and normalized relative to the year immediately preceding adoption ($t = -1$). Never-treated schools serve as the comparison group. Covariates and weighting follow the main empirical specification.

Figure A4: Event Study Estimates of Effects of Phone Pouches on Middle School Test Scores, Without Recoding 2018 and 2019

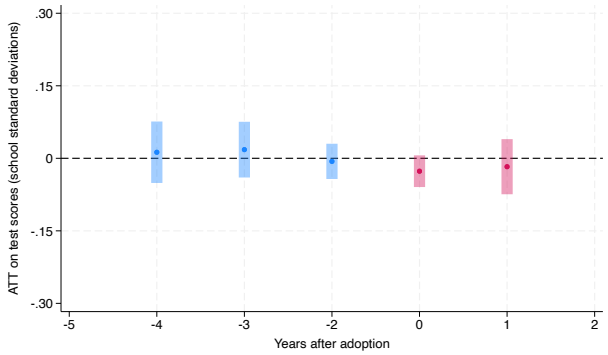


Notes: This figure reports event-study estimates of the effects of Yondr phone pouch adoption on average standardized test scores for grades 6-8 at the school level, measured in school-level standard deviations. As opposed to Figure 7, it does not recode years 2018 and 2019 to 2020 and 2021. Estimates are based on the Callaway and Sant'Anna (2021) difference-in-differences framework and plot event-time averages of group-time average treatment effects, $\text{ATT}(g,t)$, aggregated across adoption cohorts and normalized relative to the year immediately preceding adoption ($t = -1$). All specifications include the same baseline covariates and weighting scheme as in the primary test score analysis.

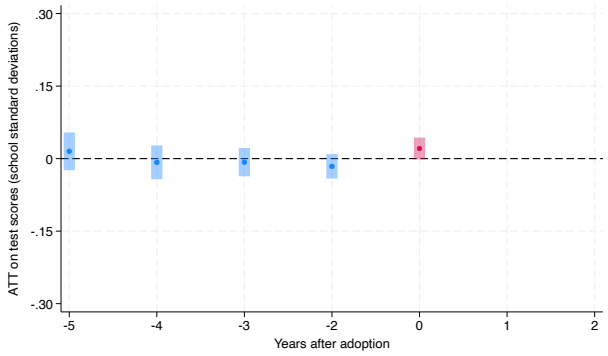
Figure A5: Event Study Estimates of Effects of Phone Pouches on Middle School Test Scores by Cohort



(a) G = 2023



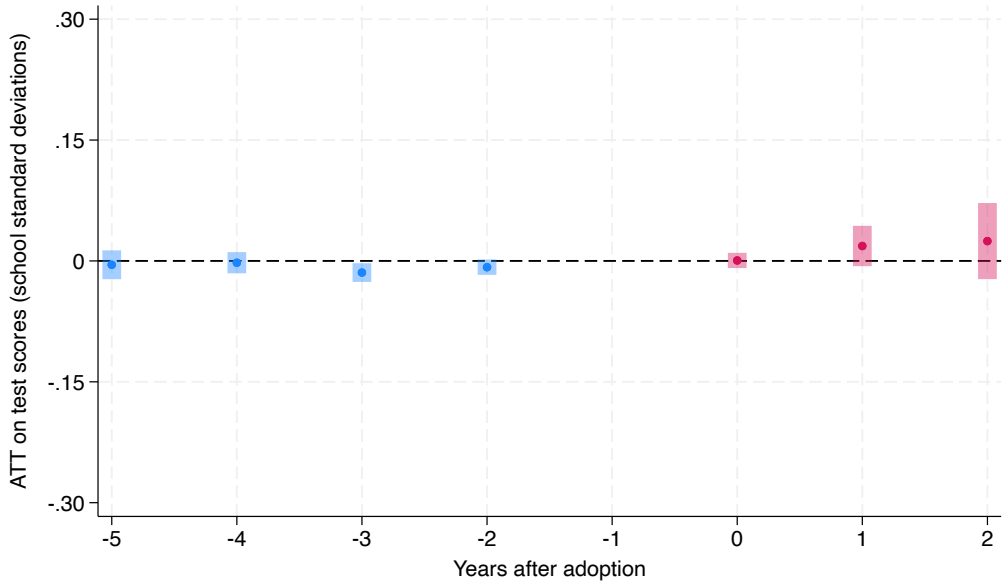
(b) G = 2024



(c) G = 2025

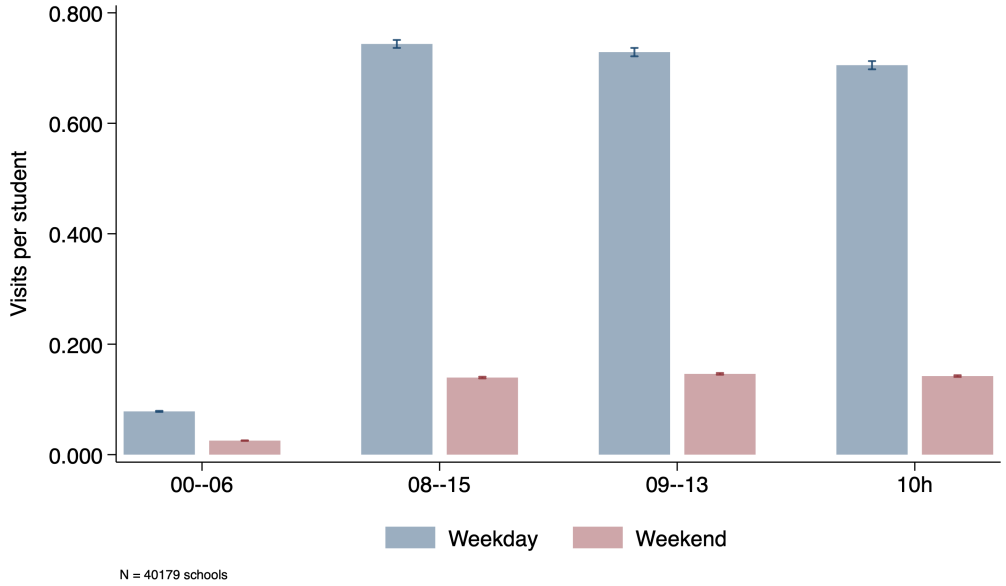
Notes: This figure reports event-study estimates of the effects of Yondr phone pouch adoption on average standardized test scores for grades 6–8 at the school level, measured in school-level standard deviations and shown separately by adoption cohort. Estimates are based on the Callaway and Sant'Anna (2021) difference-in-differences framework and plot cohort-specific group-time average treatment effects, $\text{ATT}(g,t)$, normalized relative to the year immediately preceding adoption ($t = -1$). All specifications include the same baseline covariates and weighting scheme as in the primary test score analysis.

Figure A6: Event Study Estimates of Effects of Phone Pouches on Predicted Middle School Test Scores



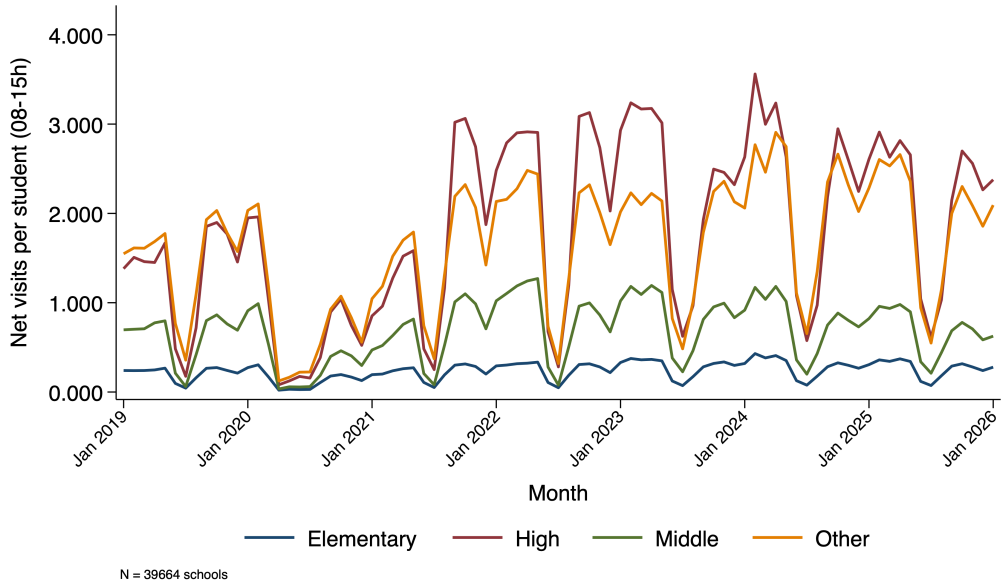
Notes: This figure reports event-study estimates of the effect of Yondr adoption on predicted test scores. The outcome is the fitted value from a regression of average test scores on the five time-varying compositional measures reported in Table A8 (total enrollment, racial/ethnic shares, and FRPL share). The overall ATT is 0.008 (SE = 0.007). Estimates are based on the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator. The plotted coefficients correspond to event-time averages of group-time average treatment effects, ATT(g, t), aggregated across adoption cohorts and normalized relative to the year immediately preceding adoption ($t = -1$). Covariates and weighting follow the main empirical specification.

Figure A7: Visits: Weekdays vs Weekends (Advan)



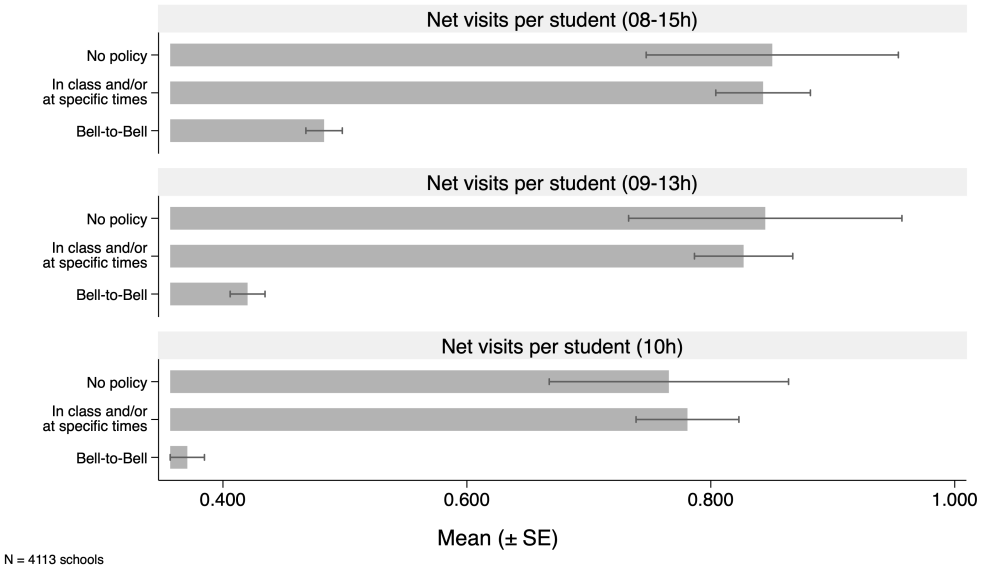
Notes: This figure compares smartphone visits during school-hour windows on weekdays and weekends using Advan GPS data. For each school, visits are measured as average device activity within school buildings during the indicated hour ranges, normalized by enrollment to obtain per-student values. The plotted bars show mean visits separately for weekdays and weekends across alternative hour windows. The plot includes all schools and all months in the sample; the average of visits per student is therefore under 1 as it includes Elementary and Combined schools, as well as summer months, which have lower activity.

Figure A8: Net Visits Time Series (Advan)



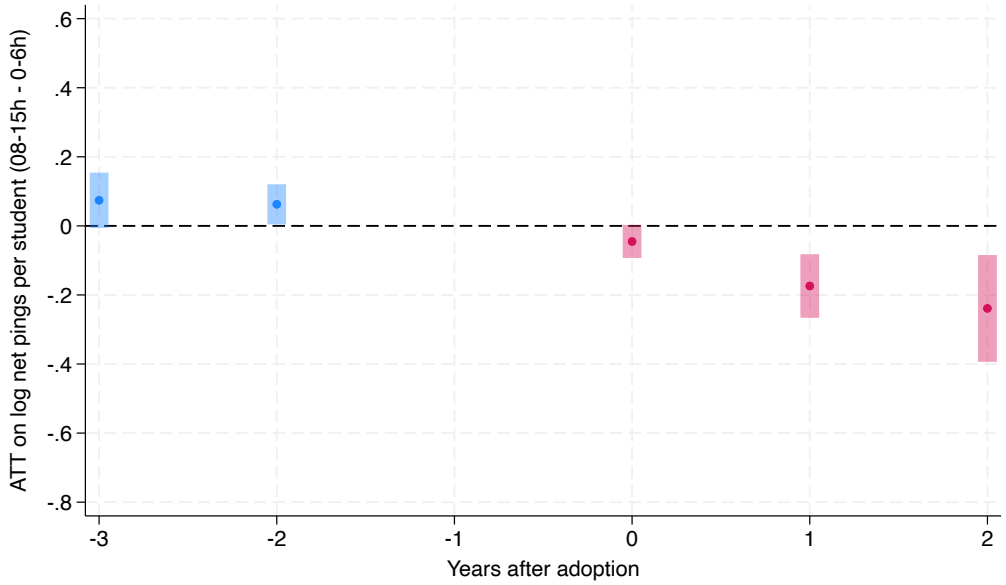
Notes: The figure reports the average net visits by school level in the NCES using smartphone location data from Advan. Net visits are defined as average weekday school-hours visits (08-15h) minus the corresponding weekend activity in the same month, further subtracting the weekday–weekend difference in nighttime activity (00:00–06:00). A school-level average for each month is computed, and the plot shows the mean of the distribution of schools by level.

Figure A9: Net visits by smartphone policy



Notes: The figure reports school-level averages constructed from the NTS and smartphone location data from Advan. Schools are grouped by the modal response to when students may have their phones (Q2), with categories corresponding to away-for-the-day (bell-to-bell) restrictions, schedule-based restrictions (e.g., use permitted during lunch or between classes), and no school-wide restriction. Net visits are defined as average weekday school-hours visits for three different intervals (08h-15h, 10h-11h, 09-13h) minus the corresponding weekend activity in the same month, further subtracting the weekday–weekend difference in nighttime activity (00:00–06:00). A school-level average for each month is computed, and the plot shows the mean and standard errors of the distribution of schools using data for November 2025 (representative of when a large share of respondents took the NTS). See Appendix C.3 for the exact wording of each survey question.

Figure A10: Treatment–Control Smartphone GPS Pings Over Time (Quadrant data)



Notes: This figure reports event-study estimates of the effect of Yondr adoption on smartphone use, measured using GPS location data from Quadrant. Estimates are based on the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator. The plotted coefficients correspond to event-time averages of group-time average treatment effects, $\text{ATT}(g,t)$, aggregated across adoption cohorts and normalized relative to the year immediately preceding adoption ($t = -1$). The outcome is the natural logarithm of net smartphone pings per student during school hours (8:00–15:00) within school buildings. Net pings are defined as average weekday school-hours pings (computed as the total ping count over the interval) minus the corresponding weekend activity in the same month, further subtracting the weekday–weekend difference in nighttime activity (00:00–06:00). The sample is restricted to middle and high schools, and never-treated schools serve as the comparison group. Covariates and weighting follow the main empirical specification.

B Additional Tables

Table A1: Comparison of NCES, NTS, and Panorama Samples

	NCES (1)	NTS (2)	Panorama (3)
Total enrollment	551	843	788
Share White	0.52	0.57	0.35
Share Black	0.15	0.15	0.18
Share Hispanic	0.24	0.20	0.36
Urban	0.27	0.24	0.42
Suburban	0.28	0.34	0.37
Rural/Town	0.45	0.42	0.20
Charter school	0.13	0.07	0.03
Schools with grades above 9	0.40	0.52	0.37
Schools with grades below 6	0.36	0.23	0.37
Income-to-poverty ratio	298	330	326
Number of schools	65,647	13,043	7280

Notes: This table reports baseline school characteristics for middle- and high-school observations in three samples: the NCES public-school universe (Column 1), the NTS (Column 2), and the subset of schools with non-missing Panorama student survey outcomes for either the classroom attention, online bullying, or subjective well-being indices (Column 3). Statistics are computed using the earliest observed year available for each school. Income-to-poverty ratio is measured using EDGE School Neighborhood Poverty Estimates and reflects the ratio of neighborhood median household income to the federal poverty threshold. All statistics are reported at the school level.

Table A2: State-specific data availability and definitions

State Name	Attendance Rate	Chronic Absenteeism	Disciplinary Measures	Middle School Test Source
Alabama	Yes (def1; a6)	Yes (def2)	No (a2)	SEDA
Alaska	Yes (def5, a7)	Yes (def1;a7)	No (a2)	No data
Arizona	No (a1)	No (a1)	No (a1)	SEDA
Arkansas	Yes (def1)	No (a1)	Yes (def3)	SEDA
California	Yes (def 3)	Yes (def1)	Yes (def2)	Zelma
Colorado	Yes (def1)	Yes (def1;a7)	No (a2)	Zelma
Connecticut	Yes (def5)	Yes (def1)	Yes (def2)	Zelma
Delaware	Yes (def1)	Yes (def1)	Yes (def3b)	Zelma
District of Columbia	No (a2)	No (a2)	No (a2)	SEDA
Florida	Yes (def1)	Yes (def;a5)	Yes (def3a;a5)	Zelma
Georgia	Yes (def1)	Yes (def1)	Yes (def3;a5)	SEDA
Hawaii	Yes (def1)	No (a1)	Yes (def2a; note1)	SEDA
Idaho	No (a1)	No (a1)	No (a1)	SEDA
Illinois	Yes (def5)	Yes (def1)	Yes (def1;a6;note1)	SEDA
Indiana	Yes (def1)	Yes (def1;a6)	Yes (def3)	SEDA
Iowa	Yes (def1)	Yes (def1)	Yes (def2; note1)	SEDA
Kansas	Yes (def5;a5)	Yes (def1;a6)	No (a1)	SEDA
Kentucky	Yes (def1;a7)	Yes (def1)	No (a1)	SEDA
Louisiana	Yes (def5)	Yes (def1;a7)	Yes (def3)	Zelma
Maine	Yes (def1)	Yes (def1)	Yes (def3a; note1)	SEDA
Maryland	Yes (def1)	Yes (def1)	No (a1)	SEDA
Massachusetts	Yes (def5)	Yes (def1)	Yes (def1;a5)	Zelma
Michigan	Yes (def1)	Yes (def1)	No (a4)	Zelma
Minnesota	No (a4)	Yes (def1; a7)	Yes (def2)	Zelma
Mississippi	No (a1)	Yes (def1)	Yes (def3)	SEDA
Missouri	Yes (def2)	Yes (def1)	No (a3)	SEDA
Montana	Yes (def1)	Yes (def1, note1)	No (a1)	Excluded (note3)
Nebraska	Yes (def1)	Yes (def1)	Yes (def1;a6; note1)	Zelma
Nevada	Yes (def5)	Yes (def1)	No (a2)	SEDA
New Hampshire	No (a2)	No (a2)	Excluded (note2)	Zelma
New Jersey	No (a1)	Yes (def1;a5)	Yes (def2;a5)	Zelma
New Mexico	No (a1)	No (a1)	No (a2)	SEDA
New York	Yes (def1;a5)	Yes (def1; a5)	Yes (def5; a5; note1)	Zelma
North Carolina	Yes (def4;a5, a6)	Yes (def1;a5)	No (a1)	Zelma
North Dakota	Yes (def5)	Yes (def1)	Yes (def3a; note1)	SEDA
Ohio	Yes (def2)	Yes (def1)	Yes (def1)	SEDA
Oklahoma	No (a4)	Yes (def1)	No (a2)	Zelma
Oregon	No (a1)	Yes (def1)	No (a1)	SEDA
Pennsylvania	Yes (def5;a5)	Yes (def1;a5)	Yes (def2)	SEDA
Rhode Island	No (a4)	Yes (def1;a6)	Yes (def4;a6)	Zelma
South Carolina	No (a4)	Yes (def1)	Yes (def3; note1)	Zelma
South Dakota	Yes (def5)	Yes (def1)	No (a1)	Excluded (note3)
Tennessee	Yes (def5)	Yes (def1; a5)	Yes (def6)	SEDA
Texas	Yes (def1;a5)	Yes (def1;a5,a7)	Yes (def1, note1)	Zelma
Utah	Yes (def1)	Yes (def1)	No (a2)	SEDA
Vermont	No (a1)	No (a1)	No (a3)	Zelma
Virginia	Yes (def1)	Yes (def1)	Yes (def3;a5,a6;note1)	Zelma
Washington	No (a1)	Yes (def1)	No (a1)	SEDA
West Virginia	Yes (def5;a5,a6)	No (a1)	No (a1)	SEDA
Wisconsin	Yes (def1;a5)	Yes (def1;a5)	Yes (def5;a5; note1)	Zelma
Wyoming	No (a1)	No (a1)	No (a4)	Zelma

Notes: For the first three columns, "Yes" and "No" indicate whether the outcome is used in the analysis for that state. Attendance definitions: def1 = aggregated days attended / aggregated days enrolled; def2 = aggregated hours attended / aggregated hours possible to attend; def3 = total days absent for any reason / (cumulative enrollment × average instructional days); def4 = average daily attendance computed for each grade level and then averaged across grades; def5 = pending confirmation. Chronic absenteeism definitions: def1 = absent for 10% or more for any reason; def2 = absent for 18 or more days for any reason; def3 = pending confirmation by the state. Disciplinary measure definitions: def1 = % of students with any disciplinary measures taken; def2 = % of students placed in any suspension; def2a = incident counts of any suspensions divided by enrollment; def3 = % of students placed in in-school suspension plus % placed in out-of-school suspension; def3a = incident counts of in-school plus out-of-school suspension divided by enrollment; def3b = % of students placed in in-school suspension plus % placed in out-of-school suspension without CDAP; def4 = % of students placed in in-school suspension; def5 = % of students placed in out-of-school suspension; def6 = def3 for 2018–2019 and def1 for 2020–2025. Data availability codes: a1 = data request pending; a2 = data request declined; a3 = majority of observations suppressed; a4 = data not available at the school level; a5 = missing 2025 data; a6 = missing years other than 2025 or 2020, and requested; a7 = missing years other than 2025 or 2020, and not available. Other notes: note1 = percentage computed using enrollment counts; note2 = excluded due to the insufficient number of Yondr adopters; note3 = no data for schools that adopted Yondr.

Table A3: Comparison of All Yondr Adopters versus Yondr Adopters in NTS

	Non-adopters (1)	Adopters (2)	Adopters in PIF (3)
Total enrollment	545	651	829
Share White	0.53	0.34	0.44
Share Black	0.15	0.27	0.26
Share Hispanic	0.24	0.30	0.22
Urban	0.26	0.50	0.44
Suburban	0.29	0.21	0.24
Rural/Town	0.45	0.29	0.33
Charter school	0.13	0.12	0.10
Schools with grades above 9	0.39	0.51	0.54
Schools with grades below 6	0.37	0.26	0.19
Income-to-poverty ratio	299	280	315
Number of schools	62,447	3200	638

Notes: This table reports baseline demographic characteristics for three samples of middle and high schools. Column (1) includes all non–Yondr-adopting public schools in the NCES Common Core of Data. Column (2) includes schools that adopted Yondr lockable pouches during the study period, excluding those with lowest-confidence implementation timing in Yondr's administrative records. Column (3) includes the subset of Yondr-adopting schools that also appear in the NTS and for which the modal educator response indicates Yondr or a similar lockable pouch as the school's phone policy, based on responses to Q3 (current policy) or Q14 (prior policy), and that match to Yondr's administrative records. All statistics are computed at the school level. Income-to-poverty ratio is measured using EDGE School Neighborhood Poverty Estimates. See Appendix C.3 for the exact wording of each question.

Table A4: Gallup Survey Demographics

Category	Variable / Level	Mean (SD)
Parent	Age	46.9 (8.3)
Parent Gender	Male	596 (38%)
	Female	962 (61%)
	Non-binary	7 (0.4%)
Parent Race	White	954 (61%)
	Black	254 (16%)
	Hispanic	227 (15%)
	Asian	94 (6.0%)
	Other	31 (2.0%)
Household Income	Less than $12,000	59 (3.8%)
	$12,000–$23,999	47 (3.0%)
	$24,000–$35,999	96 (6.2%)
	$36,000–$47,999	107 (6.9%)
	$48,000–$59,999	132 (8.5%)
	$60,000–$89,999	264 (17%)
	$90,000–$119,999	228 (15%)
	$120,000–$179,999	302 (20%)
	$180,000–$239,999	153 (9.9%)
	$240,000 and over	157 (10%)
Child	Age	15.3 (1.8)
Child Gender	Male	815 (52%)
	Female	725 (46%)
	Nonbinary	25 (1.6%)
Child Race	White	824 (53%)
	Black	284 (18%)
	Hispanic	319 (20%)
	Asian	116 (7.4%)
	Other	22 (1.4%)
Child Grade	5th Grade	1 (<0.1%)
	6th Grade	34 (2.2%)
	7th Grade	163 (10%)
	8th Grade	256 (16%)
	9th Grade	230 (15%)
	10th Grade	293 (19%)
	11th Grade	262 (17%)
	12th Grade	246 (16%)
	Not in Middle or High School	80 (5.1%)
School Type	Public school	1,125 (75%)
	Charter school	81 (5.4%)
	Private school	137 (9.2%)
	Homeschool	148 (9.9%)
Parent	Life Satisfaction (0–10)	6.8 (1.8)
Child	Life Satisfaction (0–10)	7.2 (1.7)

Notes: This table reports demographic characteristics of the parent and child dyads interviewed in the Gallup survey. These questions were only asked to the respondents who were already part of the Gallup Panel in previous survey waves; N = 1,565.

Table A5: First Stage ATT Estimates: Advan GPS Data

	Net visits 10h (1)	Net visits 09-13h (2)	Net visits 08-15h (3)
ATT	-0.195^{***}	-0.193^{***}	-0.148^{***}
	(0.052)	(0.052)	(0.046)
Observations	90,130	92,792	94,099
Schools	19523	19537	19555

Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption on smartphone pings using the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator, using data from Advan. The outcome is the log of net visits, defined as average weekday school-hours visits for each time period minus the corresponding weekend activity in the same month, further subtracting the weekday–weekend difference in nighttime activity (00:00–06:00). All specifications use an imbalanced panel and include state-by-urbanicity fixed effects and baseline covariates, following the main specification. Standard errors are clustered at the school level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A6: Robustness Checks: Difference-in-Differences Estimates of Effects of Phone Pouches on Middle School Test Scores

Specification	ATT on test scores	Observations
Baseline	−0.025**	166,471
	(0.012)	
TWFE	−0.024**	157,671
	(0.011)	
Regression adjustment only	−0.026**	166,471
	(0.012)	
Exclude strict policy change from control	−0.018	29,236
	(0.017)	
Exclude any policy change from control	−0.014	16,246
	(0.020)	
Not yet treated as controls	−0.013	9566
	(0.016)	
Exclude 2026 adopters	−0.027**	163,815
	(0.012)	
High confidence schools only	−0.025**	165,208
	(0.012)	
Exclude canceled adopters	−0.026**	165,626
	(0.013)	
Balanced panel	−0.021	91,788
	(0.014)	
Controlling for number of students tested	−0.025**	166,471
	(0.012)	
No imputed test scores	−0.022*	160,047
	(0.012)	

Notes: The table reports estimates of the average treatment effect on the treated (ATT) of Yondr adoption on standardized test scores. The baseline specification (row 1) uses the Callaway-Sant'Anna doubly robust difference-in-differences estimator with an imbalanced panel and never-treated schools as controls. Rows 2 and 3 report alternative estimators: a two-way fixed effects specification and the Callaway–Sant'Anna estimator with regression adjustment only. Rows 4–7 modify the control group. Row 4 restricts the control group to schools reporting no strict phone policy changes in the past five years; row 5 restricts to schools reporting no phone policy changes of any kind; row 6 uses only not-yet-treated schools as controls (i.e., the 2026 adopters); and row 7 excludes 2026 Yondr adopters from the control group. Rows 8 and 9 modify the treated group by restricting to high-confidence implementation dates and excluding schools with canceled accounts. Rows 10–12 address additional methodological considerations by restricting to a balanced panel, controlling for the number of students tested, and excluding imputed test scores. Standard errors are clustered at the school level. See Appendix C.3 for NTS question wording. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A7: Difference-in-Differences Estimates with Incremental Baseline Controls

	Minimum controls (1)	Demographic controls (2)	Grade level controls (3)	Charter status (4)
ATT	−0.014	−0.026**	−0.024**	−0.025**
	(0.012)	(0.012)	(0.012)	(0.012)
Observations	166,471	166,471	166,471	166,471

Notes: The table reports estimates of the average treatment effect on the treated (ATT) of Yondr adoption on standardized test scores using the Callaway and Sant'Anna (2021) doubly robust difference-in-differences estimator. Columns sequentially expand the set of baseline covariates included in the specification. Column (1) includes only state-by-urbanicity fixed effects. Column (2) adds baseline demographic characteristics, including total enrollment, racial composition, and neighborhood poverty quartiles. Column (3) further adds grade span indicators, and Column (4) adds an indicator for charter school status. All covariates are measured prior to treatment and enter interacted with year fixed effects. Standard errors are clustered at the school level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A8: Tests of Compositional Effects

	Total enrollment (1)	Share White (2)	Share Black (3)	Share Hispanic (4)	Share FRPL (5)
ATT	−4.519	0.003**	−0.001	0.007**	−0.007**
	(3.685)	(0.001)	(0.005)	(0.004)	(0.003)
Mean of dependent variable	540	0.534	0.163	0.274	0.563
Observations	166,460	162,915	156,213	162,828	147,791

Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption on school-level enrollment or student composition using the Callaway and Sant'Anna (2021) difference-in-differences estimator. Outcomes include total enrollment, racial/ethnic enrollment shares, and the share of students eligible for free or reduced-price lunch (FRPL). Differences in sample sizes reflect missing data in the NCES files. Standard errors are clustered at the school level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A9: Heterogeneous Effects of Phone Pouches on Middle School Test Scores

	High poverty (1)	Low poverty (2)	Male (3)	Female (4)	Black (5)	White (6)	Hispanic (7)	ED (8)	Not ED (9)
ATT	−0.016	−0.046**	−0.022	−0.020	−0.007	−0.043*	−0.031*	−0.013	−0.041**
	(0.014)	(0.022)	(0.014)	(0.014)	(0.023)	(0.025)	(0.018)	(0.016)	(0.021)
Observations	84,790	81,681	125,248	124,386	50,514	106,990	83,916	132,441	114,099

Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption on standardized test scores for the indicated subgroup. Columns (1) and (2) report effects at the *school level*, comparing schools with high versus low poverty, where high- and low-poverty schools are defined as those above and below the within-state median of the school neighborhood income-to-poverty ratio. Columns (3) through (9) report effects for *within-school student subgroups*, such as test scores for male versus female students or for students of different racial and ethnic groups; these outcomes are defined at the student level and aggregated to the school–year level within subgroup. ED denotes economically disadvantaged students. The number of observations varies across columns due to differences in data availability for subgroup-specific test scores, as some school–years do not report outcomes for small demographic cells. Estimates are obtained using the improved doubly robust Callaway and Sant'Anna (2021) difference-in-differences estimator and include the same baseline covariates as in the main specification. Standard errors are clustered at the school level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A10: Difference-in-Differences Estimates of Effects of Phone Pouches on Alternative Attendance Outcomes

	Chronic absenteeism rate (1)	Attendance rate CA (2)	Attendance rate net of discipline CA (3)
ATT	0.156	0.065	0.073
	(0.241)	(0.134)	(0.134)
Mean of dependent variable	24.52	92.90	92.99
Observations	229,118	39,837	39,837

Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption using the Callaway and Sant'Anna (2021) doubly robust difference-in-differences estimator. Column (1) reports effects on the chronic absenteeism rate, measured as the percentage of students classified as chronically absent. Column (2) reports effects on the overall attendance rate in California, measured in percentage points. Column (3) reports effects on the California attendance rate excluding absences due to disciplinary suspensions. All specifications use an imbalanced panel and include state-by-urbanicity fixed effects and baseline covariates. Standard errors are clustered at the school level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A11: Difference-in-Differences Estimates of Effects of Phone Pouches on Middle School Attendance, Discipline, and Student Survey Outcomes

	Attendance rate (1)	Discipline index (2)	SWB index (3)	Attention index (4)	Online bullying (5)
ATT	0.080	0.051	−0.165	−0.014	0.036
	(0.127)	(0.033)	(0.319)	(0.078)	(0.130)
Mean of dependent variable	93.19	–	–	–	–
Observations	113,662	95,328	2261	6174	3335

Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption using the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator. This table replicates Table 4 but restricts the sample to middle schools only. For Columns (1) and (2), middle schools are defined as schools with no enrollment in grades 9–12 according to the CCD. For Columns (3), (4), and (5) middle schools are defined as schools whose highest grade offered is below grade 9. Column (1) reports effects on the attendance rate (percentage points). Column (2) reports effects on the standardized discipline index. Columns (3) and (4) report effects on the subjective well-being (SWB) and classroom attention indices, measured in school-level standard deviations. Column (5) reports effects on online bullying constructed from a Panorama survey question, also measured in school-level standard deviations. All specifications use an imbalanced panel and include state-by-urbanicity fixed effects and baseline covariates. Standard errors are clustered at the school level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A12: Distribution of Schools by Yondr Adoption Confidence Tier

	Number of Schools (1)	Percent (2)
Highest	2262	49.10
High	1375	29.85
Medium	578	12.55
Low	41	0.89
Lowest	351	7.62
Total	4607	100.00

Notes: This table reports the distribution of schools across confidence tiers reflecting the strength of evidence on Yondr usage and the timing of initial adoption. Confidence tiers are based on the availability and consistency of internally reported implementation years, shipment records, and launch dates. Schools in the lowest confidence tier are excluded from the main analysis; robustness checks restrict the sample to the highest-confidence schools.

Table A13: First Stage ATT Estimates: Advan GPS Data by Sample

Panel A: Test Score Sample			
	Net visits 10h (1)	Net visits 09-13h (2)	Net visits 08-15h (3)
ATT	-0.146^{*}	-0.162^{**}	-0.152^{**}
	(0.079)	(0.075)	(0.067)
Observations	55,460	56,962	57,708
Schools	14124	14134	14161
Panel B: Attendance Sample			
	Net visits 10h (1)	Net visits 09-13h (2)	Net visits 08-15h (3)
ATT	-0.199^{***}	-0.183^{***}	-0.146^{***}
	(0.062)	(0.057)	(0.053)
Observations	55,193	56,833	57,577
Schools	12611	12616	12623
Panel C: Discipline Sample			
	Net visits 10h (1)	Net visits 09-13h (2)	Net visits 08-15h (3)
ATT	-0.193^{***}	-0.171^{**}	-0.123^{*}
	(0.073)	(0.068)	(0.065)
Observations	40,645	41,849	42,387
Schools	10441	10461	10470

Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption on smartphone pings using the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator, using data from Advan. The outcome is the log of net visits, defined as average weekday school-hours visits for each time period minus the corresponding weekend activity in the same month, further subtracting the weekday–weekend difference in nighttime activity (00:00–06:00). All specifications use an imbalanced panel and include state-by-urbanicity fixed effects and baseline covariates, following the main specification. Standard errors are clustered at the school level. Each panel subsets the sample to schools used in the main outcome analyses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A14: First Stage ATT Estimates: Quadrant GPS Data

	Net pings 10h (1)	Net pings 09-13h (2)	Net pings 08-15h (3)
ATT	-0.105^{***}	-0.097^{***}	-0.093^{***}
	(0.029)	(0.029)	(0.027)
Observations	197,723	207,731	208,623
Schools	57003	57877	58009

Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption on smartphone pings using the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator, using data from Quadrant. The outcome is the log of net pings, defined as average weekday school-hours pings (computed as the total ping count over the interval) minus the corresponding weekend activity in the same month, further subtracting the weekday–weekend difference in nighttime activity (00:00–06:00). All specifications use an imbalanced panel and include state-by-urbanicity fixed effects and baseline covariates, following the main specification. Standard errors are clustered at the school level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A15: Construction of Analysis Samples Across Outcomes

	Test scores (1)	Attendance (2)	Chronic abs. (3)	Discipline (4)
NCES Panel	806,853	806,853	806,853	806,853
Initial outcome panel	186,135	450,740	513,790	362,243
Restrict to middle/high school grades	181,360	276,685	310,577	232,047
Drop 2020 and 2021	181,360	215,248	242,548	176,828
Drop pre-2023 adopters	179,729	212,511	239,651	174,490
Drop contaminated controls	175,650	207,786	234,390	169,749
Drop low-confidence schools	174,835	206,642	233,153	168,865
Keep states with 2023–2025 adoption	173,728	206,642	233,153	168,865

Notes: The table reports the number of school–year observations remaining after sequentially applying each sample restriction for each outcome. The initial NCES panel includes all school–year observations in the Common Core of Data from 2017–18 to 2024–25. Academic years 2019–20 and 2020–21 are excluded from the attendance, chronic absenteeism, and discipline analyses to align the analysis period with the test score panel and to omit pandemic-disrupted years. The test score sample includes only grades 6 through 8; the other outcome panels also include high-school data.

Table A16: Construction of Panorama Analysis Samples

	Subjective well-being (1)	Classroom attention (2)	Online bullying (3)
NCES Panel (2019–2025)	706,168	706,168	706,168
Drop 2020 and 2021	505,092	505,092	505,092
Restrict to middle/high school grades	279,540	279,540	279,540
Require any Panorama index	20,302	20,302	20,302
Require outcome-specific index	5702	14,074	9070
Drop pre-2023 adopters	5618	13,892	8951
Drop contaminated controls	5519	13,414	8655
Drop low-confidence schools	5497	13,381	8628
Keep states with 2023–2025 adoption	3946	10,579	6041

Notes: The table reports the number of school–year observations remaining after sequentially applying each sample restriction for the Panorama-based outcomes. The initial panel includes all school–year observations sent to Panorama for matching between 2018-19 and 2024-25. Academic years 2019-20 and 2020-21 are excluded to align the Panorama samples with the main analysis period used elsewhere in the paper. Schools are restricted to those serving at least one middle or high school grade. Differences in final sample size across outcomes reflect differences in survey module and item availability across schools and years.

C Data Appendix

C.1 Additional Details on Yondr Data

This section describes the construction of the school-level Yondr adoption panel used in the analysis. The goal of this process was to (i) identify the universe of U.S. public schools that use Yondr lockable phone pouches and (ii) assign the first academic year of implementation with an explicit measure of confidence. We worked closely with Yondr staff throughout this process to understand the structure, coverage, and limitations of these data and to validate key variables.

Data Sources. We combine multiple internal administrative datasets provided by Yondr with external data from the NCES. The main Yondr data include: (i) the *School and District report*, which contains school-level account records with school names, locations, and internally reported implementation information, as well as district-level account records; and (ii) historical shipment records from two sales platforms—Airtable (Yondr's legacy sales platform) and Salesforce (Yondr's current sales platform). The school- and district-level account records represent Yondr's internal list of schools and districts believed to have used Yondr pouches. These datasets include variables reporting Yondr's internal administrative estimates of the academic years of implementation ("Years Implemented") and pouch launch dates for each of their account holders. Shipment records reflect actual shipments of Yondr pouches, including both initial and supplemental orders.

Multiple Versions of the School and District Report. Over the course of the project, Yondr provided three distinct versions of the School and District report. Earlier versions reflected interim snapshots of Yondr's internal records and differed in coverage due to ongoing account creation, data cleaning, and filtering decisions. After follow-up discussions, we requested an unfiltered extract of all school accounts in Yondr's records. The third and most recent version—used as the baseline for our analysis—was described by Yondr as their most complete and accurate record and includes a manually reviewed "Years Implemented" field. Schools that appeared in earlier versions of the report but are missing from the final version are referred to as "disappeared" schools. These cases reflect changes across file versions rather than confirmed discontinuation of Yondr usage and are treated separately in the confidence-tier classification described below.

Matching Yondr Accounts to NCES IDs. All datasets are merged to standardized NCES school and district identifiers to obtain consistent identifiers and baseline characteristics. Although Yondr provided NCES IDs for most schools, a subset of records contained missing or incorrect identifiers. We therefore supplemented and verified all NCES IDs using a combination of automated matching and extensive manual verification based on school

names, addresses, and district information. Schools identified as non-public or otherwise ineligible for inclusion (e.g., private schools, special programs, or non-U.S. entities) were excluded. After this process, all schools retained in the dataset have verified NCES school identifiers.

Shipment Records and District-Level Ambiguity. Beyond the School and District report, shipment records provide an additional signal of Yondr usage and timing. We harmonize shipment dates across Yondr's Airtable and Salesforce platforms and record the earliest observed shipment associated with each school account.

Most shipments can be directly attributed to individual schools. However, some shipments occur at the district level and cannot be reliably assigned to specific schools. When a district receives a shipment but the recipient schools cannot be uniquely identified, we flag the district and exclude it from the control group in the analysis. This avoids contamination of untreated schools that may nonetheless have been exposed to Yondr pouches.

In some cases, shipment records identify valid public schools that do not appear in the baseline school account records. When such shipments can be clearly attributed to individual schools, these schools are added to the School and District report and marked as recovered from shipping records.

Defining Adoption Timing. The treatment timing variable is defined as the first academic year in which a school is inferred to use Yondr pouches. Calendar dates are mapped to academic years using a consistent rule: shipments or launches occurring during the school year are assigned to that academic year, while late-spring and summer dates are assigned to the subsequent academic year.

Across data sources, three potential indicators of adoption timing are available: (i) the internally reported "Years Implemented" variable, (ii) shipment dates from the Airtable and Salesforce records, and (iii) administrative launch dates. Based on discussions with Yondr staff and internal validation, Yondr reports having the greatest confidence in the "Years Implemented" variable. Accordingly, when multiple timing indicators are available, we prioritize the internally reported "Years Implemented," followed by shipment dates, and finally administrative launch dates.

Confidence Tiers. To transparently reflect uncertainty in both Yondr usage and adoption timing, we classify each school into one of five confidence tiers based on the availability, consistency, and source of timing information:

- **Highest confidence:** The school record appears in the final School and District report, has a non-missing "Years Implemented" entry, and has at least one associated shipment record. The internally reported implementation year and the shipment date

correspond to the same academic year. In these cases, the "Years Implemented" variable is used to assign the start year.

- **High confidence:** The school record appears in the final School and District report and has a non-missing "Years Implemented" entry but no associated shipment record. In these cases, the "Years Implemented" variable is used directly to assign the start year. This tier remains high confidence because, based on discussions with Yondr staff, the "Years Implemented" variable was recently reviewed and manually validated, whereas shipment records—particularly those from Airtable—are believed to be less complete.

- **Medium confidence:** The school record appears in the final School and District report, has at least one shipment record, but lacks "Years Implemented" information. The start year is inferred from the first observed shipment date.

- **Low confidence:** The school record appears in the final School and District report, lacks both "Years Implemented" information and shipment records, but has a recorded launch date. In these cases, the launch date is used to assign the start year.

- **Lowest confidence:** This category includes all remaining records. Specifically, it includes schools appearing in the final School and District report with no available timing information of any kind; schools that appeared in earlier versions of Yondr's records but are missing from the final report ("disappeared" schools), regardless of whether shipment or launch date information is available; and schools recovered solely from shipment records that do not appear in the School and District report. Though these schools may have adoption timing data available, we did not feel comfortable assigning them to a higher confidence tier, given that Yondr never created a matching account for them in the School and District Report or removed them from the account data from one version of the file to the next.

Table A12 summarizes the confidence-tier classification for all schools appearing in the Yondr administrative data.

Final Analytic Sample. Our main analysis excludes schools in the lowest confidence tier, for which adoption timing is too uncertain to support credible difference-in-differences estimation. We show in robustness checks that results are similar when further restricting the sample to only the high- and highest-confidence schools.

C.2 Additional Details on GPS Data

Overview. Both Advan and Quadrant rely on opt-in location sharing through a panel of mobile applications. Neither dataset should be interpreted as observing every phone in the population; instead, each provides a convenience sample of devices whose owners have granted location access to at least one app or website in the provider's pool of partner developers. The set of observable devices can change over time due to app churn, operating-system updates, and permission changes, so location-level aggregates can mechanically reflect both true behavioral changes and changes in panel composition.

Quadrant's panel is large (the provider reports coverage on the order of hundreds of millions of devices), but most devices are observed only intermittently, often just a few times per month. Advan's panel is smaller, but it emphasizes higher-frequency devices and provides pre-processed, modeled measures intended to approximate true visits. Several features of mobile location data generate noise and sparsity. First, panel composition changes over time as users install or delete apps and change location settings. Second, event frequency depends on app behavior and platform restrictions: some apps generate many events while others generate few, and these differences can shift over time. Third, geolocation error and polygon boundaries matter, because GPS and assisted-location error can assign events to the wrong side of a boundary when creating a geofence (i.e. delimiting a location to identify devices within it), especially near tight building perimeters. Fourth, providers may apply privacy protections including suppression or other forms of processing that are more salient when counts are small. These issues motivate our emphasis on within-school changes, the use of net measures (weekday minus weekend within the same clock hours), and validation exercises.[11]

What a "ping" measures. Location events are recorded when apps with location permission register and transmit a device's position. These events do not map one-for-one to screen time: some background collection is possible (for instance through navigation or weather services, or when users grant "always" location access), so a phone that is physically present but not actively in use may still generate pings. At the same time, recent privacy settings have shifted most collection toward foreground use, implying that the majority of observed events occur while an app is actively open. Even in those cases, transmission may be delayed if observations are cached and uploaded during a subsequent app session, and certain app categories can generate events more frequently than others.

[11] For documentation, see https://docs.quadrant.io/about-quadrant-mobile-location-data, https://docs.deweydata.io/docs/faqs-advan-research, and https://advanresearch.com/downloads/Advan-Weekly-Patterns%2B.pdf. For previous work that has used these data sources, see Cook et al. (2025); Zheng et al. (2024); Goldhaber et al. (2024); Higa (2024); Figlio and Ozek (2025).

Accordingly, we interpret declines in school-hours event intensity as evidence of reduced location-generating phone activity during class time, consistent with reduced active phone use, while acknowledging that the mapping from pings to screen time is imperfect and may also partly capture the mere presence of devices (in which case a reduction should be interpreted as devices not being brought to school).

Who is being observed. We do not observe user age in either dataset, so we cannot directly restrict to minors. Coverage of school-age minors may be limited and may vary across contexts because of parental consent requirements and OS-level privacy protections, and because app and provider systems do not consistently capture age in a way that would support clean filtering. At the same time, it is not plausible to assume the data exclude minors entirely; the datasets reflect opt-in app users and apps often do not record user age.

We also cannot infer that observed devices are students. Devices could belong to staff, teachers, parents, visitors, passersby, or could reflect residual noise around the geofence boundary. This motivates focusing on within-school changes and on measures that difference out baseline activity that is less likely to reflect student presence (e.g., netting weekday school-hours activity by weekend activity in the same clock-hour window). The validation exercises reported below suggest that the school-hours measures contain meaningful signal related to on-campus activity during the instructional day. We additionally find that visits per enrolled student are often high and exhibit a ratio close to one across related measures, which is difficult to reconcile with an interpretation driven only by staff and teachers.

Advan data processing. Our main analysis uses Advan's Weekly Patterns+ product, which reports modeled device visit counts aggregated by school, date, and hour for a national panel of around 40,542 schools (January 2019 to January 2026). Advan defines a "visit" as a device being observed within the geofence boundary at any point during a given time interval, without imposing a minimum dwell-time requirement. Advan reports visits as "estimated" (scaled) counts. The scaled visit measure is constructed by scaling the number of devices observed in the geofence by the estimated size of the provider's device panel and a target population. We interpret these as model-based estimates of the number of actively used devices rather than literal counts of physical entries. Our main outcomes use the hourly series to construct measures of in-school daytime location activity during instructional hours (e.g., weekday 08:00–15:00), normalized by school enrollment. We define "net visits" as a double difference: average weekday school-hours visits minus the corresponding weekend activity within the same month, further subtracting the weekday–weekend difference in nighttime activity (00:00–06:00). This construction removes baseline non-school traffic and residual background noise around school locations, yielding a school-level measure of relative phone

activity during the school day. We then divide this number by enrollment to obtain a per-student visit rate. Finally, we build an annual school panel by averaging over the months of September, October, November, February, March, April, and May for each academic year.

Quadrant data processing. Quadrant delivers raw location events as device-level latitude/longitude coordinates with timestamps and metadata, derived from its app and partner supply chain. The data cover around 2 trillion pings observed in February and October of 2022, 2023, and 2024, as well as February 2025, for the entire United States. We map events to school polygons and construct school-by-time aggregates by counting the total number of in-polygon events within specific clock-hour windows. In contrast to Advan, our primary Quadrant measure is the total number of pings (events) in the window rather than the modeled visit estimate, which is a unique-device presence series.

This distinction matters for interpretation. Advan's hourly outputs are closer to an "active device presence" concept (devices observed at least once in the hour, with further scaling and modeling), while Quadrant ping totals are more directly tied to the event generation process itself, which can vary across apps, devices, and user settings and are therefore noisier and more sparse. On a typical day, only a small fraction of students' devices generate observable pings, and the set of observed devices changes over time as user behavior and app privacy settings change. We therefore use Quadrant to validate the direction and timing of the effects.

To process Quadrant's data, we begin by constructing school geofences. We start with school point locations from NCES and assign each school a building footprint polygon using the FEMA U.S. Structures dataset, selecting a polygon that contains the NCES point when available and otherwise selecting the nearest plausible structure within a tight radius of 200m. Second, we match pings to school polygons. For each event, we treat the recorded latitude/longitude as a point and assign it to a school if the point falls inside that school's polygon (a point-in-polygon operation). Third, we aggregate to school-by-day (and higher) outcomes. We count total in-polygon events within local-time windows that correspond to instructional hours (e.g., 08:00–15:00) and construct net measures by differencing analogous weekend windows, using the same windows as we use for the Advan data. We again divide by school enrollment to obtain per-student indicators. We then build school-year panels by taking the mean between the February and November averages for each academic year.

Validation exercises. To assess whether the constructed measures contain meaningful signal, we present three sets of validation exercises. A first validation check is whether observed activity is concentrated at times when students are plausibly present and phones are

plausibly in use. Figure A7 summarizes aggregate Advan activity over time and illustrates that measured activity is substantially higher during school hours than at night, and higher on weekdays than weekends, consistent with the interpretation that the measure is capturing school-related device presence and activity rather than purely mechanical background location signals.

A second validation check, shown in Figure A8, is whether activity scales with school size and differs across levels of schooling. Phone activity is higher in months where schools are in session and lower in summer months, increases with enrollment, yielding values of visits per student that are close to or above 1, per-student activity is systematically higher in high schools than in middle schools, and both are much higher than in elementary schools. These patterns are consistent with basic behavioral predictions: older students are more likely to own smartphones, to carry them throughout the school day, and to generate usage-related location events.

A final validation check links GPS outcomes to independent, survey-based measures of phone use and policy strictness from the NTS. Figure A9 shows that GPS activity is systematically lower in schools with more restrictive phone policies (e.g., bell-to-bell) and higher in schools with weaker policies. We then relate GPS-based measures of net smartphone activity to teacher-reported in-class phone misuse from the survey data. The correlations between net visits as measured in the Advan data and in-class usage as reported in NTS are positive and statistically significant for different hour-window definitions, indicating that schools with higher survey-reported classroom phone use also exhibit larger GPS-measured changes in in-school smartphone activity. For the 8:00–15:00 measure, the correlation is 0.157 ($p=< 0.001$); for the 9:00–13:00 measure, the correlation is 0.172 ($p=< 0.001$); and for the 10:00 single-hour measure, the correlation is 0.165 ($p=< 0.001$). These correlations provide external validity for the GPS-based measures by demonstrating that they covary with independently measured constructs in expected directions.

Additional first stage results. Finally, we present additional evidence on the effect of Yondr adoption on GPS-measured smartphone activity. The main text focuses on the Advan outcome defined using the 10:00–11:00 school-hours window and the netting procedure described above. The results in Tables A5, A13, and A14, as well as Figure A10, show that the first-stage findings are not sensitive to reasonable alternative choices. Across definitions, including other time-windows for school-hours, other samples of schools that match our main analyses (for test scores, attendance, and discipline outcomes), and data providers (replicating the analysis with Quadrant data), we find the same qualitative pattern of a sustained decline in smartphone activity after adoption.[12]

[12]We cannot conduct an analogous robustness exercise for the Panorama sample because, as described

C.3 Wording of NTS Questions Used in the Analysis

This section reproduces the wording of the key survey questions used in the analysis. Question numbering reflects the original survey instrument; only the questions used in this paper are included below.

Current Phone Policy

Q2. Officially, when is the personal use of phones (e.g., for texting, social media) restricted?

- Away for the day (i.e., students are prohibited from using their phones during the entire school day)
- Schedule-based restriction (e.g., students can only use phones during lunch or between classes)
- No school-wide restriction (e.g., Individual teachers and staff decide when students can use phones for their class)

Q3. Officially, where are students allowed to keep their phones?

- No phones in building or on campus (students must leave their phones at home)
- Centralized collection (students must deposit their phones in one place, like the main office, at the beginning of the school day)
- Yondr pouches or similar (students must keep their phones in a lockable pouch)
- Classroom collection by all or most teachers (students must deposit their phones at the beginning of every class)
- Lockers only (students must keep their phones in their lockers all day)
- "No show" (students can keep their phones in backpacks, pockets, etc.—but phones must be out of sight)
- No school-wide policy (Individual teachers and staff decide where students can keep phones)

Phone Use Measures

Q5. Between classes, about how many students in your school are using their phones? Give your best guess.

in Section 3, Panorama matched school identifiers to survey responses prior to anonymization, preventing linkage to additional school-level samples.

Q6. During class, about how many students are using their phones for personal reasons (for example, texting, social media)? Give your best guess.

For Q5 and Q6, response options ranged from 0% (none of the students) to 100% (all of the students) in 10 percentage-point increments.

Policy Evaluation

Q8. Overall, how satisfied are you with your school's official cell phone policy?

Response options ranged from 0% (not at all satisfied) to 100% (completely satisfied) in 10 percentage-point increments.

Q9. In your opinion, should your school's cell phone policy be more or less restrictive than it is now?

- The policy should be much more restrictive
- The policy should be a little more restrictive
- The policy is just right
- The policy should be a little less restrictive
- The policy should be much less restrictive

Policy Change and Prior Policy

Q10. Have there been any major changes in your school's official cell phone policy in the last 5 years?

- Yes, the policy has changed
- No, the policy has not changed
- Honestly, I recently joined this school and don't know

(If yes:)

Q11. Approximately when did your school's cell phone policy change?

- Month (Dropdown list: January – December)
- Year (Dropdown list: 2021 – 2025)

Q13. Before the change, when was the personal use of phones (e.g., for texting, social media) restricted?

Q14. Before the change, where were students allowed to keep their phones?

Q13 and Q14 used the same response options as Q2 and Q3, respectively.

Q16. Before the change, about how many students were using their phones between classes? Give your best guess.

Q17. Before the change, about how many students used their phone during class for personal reasons (for example, texting, social media)? Give your best guess.

Q19. Before the change, how satisfied were you with your school's official cell phone policy?

Q16, Q17, and Q19 used the same percentage-scale response options as Q5, Q6, and Q8.

Q20. Other than the policy for cell phones, were there any other major changes at your school (e.g., change in leadership, new curriculum) that happened around the time of the phone policy change that might have changed how students behaved, performed in class, etc.?

- Not that I can think of
- Yes: _____ (open text response required)

C.4 Wording of Gallup Questions Used in the Analysis and Weighting Scheme

This section reproduces the wording of the key Gallup questions used in the analysis. Question numbering reflects the original survey instrument; only the questions used in this paper are included below.

Hypothetical Phone Ban Scenario

W3PTEXT1. Imagine that a school changes its cell phone policy. Before, it had no school-wide restrictions. After, it bans cell phones by making students keep their phones in a lockable pouch that prevents cell phones from being used during the school day.

W3P3. How would this phone ban affect standardized test scores?

- Scores would go down
- Scores wouldn't change
- Scores would go up

W3P4. How would this phone ban affect how well students get along?

- Students would get along worse

- Students would get along about the same
- Students would get along better

W3P5. How would this phone ban affect student mental health?

- Mental health would worsen
- Mental health wouldn't change
- Mental health would improve

W3P6. Would you support or oppose schools banning cell phones and making students keep them in a lockable pouch?

- I oppose schools banning cell phone use by making students lock them in a pouch
- I'm neutral
- I support schools banning cell phone use by making students lock them in a pouch

Sample weighting. The Gallup data come from a probability-based sample drawn from the Gallup Panel and fielded from February 10 to February 17, 2026. The cross-sectional data from this survey wave include 4,449 completed interviews, comprising 1,801 parent-child dyads and 2,648 adult Gen Z respondents. Gallup provides separate survey weights for youth/young adult responses and for parent responses. These weights begin with panel base weights that incorporate respondents' probabilities of selection into both the Gallup Panel and this specific survey wave; for children in households with multiple eligible children, the base weight also adjusts for within-household child selection. Gallup then applies post-stratification using a raking procedure to align the sample with U.S. population benchmarks from the American Community Survey on age, gender, education, ethnicity, race, and region. Accordingly, all estimates used in this paper use the appropriate cross-sectional weight for the respondent type being analyzed, so that results are nationally representative of U.S. youth and young adults ages 12–29 or of parents of youth ages 12–18.

C.5 Construction of Test Score, Attendance, and Discipline Samples

This section documents the construction of the test score, attendance, and discipline analytic samples. All analyses are conducted at the school–year level and use the largest available set of observations for which the relevant outcome measures can be consistently constructed.

Table A15 summarizes the construction of each analysis sample and reports how each sample restriction affects the number of school–year observations.

Standardized Test Score Analysis Sample

The initial test score panel combines data from the EDC and SEDA and includes 186,135 school–year observations with non-missing English Language Arts or mathematics test scores averaged across grades 6 through 8. The sample covers 40,350 unique public schools in 49 states and the District of Columbia over the 2018, 2019, and 2022 through 2025 academic years.

We apply a sequence of common sample restrictions designed to focus on a well-defined treatment period and to ensure credible identification. Specifically, we exclude schools that adopted Yondr prior to the 2023 academic year, schools located in districts identified as potentially contaminated control districts—districts that may have adopted Yondr but for which school-level adoption timing cannot be reliably determined—and Yondr-adopting schools for which implementation timing is classified as lowest confidence in the Yondr administrative data. To standardize the sample definition across outcomes, we include only schools that had middle and/or high school grades (6–8 and 9–12, respectively) in the first year they appear in the panel. We further restrict the sample to states with at least one Yondr adoption during the 2023–2025 adoption window, as states without adoption during this period provide no identifying variation.

After applying these restrictions, the main test score analysis sample consists of 173,728 school–year observations, representing 37,051 unique schools across 47 states and the District of Columbia.[13] Of these schools, 1,385 adopt Yondr during the 2023–2025 adoption window, with the remainder serving as comparison schools.

The baseline test score specifications use an unbalanced panel: some schools are observed in all outcome periods used in the analysis, while others are observed for a subset of years. This reflects differences in data availability across assessment sources as well as genuine entry and exit of schools over time. In particular, EDC state assessment data extend through the 2025 school year for most states, while SEDA-based achievement measures are available only through the 2024 school year. Table A2 lists the data source used for each state.

As a robustness check, we also report estimates from a fully balanced panel of schools observed in all six outcome periods. This balanced-panel sample includes 15,298 schools, representing approximately 41 percent of the baseline test score sample. Estimated effects in the balanced-panel specifications are similar in magnitude and direction to the baseline results.

[13]The panel does not include data for Alaska, Montana, and South Dakota.

School-Level Attendance Analysis Sample

The initial attendance panel includes 450,740 school–year observations, covering 70,853 unique public schools. The initial attendance panel includes all school–year observations from 2018 through 2025 in states with substantial Yondr adoption for which a conventional school-level attendance rate is reported by the state education agency. We begin with states ranked highest by the number of Yondr-adopting schools and work downward to include as many states as possible with comparable attendance rate definitions. Some states are excluded at this stage because they do not report traditional attendance rates at the school level (e.g., reporting only chronic absenteeism).

We first restrict the sample to schools serving at least one middle or high school grade, where smartphone use is more prevalent and where Yondr adoption is concentrated. We then apply outcome-availability and treatment-related restrictions analogous to those used in the test score analyses, including requiring a non-missing attendance rate measure, excluding pre-2023 adopters, excluding potentially contaminated control districts, and dropping low-confidence Yondr adopters. We also restrict the sample to states with at least one Yondr adoption during the 2023–2025 adoption window. To align the attendance analysis period with the test score sample and to exclude pandemic-disrupted years, we drop academic years 2020 and 2021. After applying these restrictions, the resulting final attendance analysis sample consists of 206,642 school–year observations, representing 41,834 unique schools.[14]

As in the test score analyses, the baseline attendance specifications use an unbalanced panel, reflecting differences in attendance reporting practices across states and years as well as genuine entry and exit of schools over time.

School-Level Disciplinary Outcomes Analysis Sample

The initial disciplinary panel includes 362,243 school–year observations, covering 60,372 unique public schools across 22 states with at least one school-level disciplinary measure reported during the academic year.[15]

States differ substantially in the disciplinary outcomes they report. Some states provide a broad measure of the share of students disciplined during the academic year, which may include suspensions, expulsions, and other disciplinary actions. Other states report only suspension-based measures, such as the share of students receiving any suspension or the shares receiving in-school or out-of-school suspension. To accommodate this heterogeneity,

[14]The states included in the final attendance analysis sample are listed in Table A2. Coverage varies by state and year.

[15]The disciplinary dataset is assembled from state education agency administrative records in states with substantial Yondr adoption for which school-level disciplinary measures can be consistently constructed.

we construct a unified disciplinary intensity measure using the most comprehensive disciplinary measure available in each state–year observation.

Specifically, we prioritize the following measures: (i) the percentage of students disciplined during the academic year; (ii) the percentage of students receiving any suspension (in-school or out-of-school); (iii) the combined percentage receiving in-school or out-of-school suspension, calculated as the sum of the in-school and out-of-school suspension rates, recognizing that this may "double-count" some students; (iv) the percentage receiving in-school suspension only; and (v) the percentage receiving out-of-school suspension only. We use the first available measure in this order. In the final analytical sample, approximately 26 percent of observations use measure (i), 66 percent use measures (ii) or (iii), and 7 percent use measures (iv) or (v). We standardize the resulting disciplinary rate within state–year cells so that differences in reporting definitions across states do not mechanically drive the estimates.

We apply the same grade-level and treatment-related sample restrictions used in the test score and attendance analyses, restricting attention to schools serving at least one middle or high school grade, excluding pre-2023 Yondr adopters, potentially contaminated control districts, and low-confidence Yondr adopters, and retaining only states with Yondr adoption during the 2023–2025 window. To align the discipline analysis period with the test score and attendance samples and to exclude pandemic-disrupted years, we drop academic years 2020 and 2021.

After applying all restrictions and excluding observations with missing standardized discipline measures, the final disciplinary analysis sample consists of 168,865 school–year observations, representing 35,605 unique schools across 22 states.[16] As in the other analyses, the baseline disciplinary specifications use an unbalanced panel, reflecting differences in data availability across states and years as well as genuine entry and exit of schools over time.

C.6 Additional Details on Panorama Data and Analysis Sample

We use student survey data provided by Panorama Education, an education technology company that develops and administers surveys for K–12 schools. Panorama surveys are widely used by school districts to measure aspects of student experience—such as engagement, classroom climate, and socio-emotional well-being—that are not captured by administrative records or standardized tests.

We devote a separate appendix section to the Panorama student survey data because these data differ in important ways from the other data sources used in the paper. Unlike

[16]The final list of states included in the analysis are listed in Table A2.

the administrative records, surveys, and GPS data assembled directly by the research team, access to Panorama data required a mediated matching process conducted internally by Panorama to protect the confidentiality of its client schools. As a result, the structure, timing, and coverage of the Panorama data differ from those of the other datasets.

This appendix provides additional detail on the Panorama data, including the structure and coverage of the Panorama survey data, the pre-specified process used to select and aggregate primary outcomes, and the construction of the Panorama-specific analysis samples.

Panorama's student survey platform covers a broad range of topics related to students' academic experiences, engagement, and socio-emotional well-being. Across survey modules and grade levels, Panorama administers items addressing classroom engagement, attention, effort, school climate, belonging, relationships with teachers and peers, and emotional states (such as happiness, anxiety, and frustration). In total, the platform includes hundreds of distinct survey questions, many of which are grouped into conceptual domains such as classroom engagement, emotion regulation, sense of belonging, and positive and challenging feelings.

To protect the privacy of Panorama's client schools, Panorama does not disclose school identities or client identifiers to external researchers. Instead, Panorama provided data through a controlled matching process designed to preserve anonymity while enabling our school-level analysis. Specifically, we provided Panorama with a list of NCES school identifiers corresponding to the universe of schools included in our analysis, along with a set of school-level covariates required for our research design. Panorama internally matched this list to their client database and appended aggregated survey outcomes, constructing a school-by-year file containing mean responses to individual survey items. For each school–year in which Panorama surveys were administered, Panorama computed the mean response to each survey item across all responding students in that school and year.

The resulting dataset therefore contains school-by-year means of item-level survey responses, with no student-level data and no school names or direct identifiers. Schools are represented only by anonymized identifiers that are consistent across years but cannot be linked back to identifiable institutions by the research team.

The Panorama data span academic years 2019 through 2025. In total, the Panorama file includes 61,570 school–year observations with non-missing survey data, representing 17,429 unique schools across 46 states. Across all middle- and high school–year observations in the NCES panel during this period, 8.7 percent have at least one Panorama survey measure observed in a given year. Coverage varies substantially across states, schools, and years, reflecting the fact that Panorama surveys are administered at the discretion of districts and schools rather than through a nationally representative sampling design.

In addition to variation in whether schools administer Panorama surveys at all, there is

substantial heterogeneity in which survey modules and items are administered even among participating schools. Districts and schools select which Panorama survey instruments they want their students asked. As a result, conditional on observing any Panorama data for a given school–year, the specific set of survey items available can vary widely across schools and across years. This feature of the data implies that different items are observed for different subsets of schools.

Panorama survey administration is also uneven over time. Among school–year observations with survey data, coverage increases steadily after 2019, peaks in 2023–2024, and remains substantial through 2025. The number of school–year observations in each year ranges from roughly 4000 in 2019 to over 10,000 in 2025. Because survey participation and survey content are determined by school and district adoption and implementation decisions, the resulting sample is not nationally representative and should be interpreted as a large but plausibly selected set of schools.

Nature of the Measures and Selection of Primary Outcomes. Panorama survey items are typically measured on ordered five-point response scales. While exact wording varies slightly across survey modules, most items use conceptually similar scales capturing frequency (e.g., "Almost never" to "Almost always"), intensity (e.g., "Not at all focused" to "Extremely focused"), or agreement. Responses are coded on a numeric scale from 1 to 5. When necessary, items are reverse-coded so that higher values consistently correspond to more desirable outcomes. For example, items asking how often a student felt sad, worried, or frustrated are reverse-coded so that higher values indicate better emotional well-being.

Because Panorama offers a very large number of distinct survey items—234 unique items in our school–year file—any analysis of student survey outcomes raises concerns about multiple testing, specification searching, and attenuation from measurement error. These concerns are particularly salient in our setting because survey coverage is uneven across schools and years, schools select which survey modules to administer, and different items are observed for different subsets of schools. For these reasons, it was essential to commit ex ante to a small number of primary outcomes and to a transparent, pre-specified procedure for selecting and aggregating survey items.

We adopted a structured, multi-stage process to select our primary Panorama outcomes. We began with the universe of Panorama survey items and first restricted attention to closed-ended Likert-scale questions, excluding free-response items. We further limited the sample to student-reported items, excluding all staff- and parent-reported questions. We then removed duplicate or near-duplicate items that appeared across survey versions with only minor wording differences. Next, we excluded items belonging to topics with extremely low coverage, defined as appearing in fewer than 0.01 percent of school–year observations, as

such items provide little usable variation. Finally, the research team excluded items judged to be highly unlikely to respond to changes in student phone access during the school day. This filtering process yielded a set of 104 distinct student survey items for further consideration.

The first primary outcome was selected on conceptual grounds. A central motivation for school smartphone restrictions is concern about student mental health and emotional well-being, which has featured prominently in public discourse and policy debates surrounding phone bans. Panorama's survey platform includes two clearly defined constructs that map directly onto this concern: *Positive Feelings* and *Challenging Feelings*. These constructs consist of a small set of items asking students how often, during the past week, they experienced specific emotional states. The Positive Feelings items ask about feeling excited, happy, hopeful, loved, and safe, while the Challenging Feelings items ask about feeling angry, frustrated, lonely, mad, sad, and worried. These items are conceptually transparent, consistently worded, and widely used, making them a natural choice for a primary outcome capturing subjective well-being.

Selecting additional primary outcomes required a more deliberate filtering process. Beyond emotional well-being, claims about phone bans often emphasize improvements in students' ability to focus, pay attention, and engage in classroom activities, as well as changes in peer interactions, including online behavior. At the same time, the Panorama data include many constructs that are plausibly related to academic outcomes but are less directly tied to the mechanisms through which phone restrictions are expected to operate in the short run. To discipline this choice, we grouped the remaining items into thirteen candidate constructs that plausibly might respond to changes in phone access at school, including classroom attention, engagement, effort, peer interactions (including online bullying), and classroom climate.

We then solicited external input through a survey of secondary school teachers and administrators recruited via Prolific. Respondents were asked to imagine a school that shifted from having no school-wide phone restrictions to implementing a complete phone ban, with students completing the same survey before and after the policy change. For a randomized subset of approximately 30 survey items, respondents rated how likely each item was to change following the phone ban, using a three-point scale ranging from "not very likely to change" to "very likely to change." Items were randomly assigned across survey forms to limit cognitive burden. In total, 152 respondents completed the survey.

We averaged these ratings within each candidate construct and selected the construct with the highest average perceived responsiveness to phone restrictions. Across respondents, items related to students' in-class attention and perceived online bullying consistently emerged as among the most likely to respond to a school-wide phone ban. Based on this process,

and prior to examining treatment effects, we pre-registered classroom attention and online bullying as additional primary Panorama outcomes.

The classroom attention construct consists of seven items drawn from multiple Panorama survey modules, all of which capture students' ability to focus, pay attention, and exert effort during class. These include questions about how hard students try to pay attention when the teacher is speaking, how focused they are on class activities, how much effort they put forth during class, and how often they pay attention or follow directions in class. Although drawn from different modules, these items share a common conceptual focus and were consistently rated as highly responsive to phone restrictions by educators.

The online bullying outcome is based on a single survey item asking students how likely it is that someone from their school would bully them online. This item was rated by educators as highly likely to respond to changes in school-wide phone restrictions, consistent with the view that smartphones may facilitate real-time peer harassment during the school day.

Together, subjective well-being, classroom attention, and perceived online bullying capture three central channels through which smartphone restrictions are hypothesized to affect students: emotional experience, in-class focus, and peer behavior in online settings.

Index construction. We construct outcome indices using a pre-specified procedure designed to reduce measurement error and summarize information from multiple related survey questions in a transparent and parsimonious way. All steps in this procedure were determined prior to examining treatment effects. We first restrict attention to survey items with sufficient coverage, dropping any item observed in fewer than 0.5 percent of school–year observations. This drops three of the seven items in the classroom attention construct. Remaining items are grouped into outcome topics based on our pre-registered classification of survey items into conceptual constructs. Panorama provides survey data aggregated to the school level. When a school administers the same item in multiple survey waves within a single academic year (e.g., fall and spring), Panorama reports wave-specific school-level means; in these cases, we average across waves to obtain a single school–year value for that item.

Each item's school-year mean is first coded so that higher values correspond to more desirable outcomes and then standardized within survey year to have mean zero and unit variance. For each school-year, we compute outcome indices as the unweighted average of all available standardized items within the relevant construct for SWB and classroom attention. The online bullying outcome is constructed from a single standardized survey item. Finally, composite indices are re-standardized by year to have mean zero and unit variance. As a result, all Panorama outcomes are expressed in school-level standard deviation units, facilitating interpretation of effect sizes and comparison across outcomes.

Analysis sample. This subsection documents the construction of the analysis samples used for the Panorama-based outcomes. As with the other outcomes in the paper, all analyses are conducted at the school–year level and apply a common set of treatment-related restrictions, combined with outcome-specific availability requirements.

We begin with the universe of school–year observations sent to Panorama for matching, which includes 706,168 school–year observations spanning academic years 2019 through 2025. This initial panel corresponds to all public school–year observations in the NCES frame during this period. As in the test score, attendance, and discipline analyses, we exclude the pandemic-affected academic years 2020 and 2021 to maintain consistency across outcomes and to avoid years in which survey administration and school operations were highly disrupted. This restriction removes 201,076 school–year observations.

We next restrict the sample to schools serving at least one middle or high school grade, where smartphone use is more prevalent and where Yondr adoption is concentrated. After this restriction, we retain only school–year observations for which at least one of the three primary Panorama indices is observed. Because Panorama survey content varies substantially across schools and years, this step sharply reduces the sample relative to the initial NCES panel.

We then apply the same set of treatment-related sample restrictions used in the main analyses. Specifically, we drop schools that adopted Yondr prior to the 2022–23 academic year, schools located in districts identified as potentially contaminated control districts (i.e., districts that may have adopted Yondr but for which school-level adoption timing cannot be reliably determined), and Yondr-adopting schools for which implementation timing is classified as lowest confidence in the Yondr administrative data. Finally, we restrict the sample to states with at least one Yondr adoption during the 2023–2025 adoption window, ensuring that identification relies on within-state variation in adoption timing.

After applying all restrictions, the final subjective well-being analysis sample consists of 3946 school–year observations. The corresponding classroom attention analysis sample is larger, reflecting broader item coverage, and consists of 10,579 school–year observations. The online bullying analysis sample consists of 6041 school–year observations. Differences in sample size across the three outcomes arise solely from differences in item availability across schools and years; all treatment-related and institutional restrictions are identical.

As with the test score and attendance analyses, the Panorama samples are unbalanced panels. Schools enter and exit the sample depending on whether the relevant survey modules were administered in a given year. Table A16 summarizes the construction of the Panorama analysis samples.

D Imputation of Missing Test Scores

To compute average test scores when either English Language Arts (ELA) or mathematics scores are missing, we impute the missing subject-specific score at the school–year level using a trend-preserving procedure that combines (i) subject-specific time-series information and (ii) the year-specific deviation of the *observed* subject from its own local trend. We never alter observed values, and because the procedure requires one subject to be observed in the relevant year, we impute at most one subject for any school–year observation. Importantly, we use imputation solely to construct average test scores and apply it only to school–year observations in which one subject is missing; when we analyze ELA and mathematics scores separately, we do not impute missing values. In total, imputation affects 4,499 observations out of the 173,728 school–year observations in the final test score analysis sample.

Step 1: Identifying nearby observed values. For each school i, year t, and subject $s \in \{\text{Math}, \text{ELA}\}$, we identify:

- the nearest previously observed value $Y_{i,t',s}$ with $t' < t$, and
- the nearest subsequently observed value $Y_{i,t'',s}$ with $t'' > t$.

If both neighboring observations exist, we construct a subject-specific local baseline as their average:

$$\bar{Y}_{its}^{\text{surround}} = \frac{Y_{i,t',s} + Y_{i,t'',s}}{2}.$$

If only one neighboring observation exists, that value is used as the baseline. If neither exists, no baseline can be constructed for subject s in year t.

Step 2: Computing the deviation of the observed subject. Let $-s$ denote the subject opposite s (Math versus ELA). When $Y_{it,-s}$ is observed, we compute its deviation from its own local baseline:

$$D_{it,-s} = Y_{it,-s} - \bar{Y}_{it,-s}^{\text{surround}}.$$

If either the observed value or its corresponding baseline is missing, no deviation can be computed in that year.

Step 3: Imputing the missing subject. If subject s is missing in year t, and both of the following are available:

1. a subject-specific local baseline $\bar{Y}_{its}^{\text{surround}}$, and
2. the deviation $D_{it,-s}$ of the other subject in the same year,

we impute subject s as:

$$\widehat{Y}_{its} = \bar{Y}^{\text{surround}}_{its} + D_{it,-s}.$$

This construction preserves the level implied by subject s's nearby observations while incorporating the year-specific shock reflected in the observed subject. If either component is unavailable, the value remains missing. Because the deviation term requires the other subject to be observed in year t, the procedure never imputes both subjects in the same school–year.

Step 4: Constructing the average test score. Let $\widehat{Y}_{it,\text{Math}}$ and $\widehat{Y}_{it,\text{ELA}}$ denote the subject-specific values after imputation (observed or imputed). The average test score is constructed as:

$$\widehat{Y}^{\text{avg}}_{it} = \begin{cases} \frac{\widehat{Y}_{it,\text{Math}} + \widehat{Y}_{it,\text{ELA}}}{2}, & \text{if both subjects are available,} \\ \widehat{Y}_{it,\text{Math}}, & \text{if only Math is available,} \\ \widehat{Y}_{it,\text{ELA}}, & \text{if only ELA is available.} \end{cases}$$

If neither subject is observed nor can be imputed, the average score remains missing; this case does not occur in the analysis sample.

E Additional Robustness Checks for Test Score Estimates

Table A6 examines the sensitivity of our estimated treatment effects to alternative estimators, control group definitions, treated group restrictions, and outcome constructions.

Alternative estimators. We first assess sensitivity to the estimation approach. In addition to the baseline Callaway–Sant'Anna doubly robust estimator, we estimate a weighted two-way fixed effects specification that applies inverse probability weights constructed from the same baseline covariates and includes those covariates interacted with year fixed effects. We also implement the Callaway–Sant'Anna framework using regression adjustment only, omitting the inverse probability weighting component. This latter specification is motivated by the relatively small number of Yondr-adopting schools in the data, which could in principle generate small estimated propensity scores for some treated observations. Using regression adjustment alone provides a diagnostic check that the results are not driven by the weighting component of the doubly robust estimator.

Modifications to the control group. We next modify the composition of the control group. Using information from the NTS, we restrict the control group to schools whose modal respondent reports no strict phone policy changes in the past five years, and, in a more stringent specification, to schools reporting no phone policy changes of any kind over that period. These specifications assess whether the baseline results are sensitive to contemporaneous policy activity among non-Yondr schools.

We also present a specification that excludes schools scheduled to adopt Yondr in 2026 from the control group. In the baseline specification, these schools are included among never-treated schools. Excluding them addresses the possibility that future adopters differ systematically from schools that never adopt.

Finally, we report estimates that rely solely on within-adopter timing variation by restricting the sample to schools that eventually adopt Yondr and using later adopters as not-yet-treated controls. Although this specification eliminates comparisons to never-treated schools, it substantially reduces identifying variation because the number of schools scheduled to adopt in 2026 (approximately 528 in the test score sample) is limited.

Modifications to the treated group. We next restrict the treated sample to schools with high- and highest-confidence Yondr implementation dates. We also exclude schools whose Yondr accounts are recorded as canceled or associated with building closures. The baseline framework assumes that once a school adopts Yondr, it remains treated. However, Yondr's administrative data occasionally indicate that an account was later canceled or that a school building closed. Although Yondr data staff indicated that the quality of this indicator is imperfect, excluding these observations provides a check on whether potential treatment reversals or partial implementation affect the estimated effects.

Panel restrictions and outcome construction. We then examine several additional methodological considerations. First, we restrict the analysis to a fully balanced panel of schools observed in all six outcome years (2018, 2019, and 2022 through 2025). This is a particularly informative robustness check given that our main analysis combines test score data from two sources: SEDA, which does not extend through the 2024–25 school year, and state assessment data compiled by EDC, which does. By construction, restricting to a fully balanced panel implicitly limits the sample to schools with outcomes observed in all years, which in practice means relying only on the EDC data.

Second, we control for the log of the number of students tested to address the possibility that changes in test participation affect measured outcomes. Finally, we restrict the sample to observations with both Math and English Language Arts scores observed, excluding the

small share of average test score observations that rely on imputed subject-specific scores.

Across all specifications, the estimated effects remain small, negative, and relatively similar in magnitude to the baseline estimate. None of the alternative specifications yields evidence of measurable positive impacts of Yondr adoption on middle school standardized test scores over the period we observe.