Stanford | Institute for Economic Policy Research (SIEPR)

THE EFFECTS OF SCHOOL PHONE BANS: NATIONAL EVIDENCE FROM LOCKABLE POUCHES

Hunt Allcott
Stanford University and NBER

Thomas Dee
Stanford University and NBER

Matthew Gentzkow
Stanford University And NBER

E. Jason Baron
Duke University and NBER

Angela L. Duckworth
University of Pennsylvania

Brian Jacob
University of Michigan and NBER

APRIL, 2026

Working Paper No. 26-12

NBER WORKING PAPER SERIES

THE EFFECTS OF SCHOOL PHONE BANS: NATIONAL EVIDENCE FROM LOCKABLE POUCHES

Hunt Allcott
E. Jason Baron
Thomas Dee
Angela L. Duckworth
Matthew Gentzkow
Brian Jacob

Working Paper 35132
http://www.nber.org/papers/w35132

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2026

We acknowledge funding from Arnold Ventures, the Bezos Family Foundation, the National Governors Association, the Stanford Impact Labs, the Stuart Foundation, and the Walton Family Foundation. We thank Lila DiMasi, Kelly Flynn, Luca Moreno-Louzada, Liz Reosti, Chihiro Tanigawa, and Daniel Werner for excellent research assistance. We are grateful to Yondr for their partnership and for providing access to administrative data. We also thank Sahaan Sozhamannan, Peter Stiepleman, and Bianca Larry at Yondr for helpful collaboration and support. The study was pre-registered, and the pre-analysis plan is publicly available at https://osf.io/8twh9/overview. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w35132

The Effects of School Phone Bans: National Evidence from Lockable Pouches
Hunt Allcott, E. Jason Baron, Thomas Dee, Angela L. Duckworth, Matthew Gentzkow, and Brian Jacob
NBER Working Paper No. 35132
April 2026
JEL No. I28, I31, L86

ABSTRACT

Schools across the U.S. have sharply restricted student use of phones during the school day. We evaluate one type of restriction—lockable phone pouches—using nationwide data combining large-scale surveys, GPS pings, standardized test scores, and school administrative records, along with sales records from the largest pouch provider. Using a staggered difference-in-differences design, we find that pouch adoption substantially reduces phone use as measured by GPS pings and teacher reports. In the first year after adoption, disciplinary incidents increase and student subjective well-being falls, consistent with short-term disruption. However, effects on well-being become positive in later years and disciplinary effects fade. For academic achievement, average effects on test scores are consistently close to zero. High schools see modest positive effects, particularly in math, while middle schools see small negative effects. We find little evidence of effects on school attendance, self-reported classroom attention, or perceived online bullying.

Hunt Allcott
Stanford University
and NBER
allcott@stanford.edu

E. Jason Baron
Duke University
Economics Department
and NBER
jason.baron@duke.edu

Thomas Dee
Stanford University
Department of Economics
and NBER
tdee@stanford.edu

Angela L. Duckworth
University of Pennsylvania
Department of Psychology
duckwort@wharton.upenn.edu

Matthew Gentzkow
Stanford University
Department of Economics
and NBER
gentzkow@stanford.edu

Brian Jacob
University of Michigan
Gerald R. Ford School of Public Policy
and NBER
bajacob@umich.edu

1 Introduction

The impacts of phones on young people are hotly debated. An important part of this debate concerns phone use in schools. Scholars have argued that phone access during the school day disrupts attention and learning, weakens in-person relationships, and contributes to rising anxiety and depression among students (Haidt, 2023). Policymakers have responded by sharply restricting phone 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 (D'Addio, 2025; Prothero, Langreo and Klein, 2026). At the same time, others have questioned the effectiveness of these policies, emphasizing the potential educational benefits of phones, classroom disruptions created by enforcing limits, and limited causal evidence on the benefits of phone restrictions (Hess, 2026; Livingstone, 2026; Oster, 2026).

In this paper, we provide a nationwide, quasi-experimental analysis of the effects of phone restrictions in U.S. schools, focusing on one well-defined and rapidly expanding intervention: lockable phone pouches designed to limit students' access to phones during the school day. Under this approach, students place their phones in magnetically sealed pouches upon arrival at school and are generally not allowed to access them until dismissal. The intervention has been widely adopted, with nearly 5000 schools using pouches by 2026. It is an example of the kind of stringent, physically binding restrictions that proponents of limits on phones have advocated (Haidt, 2023). We use administrative data from Yondr, the leading provider of lockable phone pouches.

We implement a staggered difference-in-differences design that compares changes in schools that adopt pouches to those in observationally similar schools that do not adopt. We show that adopting schools differ along some observable baseline characteristics such as size, urbanicity, and student demographics. However, 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 observables. We examine pre-adoption trends and show that treated and comparison schools exhibit similar outcome trajectories prior to adoption. We use 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. We use the doubly robust implementation of this estimator, which combines inverse probability tilting and outcome regression. The details of our analysis were publicly pre-registered, including empirical specifications and outcome definitions; the pre-analysis plan is available here.

Our outcome measures draw on both administrative records and surveys. To measure changes in phone use, we use GPS-based data that track device activity during school hours.

We add school-level administrative outcomes for middle and high schools—including standardized test scores, attendance, and disciplinary incidents—and student survey data from Panorama Education, an education technology company that fields surveys measuring students' subjective well-being, classroom attention, and perceived online bullying. Finally, we supplement these data with two new national surveys we fielded between April 2025 and February 2026 that provide additional information on policy implementation and views of teachers, parents, and students. Because our data are at the school level, we standardize outcomes using the standard deviation of school-level means. To facilitate comparison with other interventions, we report corresponding student-level effect sizes using standard assumptions about the relationship between school- and student-level variation (What Works Clearinghouse, 2020).

We begin with evidence from our surveys on the overall landscape of U.S. phone policies. Understanding this landscape is important, because our estimator captures the effect of Yondr adoption relative to adopting schools' prior policies and the mix of policies in the control group. Overall, there is wide variation, even in states that have passed legislation requiring stringent policies. Some schools restrict usage only during class periods, while others restrict it throughout the school day. Implementation may be left to individual teachers, based on school-wide requirements that students keep their phones in their backpacks or lockers, or based on physical restrictions such as lockable pouches or centralized collection. However, two regularities emerge that are important for our analysis. First, most schools that adopt Yondr pouches previously operated under relatively lenient rules such as teacher discretion or "no-show" policies (under which students may carry phones but are prohibited from displaying or using them). Second, only a small share of control schools adopted stringent physical restrictions other than Yondr during the period we study. We show that our main results are robust to restricting the control group to schools reporting no recent phone policy changes, although our estimates become less precise in this case.

Our surveys also provide some initial evidence on the perceived impact of pouches. Teachers in schools that adopt pouches report reduced frequency of inappropriate usage by students in class and increased satisfaction with their schools' overall policy approach. Parents express support for pouches and expect improvements in academic, social, and mental health outcomes. Students tend to oppose pouches and expect less change in these outcomes.

Our first set of main results examines the effect of Yondr adoption on in-school phone use. This effect is of independent interest, since critics have cited student circumvention of Yondr pouches as a key limitation and suggested that their overall impact on usage may be limited (Kircher and Holtermann, 2026). Because no available dataset directly and comprehensively measures student phone use during the school day, we draw on two complementary sources of evidence. First, using GPS-based measures of device activity, we observe a large and persistent decline in phone activity on school campuses during school hours following adoption. The magnitude of this decline—roughly 30 percent in total GPS pings by the third year after adoption—cannot be read as a direct measure of the change in student phone use, since it also includes use by adults and because pings are often recorded when phones are on but not in use. But it confirms that the impact on student use is substantial and it can be read as a conservative lower bound on the magnitude. Second, teacher reports from our survey indicate large declines in student phone use following adoption. The share of students reported to be using their phones in class for personal reasons falls from 61 percent to 13 percent following Yondr adoption (a drop of roughly 80 percent), a substantially larger decline than in non-adopting schools.

Together, these results indicate that Yondr adoption meaningfully reduces in-school phone use, even if enforcement may be imperfect and the exact magnitude of the reduction cannot be read directly from any single measure. The treatment effects we estimate on downstream outcomes capture the effect of Yondr pouches as implemented in practice. This "reduced form" parameter speaks directly to policy, as Yondr is one of the most widely adopted forms of stringent phone restriction. We discuss below the extent to which our estimates can be extrapolated to other parameters of interest, such as the impact of a hypothetical perfectly enforced ban.

We next turn to the effects of Yondr adoption on student behavior and well-being. Disciplinary incidents increase in the year of adoption by approximately 0.03 student-level standard deviations—corresponding to roughly a 16 percent increase in suspension rates (in-school or out-of-school)—but this effect fades in subsequent years. Student-reported subjective well-being also exhibits a dynamic response: it declines in the year of adoption and then rebounds, becoming positive by the second post-adoption year. The decline is roughly 0.2 student-level standard deviations, followed by an increase of 0.16 standard deviations. For comparison, Allcott et al. (2020) find that deactivating Facebook for four weeks increased subjective well-being by 0.09 standard deviations. These patterns are consistent with some short-term disruption as Yondr pouches are adopted, followed by adaptation and subsequent improvements in well-being. It is important to note, however, that the relatively short panel limits our ability to fully separate dynamic effects from cross-cohort heterogeneity.

Turning to academic achievement, we find average effects on test scores that are close to zero over the first three years following adoption: we can rule out improvements larger than approximately 0.008 student-level standard deviations. These averages reflect small and precisely estimated effects across subjects, but mask interesting heterogeneity across grade levels. In high schools, we find modest positive effects—particularly in Math—of roughly 0.024 student-level standard deviations. For context, this corresponds to roughly a 0.9 percentile point increase in the test score distribution (von Hippel, 2025), or about one-fifth the achievement gains associated with a higher value-added teacher (Chetty, Friedman and Rockoff, 2014). In contrast, effects in middle schools are generally negative and even smaller in magnitude—roughly half as large.

We also find little evidence that restricting phone access improves other aspects of the school environment. Effects on the attendance rate are precisely estimated and close to zero: we can rule out improvements larger than 0.056 percentage points. Relative to a mean attendance rate of 93 percent, this corresponds to a change of less than 0.1 percent. We find similarly precise and small effects for chronic absenteeism and no measurable improvements for perceived online bullying or self-reported classroom attention. For classroom attention, we observe a negative and statistically significant estimate in the second post-adoption year; however, estimates for this specific outcome should be interpreted with caution, as there is some evidence of differential pre-trends in the subsample used for this measure.

Several mechanisms could plausibly explain these results. The increase in disciplinary incidents may reflect the enforcement of new phone restrictions (Figlio and Ozek, 2025). Alternatively, students may substitute from phones to other disruptive behaviors, including peer interactions that could lead to more disciplinary incidents (Jacob and Lefgren, 2003). The contrasting test score effects across middle and high schools may reflect a combination of behavioral responses and institutional constraints. Younger students may have more limited impulse control, making them more likely to substitute toward other disruptive behaviors when phones are removed. Differences in the magnitude of phone use reductions may also contribute: in our data, phone activity declines more in high schools than in middle schools, so middle schools may bear the costs of introducing and enforcing the restriction while generating smaller benefits.[1] Of course, these interpretations are necessarily speculative and not exhaustive.

Our analysis has several important limitations. First, our outcome measures capture only certain dimensions of student performance and school climate. While test scores and validated survey instruments are widely used in the education literature, they may not reflect all potential effects of phone restrictions. Second, because restrictions have expanded only recently, we observe outcomes for at most three years beyond adoption; longer-run effects remain an open question. Third, while we focus on a well-defined form of phone restriction, schools also use other strict approaches, including locker storage requirements or complete campus-wide bans, which may differ in both implementation and effects.

[1] Consistent with this interpretation, 72 percent of U.S. high school teachers report phone distraction as a major problem, compared with 33 percent of middle school teachers (Hatfield, 2024).

Our paper's primary contribution is to provide large-scale, nationwide quasi-experimental evidence on the effects of school phone restrictions in the U.S. We build on a small and mixed literature on school phone bans, most of which comes from outside the U.S. Beland and Murphy (2016) 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. Lichand et al. (2026) finds that a district-wide ban in Rio de Janeiro led to improvements in standardized test scores. Beneito and Vicente-Chirivella (2022) finds that bans in Spain led to improvements in PISA scores and reductions in bullying. Abrahamsson (2026) finds that bans in Norwegian middle schools improved girls' GPA, reduced mental health consultations, and lowered bullying.[2]

A recent study of Florida's statewide phone ban is perhaps most relevant for our context. Using administrative student records and phone 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 phone activity. They document short-run increases in disciplinary incidents followed by improvements in test scores and attendance in the second year after the ban. The estimated achievement gains are on the order of 0.6 percentile points overall and roughly 0.9 in middle and high schools, similar in magnitude to the 0.9 percentile point effect we estimate for high school math.

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, 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 phone activity, we leverage original survey data to enrich the analysis. Our teacher survey allows us to characterize the policy environment in both treated and control schools, document enforcement practices, and measure teacher-reported outcomes. 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 phone restrictions. Together, this approach provides a comprehensive assessment of how restrictive phone policies affect student outcomes.

Our paper is also related to research on the effects of personal digital devices and social media in adult populations. Experimental and quasi-experimental studies show that reducing

[2]In related work, Shi and Villarroel (2026) find that classroom-level restrictions in Chile modestly reduced in-class phone use but did not measurably improve test scores. 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.

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 phones during school hours.[3] 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 phone use in many ways—ranging from "no-show" policies 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 precise information on adoption timing and broad national coverage. This reduces measurement error in the key treatment variable, enables the construction of a large nationwide panel of schools, and therefore provides substantial statistical power. The breadth of the data also enhances the external validity of our findings and allows us to examine treatment effect heterogeneity across a wide range of school contexts. 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 phones during the school day and lower the enforcement burden on teachers and administrators. As a

[3] When the company was founded in 2014, it was mostly known by the public for providing pouches for concerts and music venues.

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 phone restrictions.

2.2 Conceptual Framework and Hypotheses

School phone restrictions are often justified on the grounds that constant access to digital devices undermines academic engagement and learning, disrupts student behavior and in-person interactions, and contributes to rising anxiety and depression among students (Haidt, 2023). Our goal is to evaluate whether restricting phone access in schools achieves these intended objectives. To do so, we examine outcomes that map onto these constructs.

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 phone restrictions could plausibly influence (Duckworth, Quinn and Tsukayama, 2012), we use them for several reasons. First, although state testing systems differ, standardized test scores can be collected and standardized within state-by-year cells to construct measures that are comparable across schools 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 phone access impacts academic performance and engagement, alters behavioral dynamics, or affects students' psychological well-being.

The expected effects of restricting phone access on academic performance and engagement are ambiguous ex ante. On the one hand, reducing access to phones 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 phones 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 phone 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 phones 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 (Kamenetz, 2025). Greater peer engagement could strengthen social relationships (Yondr, 2026), 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 phones 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 uncertain.

The expected effects on subjective well-being are similarly ambiguous. On the one hand, reducing access to phones 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 phone 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' 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.

More broadly, the effects of phone restrictions may also vary systematically across school contexts. In particular, differences between middle and high schools may shape both the magnitude of the reduction in phone use and the behavioral responses that follow. Phone use is generally more prevalent in high schools, suggesting greater scope for reductions in digital distraction, as we show below using GPS-based measures of phone activity. At the same time, younger students may have less developed self-regulation and may be more likely to substitute toward in-person peer interaction or other non-academic activities when phone access is restricted. Institutional features may also differ: classroom structure, enforcement practices, and the role of phones in instruction may vary across grade levels. As a result, the net effects of phone restrictions depend on the relative magnitude of these forces and could plausibly differ across settings even if the policy itself is similar.

3 Data

A key challenge for national research on school phone policies is the absence of standardized records documenting policy adoption and implementation. Although most school districts have some form of phone policy, official records are not systematically maintained or easily accessible, 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' subjective 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, we combine administrative adoption data from Yondr with phone 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.

Phone Location Data. To measure impacts on phone use, we use phone GPS data from Advan Research. These data are based on time-stamped location "pings" generated when mobile devices interact with partner apps or websites. Advan provides pre-processed school-level aggregates at an hourly resolution. It 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. 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).[4] We then divide by enrollment and form an academic-year panel by averaging these monthly measures over September, October, November, February, March, April, and May.

We interpret these measures as proxies for relative phone use across schools and periods, rather than direct measures of overall usage levels. 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. We therefore expect ping counts to fall when phone use declines.[5] That said, the GPS data are inherently noisy proxies for phone use: location events depend on app adoption, operating-system settings, permission choices, and the set of observable device changes over time as users install/uninstall apps or opt in/out. Moreover, we cannot distinguish pings generated by student phones from those produced by staff, teachers, or parents. We discuss these caveats as well as other data processing steps in more detail in Appendix C.2.

Despite these limitations, we present several validation exercises that suggest the data contain meaningful signal (see Appendix C.2). 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 Nationwide Teacher Survey (NTS), described next.

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 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 108,000 responses from educators in all 50 states. Throughout

[4]More specifically, the outcome variable in our regressions is $\log\Big(\big(V_{\text{weekday, day}} - V_{\text{weekend, day}}\big) - \big(V_{\text{weekday, night}} - V_{\text{weekend, night}}\big)\Big)$, where each V term denotes visits per enrolled student in the corresponding time period and day type.

[5]Advan's data are provided as scaled estimates based on the number of pings. These pings are recorded 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 paper, we restrict attention to responses from middle and high schools; in these settings, phone use is most prevalent and Yondr adoption is concentrated. The NTS sample includes roughly 50 percent of all middle and high schools in the country.

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 (e.g., throughout the whole school day, or just during classes) and where students are permitted to keep their phones (e.g., on their person, in lockers, or collected by staff). 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.

The survey further documents policy changes over time. Teachers report whether their school has experienced a major change in its phone 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 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 three primary purposes in our analysis. First, we use the CCD to define the analysis sample, restricting attention to middle and high schools. We classify schools as high schools if they offer grade 9 or above, and as middle schools if they offer grade 6 or above but not grade 9 or above. We also omit schools that CCD does not classify as "regular." A regular school is defined as a public institution that provides instruction leading to a standard high school diploma and does not primarily focus on special education, vocational/technical education, or alternative education.

Second, 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, 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, allowing 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.

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 use 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 indicators.

Third, 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 rates, and racial composition.

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 measure student achievement in middle schools, we combine state assessment data for grades 6 through 8 compiled by the Education Data Center (EDC) with test score estimates from the Stanford Education Data Archive (SEDA). All states administer standardized assessments in these grades, allowing us to obtain comparable achievement data across the country and over time.[6]

We begin with data from EDC's State Assessment Data Repository, 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 school-level English Language Arts (ELA) and math results for grades 6 through 8. 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

[6]The EDC's State Assessment Data Repository is sometimes informally referred to as "Zelma," the name of its AI interface for querying the data.

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. We then construct school-level achievement measures by taking weighted averages across grades, where weights correspond to the number of students tested on each exam. 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, 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-by-subject achievement estimates in student standard deviations, which we aggregate to the school-year level using the same procedure applied to the EDC data.[7]

Our main analysis combines these two sources to maximize coverage. As a robustness check, we also present results using the subset of states for which average scale scores are available from EDC, excluding SEDA-based measures. For each school-year, we construct an average achievement measure by combining standardized math and ELA scores. When one subject is missing (in roughly 3% of school–year observations), we construct the average using an imputation procedure described in Appendix D. The middle school test score sample includes 1,341 Yondr-adopting schools, representing 98 percent of the set of potential adopters available for this analysis. As discussed below, our main specifications focus on schools adopting Yondr between 2023 and 2025; accordingly, the relevant set of potential adopters for this analysis consists of middle school adopters in the 2023–2025 cohorts.

In contrast to assessments in grades 6 through 8, 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. Moreover, there is no centralized repository of high school assessment scores comparable to EDC or SEDA. For this reason, we compiled high school assessment data ourselves, relying on publicly available files on state department of education websites when available, and making requests to state departments to obtain information from other states. Ultimately, we were able to obtain high school assessment

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

data from 26 states representing 64 percent of high school students across the country. The high school test score sample includes 656 Yondr-adopting schools, representing 69 percent of the set of potential adopters available for this analysis.

Mirroring our data collection strategy for grades 6 through 8, we focused on school-level average scores for math and ELA assessments. In general, states utilize a single type of test—either end-of-course, end-of-grade, or college entrance exams—for official accountability purposes. However, in the few cases where a state administered more than one type of test, we incorporated data from all assessments. Appendix Table A3 details which high school assessments were administered in each state over our sample period. Importantly, regardless of the specific exam, these assessments were required for all students in a given grade or course, including ACT or SAT exams, which were typically administered to all grade 11 students in the state. For this reason, there should be no selection into testing beyond what one would expect in other mandatory assessment systems. Consistent with this, we show below that controlling for the number of students taking high school assessments in a given school–year does not affect our results.

As with grade 6 through 8 assessments, we first standardize school-level average scores by state-year-exam. While end-of-grade and college entrance exams are administered to students in a single grade, end-of-course exams can be taken by students in different grades due to when students take the particular course (e.g., some students may take Geometry in grade 9 while others take the course in grade 10). For this reason, we construct school-level achievement measures by taking weighted averages across *exams* within math or ELA using the number of students tested on each exam as weights. We follow the same imputation procedure as for middle school test scores when one of the two subjects is missing.

A small share of schools in our data serve both middle and high school grades and therefore appear in both the middle school (grades 6–8) and high school (grades 9–12) test score samples. This overlap accounts for approximately 7.1 percent of school–year observations in the test score data. In our baseline pooled specification, we assign such schools to the middle school sample. This choice reflects the fact that middle school test scores are more consistently available across states and years, providing a more standardized measure of achievement. As a robustness check, we instead construct a "stacked" dataset in which these observations are duplicated and treated as separate school-by-year-by-level observations, allowing both middle and high school test scores to enter the pooled analysis. As we show below, the resulting estimates are nearly identical to the baseline specification, indicating that our results are not sensitive to how overlapping schools are handled.

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 middle school 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.

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. For California, we therefore construct an additional 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 31 states between 2018 and 2025; for chronic absenteeism, data are available for 38 states. Appendix C.5 provides additional details on dataset construction, including the states included for each outcome; Tables A4 and A5 show data availability and definitions for each state. The attendance sample includes 1,711 Yondr-adopting schools, representing 84 percent of the set of potential adopters available for this analysis. For chronic absenteeism, it includes 1,786 adopters, representing 88 percent of the potential schools.

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.

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 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; (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.[8]

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 details the composition of the discipline index, including the share of observations derived from each underlying disciplinary measure. Data availability varies across states and years; Table A6 shows data availability and definitions for each state. The final analytic sample includes 22 states with discipline data between 2018 and 2025. The discipline sample includes 1,557 Yondr-adopting schools, representing 77 percent of the set of potential adopters available for this analysis.

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

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 or respondent 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.

[8] 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.

The Panorama data span academic years 2019 through 2025. In total, the Panorama file includes 62,289 school–year observations with non-missing survey data during this period, representing 16,285 unique schools across 46 states. Across all school–year observations in the NCES panel during this period, 6.9 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 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 phone restrictions are hypothesized to affect students.

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). 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. Finally, we restrict our analysis to regular schools only, as discussed above.

We then apply outcome-specific restrictions. For test score analyses, we require non-missing standardized math or ELA scores in that school-year. For attendance analyses, we require a non-missing 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 sample.

3.3 Interpreting School-Level Effect Sizes

Finally, because our outcomes are measured at the school level, estimated effects are expressed in units of the standard deviation of school-level means. These effects tend to overstate magnitudes relative to student-level effect sizes commonly reported in the education literature. To facilitate interpretation, we translate school-level estimates into approximate student-level standard deviations using standard assumptions about the relationship between within- and between-school variation (What Works Clearinghouse, 2020).

Specifically, we use intraclass correlation coefficients (ICCs) that vary by outcome type. For behavioral and survey-based outcomes, we use an ICC of 0.1, implying that student-level effect sizes are approximately one-third of the corresponding school-level estimates. For test scores, we use an ICC of 0.2, implying that student-level effect sizes are approximately half as large as the corresponding school-level estimates. Throughout the paper, we report treatment effects in school-level standard deviations and, where useful for interpretation, translate these into approximate student-level effect sizes using these conversions.

4 Descriptive Facts on School Phone Restrictions

This section presents descriptive facts on the landscape of school phone restrictions in the U.S. We draw on data from a nationally representative Gallup survey. In February 2026, we partnered with Gallup and the Walton Family Foundation to field a survey of approximately 2,000 adolescents ages 12–18, along with a parent in the same household.

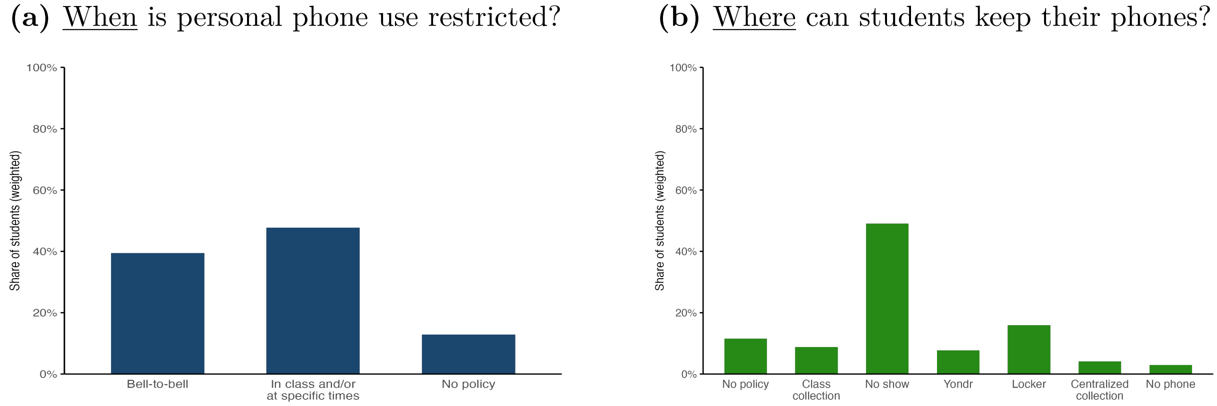
4.1 Schools Have a Wide Variety of Specific Phone Policies

Schools currently regulate student phone use in a wide variety of ways, both in terms of when use is restricted and where students are permitted to keep their devices during the school day. Panel A of Figure 1 shows the distribution of policies governing *when* personal phone use is restricted. 39 percent of students report their school has bell-to-bell restrictions (i.e., phones have to be away for the entire school day), while 48 percent say it allows phone use during non-instructional periods such as passing time or lunch. A nontrivial share of students—13 percent—report having no formal restrictions on personal phone use at their school. 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 49 percent of students, under which they may keep phones in backpacks or pockets but are required to keep them out of sight. 16 percent of students are required to store phones in lockers and 8 percent report using lockable pouches such as Yondr.

4.2 Parents Support Phone Bans While Students Oppose Them

Figure 2 displays results on parent and child attitudes toward phone policies. Parents and students were asked to imagine a policy in which schools require students to keep phones in lockable pouches throughout the school day. On a 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

Figure 1: Distribution of Phone Policies in the U.S.



Notes: This figure summarizes school phone policies using data based on student responses from the Gallup Survey. Panel A reports when phones are restricted during the school day, 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, with categories indicating whether phones are prohibited on campus, centrally collected, stored in lockers, secured in lockable pouches such as Yondr, kept out of sight ("no-show"), collected in classrooms, or governed by no school-wide policy. See Appendix C.4 for the exact wording of each question and for more detail about the data. Means are weighted to achieve a nationally representative sample of youth aged 12 to 29. The sample is restricted to middle and high school students in public or charter schools; N = 1198 for when; N = 1199 for where.

test scores (0.167), student relationships (0.141), and mental health (0.223), and on average tend to oppose lockable pouches (−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 sharply on support for the ban itself.

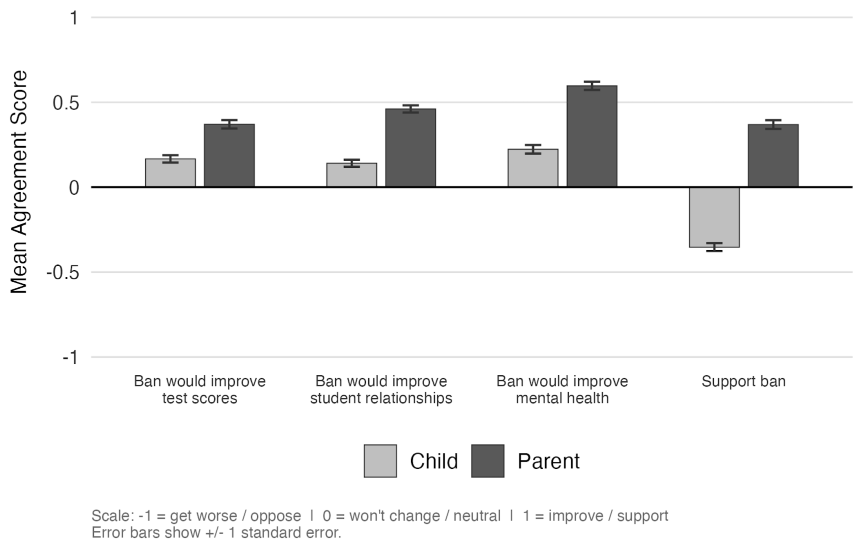
5 Difference-in-Differences Empirical Strategy

Estimating the causal impact of restricting student phone 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 first describe the standard two-way fixed effects (TWFE) specification that motivates

Figure 2: Parent and Student Attitudes Towards Phone 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 and child responses 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 wording of each question. Table A7 shows unweighted sample demographics.

our analysis. The TWFE analogue of our event-study specification can be written as:

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

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 test of the identifying assumption that, absent Yondr adoption, outcomes in treated and comparison schools would have followed parallel trends. 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 × 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 (e.g., see Roth et al. (2023)). 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 observed characteristics, which is important in our setting where adopters differ substantially from non-adopters.[9]

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 and never-treated schools serve as the comparison group.

The estimands of interest are group-time average treatment effects, denoted $\mathrm{ATT}(g,t)$, 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 (2023, 2024, or 2025). Once the $\mathrm{ATT}(g,t)$'s 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, which averages treatment effects across all post-adoption years and cohorts.

Finally, given the number of outcomes and estimates we consider, multiple hypothesis testing is an important concern. We address this in several ways. First, as discussed above, the analysis was pre-registered. Second, we focus interpretation on a relatively small set of ATTs for our primary outcomes. Third, we treat other estimates—including individual event-study coefficients and subgroup analyses—with greater caution. Throughout, we encourage readers to interpret these more granular results accordingly.

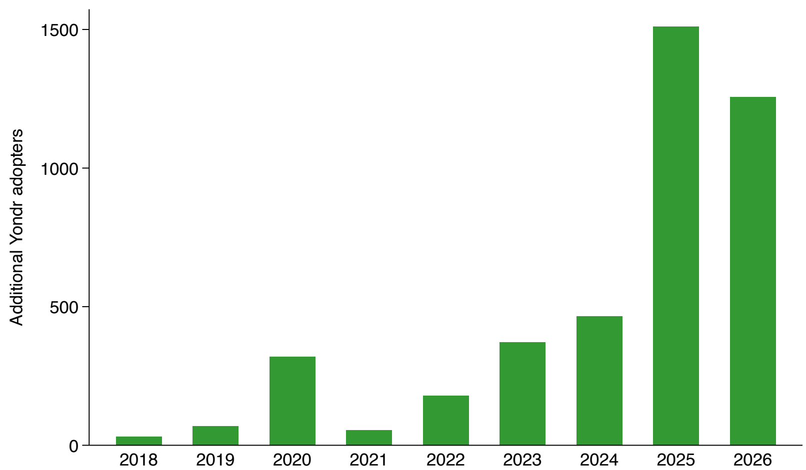
5.1 Characterizing Yondr Treatment and Control

Treatment. Figure 3 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

[9] We include the covariates listed above. The Callaway and Sant'Anna (2021) estimator conditions on baseline covariates to allow for differences in trends across groups defined by these covariates.

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.

Figure 3: 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. Each school appears once, in its initial year of adoption.

Table 1 reports summary statistics for schools in the pooled (middle and high school) test score sample, comparing Yondr adopters to never-treated schools. Columns (1) and (2) show means in the raw sample. Relative to never adopters, Yondr-adopting schools are larger, more likely to be located in urban areas, and serve student populations that have larger shares of Black and Hispanic students. 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 (excluding 2018 test scores). 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. Income-to-poverty ratios enter the weighting procedure as state-specific quartiles.

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	630	697	0.000	690	697	0.593
Share White	0.52	0.28	0.000	0.28	0.28	0.562
Share Black	0.15	0.30	0.000	0.31	0.30	0.137
Share Hispanic	0.25	0.34	0.000	0.33	0.34	0.393
Urban	0.27	0.58	0.000	0.59	0.58	0.537
Suburban	0.31	0.22	0.000	0.21	0.22	0.652
Rural/Town	0.42	0.20	0.000	0.19	0.20	0.745
Charter school	0.14	0.13	0.194	0.14	0.13	0.095
Schools with grades above 9	0.32	0.38	0.000	0.39	0.38	0.818
Schools with grades below 6	0.43	0.31	0.000	0.30	0.31	0.452
Income-to-poverty ratio	308	272	0.000	268	272	0.290
Avg. test score index	0.05	-0.45	0.000	-0.43	-0.45	0.441
Avg. Math score	0.05	-0.44	0.000	-0.41	-0.44	0.408
Avg. English score	0.05	-0.44	0.000	-0.40	-0.44	0.287
Number of schools	42,920	1777		39,825	1777	

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 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. While 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 and Sant'Anna (2021) estimator, which incorporates these covariates in a doubly robust way.

Table A8 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 A9). 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 62 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 23 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.

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 23 percent of NTS Yondr adopters report that the policy change coincided with other major school-level changes—such as changes in leadership, new curricula, or other institutional reforms—based on responses to Q20 of the NTS, which asks whether any non-phone-related changes occurred around the time of the policy change (see Appendix C.3 for the exact wording). This suggests 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.

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 associate with relatively low reported phone use and high teacher satisfaction in the NTS). 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 refine the control group in three ways: (i) restricting to schools that report no strict phone policy changes, (ii) restricting to schools that report no phone policy changes of any kind in the last five years, and (iii) restricting to schools that never implement a strict policy.

6 Difference-in-Differences Estimates

6.1 Effects on In-School Phone Use

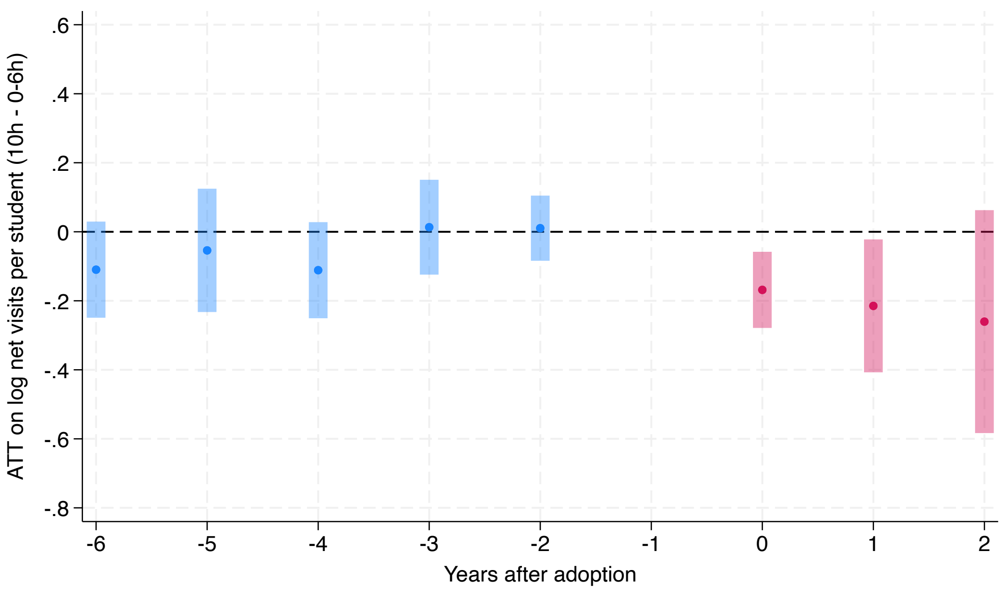
We begin by examining the effects of Yondr adoption on in-school phone use using two complementary sources of evidence: GPS-based measures of device activity and teacher-reported measures from the NTS.

Table A10 reports estimates of the ATT of Yondr adoption averaged across post-adoption periods and cohorts, based on the Callaway and Sant'Anna (2021) doubly robust estimator. The outcome is the natural log of net phone visits per student during instructional hours. In Column 1, our preferred specification, 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. Panel A shows that Yondr adoption reduces net phone visits by approximately 0.19 log points in the pooled sample (across middle and high schools).

These estimates are robust to alternative definitions of instructional hours, including broader windows such as 9:00–13:00 and 8:00–15:00 (Columns 2–3). Panels B and C present results separately for middle and high schools, showing larger effects in high schools. Appendix C.2 reports additional robustness checks using alternative samples. Across measures and outcome-specific samples—including the middle school and high school test score samples, as well as the attendance and discipline samples—the results consistently indicate a meaningful decline in in-school phone activity following Yondr adoption (Table A11).

Figure 4 presents event-study estimates of the ATT of Yondr adoption on the natural log of net phone visits per student between 10:00 and 11:00. The event-study coefficients show no evidence of differential pre-trends: estimates in the years prior to adoption are all statistically indistinguishable from zero. Following adoption, phone 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 0.21 log points in $t+1$ and nearly 0.3 log points by $t+2$.

Figure 4: Event Study of the Effects of Phone Pouches on Pooled High School and Middle School Phone GPS Activty



Notes: This figure reports event-study estimates of the effect of Yondr adoption on pooled (across middle and high schools) phone 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, 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 phone 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.

These estimates cannot be read as a direct measure of the change in student phone use and should be interpreted with caution. As discussed in detail in Appendix C.2, these GPS-based measures are best seen as proxies of student phone activity for several reasons.

First, we cannot distinguish student phone activity from that of teachers, staff, parents, or passersby. Second, these measures are inherently noisy because they depend on location signals generated through specific apps and partner data sources, which vary with user behavior, privacy settings, and other factors. Third, devices may record pings while they are turned on even if not accessible (e.g., in a locker or pouch). Fourth, Advan reports scaled estimates of the number of devices present in each location based on the underlying sample of observed pings, so true changes may be attenuated by their scaling and smoothing procedures. As a result, they likely provide a conservative lower bound on the true reduction in student phone use.

We next turn to survey evidence from the NTS. 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 experienced policy changes and others of which did 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 with the school's official cell phone policy increases from approximately 26 percent to 75 percent (48 percentage points). Reported in-class student phone use for personal reasons falls from 61 percent to 13 percent (48 percentage points), and reported phone use between classes declines by 53 percentage points. In contrast, among non-Yondr schools, satisfaction increases by 24 percentage points, while reported in-class phone use declines by 25 percentage points and between-class use declines by 20 percentage points. As discussed above, moves to strict restrictions among non-Yondr schools are relatively rare. However, the sizable changes observed among these schools suggest that many may be adopting less stringent alternatives—such as "no-show" rules or classroom-level collection—that nonetheless reduce phone use. Even so, the changes associated with Yondr adoption are substantially larger across all measures.

We emphasize that these comparisons are descriptive and rely on retrospective teacher reports. In addition, as discussed above, the Yondr schools observed in the NTS data represent only the subset of adopters that overlap with both our survey sample and the administrative adoption records and may not be fully representative of all Yondr schools nationwide. Despite these limitations, the survey evidence suggests relatively large reductions in student phone use following adoption. Together with the GPS-based results, these findings indicate that Yondr adoption meaningfully reduces in-school phone use, even if enforcement may be imperfect and the exact magnitude of the reduction cannot be measured directly.

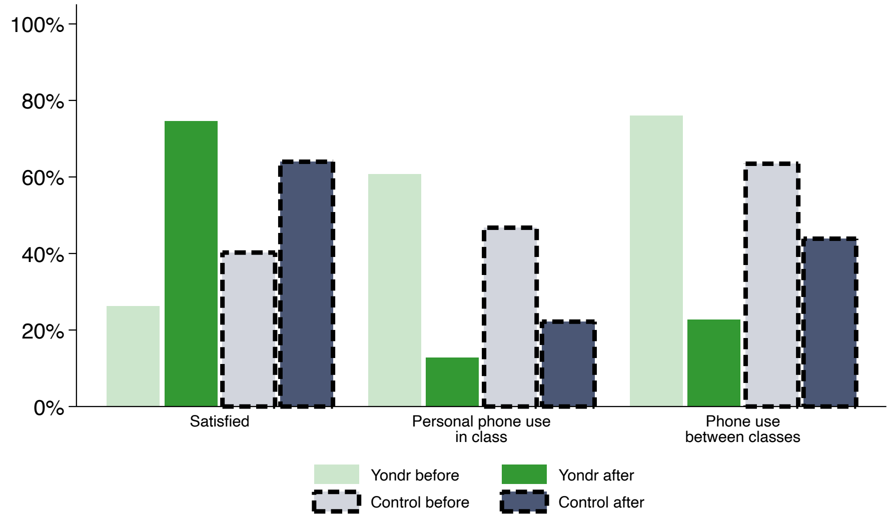
6.2 Effects on School Discipline, Attendance and Student Survey Outcomes

Having established that Yondr adoption leads to reductions in student phone use, we next examine its effects on disciplinary outcomes, student attendance, 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 relatively broad picture of how phone restrictions affect the school environment and student well-being. As discussed in Section 3, Panorama surveys are administered at the discretion of schools and are observed for a relatively small share of school-year observations, so these estimates reflect a potentially selected subset of schools and should be interpreted accordingly.[10]

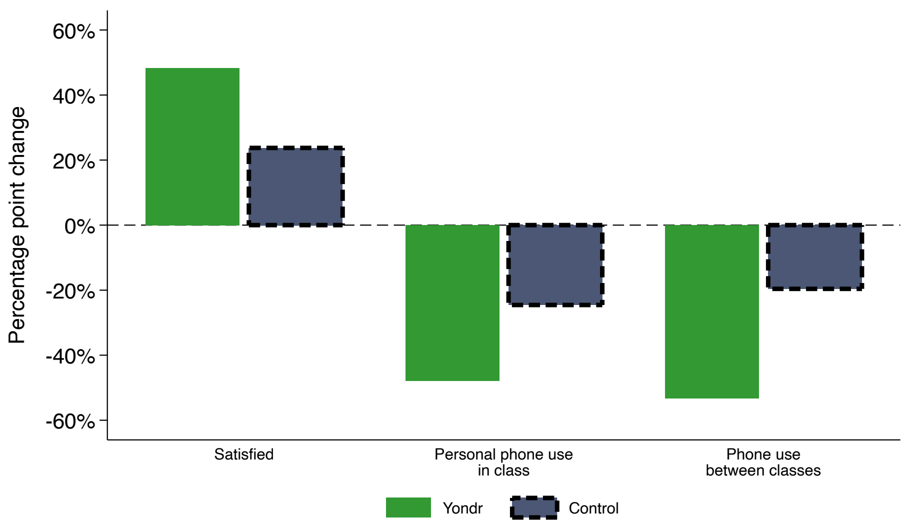
Consistent with our discussion of multiple hypothesis testing, we focus primarily on estimates of the ATT in the pooled sample (across middle and high schools) that average across post-treatment periods and cohorts, and use event-study specifications to illustrate dynamic patterns over time and assess the parallel trends assumption. Because several of the student outcome measures are unavailable during the 2020 and 2021 academic years due to the COVID-19 pandemic, we exclude these two years from the analysis. To maintain a consistent event-time structure, we relabel the 2018 and 2019 school years as 2020 and 2021, respectively, so that event times $t-5, \ldots, t-1$ correspond to successive 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.

[10]Moreover, 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 and Sant'Anna (2021) 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 below for test scores, the two approaches yield nearly identical estimates in settings where both can be implemented.

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. The sample is restricted to middle and high schools. 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.

Discipline and Attendance. Column (1) of Table 2 reports the ATT of Yondr adoption on the school-level disciplinary index. Adoption increases the disciplinary index by 0.059 (SE = 0.024) school-level standard deviations, corresponding to approximately 0.02 student-level standard deviations (see Section 3.3 for details on this conversion). To provide additional context, we translate these effects into percentage points on suspension rates (in-school or out-of-school). The estimates imply an increase of roughly 0.33 percentage points, corresponding to about an 11 percent increase relative to baseline levels.[11] Figure 6 (Panel A) reports event-study estimates for the disciplinary index. At adoption, disciplinary incidents rise sharply, by approximately 0.085 (SE = 0.020) school-level standard deviations, corresponding to roughly 0.03 student-level standard deviations or about a 16 percent increase in suspension rates relative to baseline levels. The point estimate moves toward zero in $t + 1$ and turns negative by $t + 2$, though these later estimates are imprecisely estimated.

Table 2: Difference-in-Differences Estimates of Effects of Phone Pouches on Pooled High School and Middle School Discipline, Attendance, and Student Survey Outcomes

	Discipline index (1)	Attendance rate (2)	SWB index (3)	Attention index (4)	Online bullying (5)
ATT	0.059**	−0.138	−0.305	−0.088	−0.039
	(0.024)	(0.099)	(0.201)	(0.085)	(0.092)
Mean of dependent variable	-	92.62	-	-	-
Observations	153,968	201,483	3382	9247	5648

Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption. Columns (1) and (2) use the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator, whereas Columns (3)–(5) use the regression adjustment implementation of the estimator. Column (1) reports effects on a standardized school-level discipline index constructed from suspension and discipline measures and standardized within state-by-year cells. Column (2) reports effects on the attendance rate. 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 (2) are in percentage points, and the estimates in all other columns are in school-level standard deviations. Never-treated schools serve as the comparison group. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Column (2) reports effects on the attendance rate. The estimate of -0.138 percentage points (SE = 0.099) allows us to rule out even modest improvements relative to the mean attendance rate of 93 percent. Figure 6 (Panel B) reports the corresponding event-study

[11] This calculation uses the within-state standard deviation of suspension rates to map standardized effects into percentage points. We illustrate this using California, which reports the share of students placed in any suspension. In California, the school-level standard deviation of suspension rates is approximately 5.6 percentage points. Multiplying this by the estimated effect of 0.059 implies an increase of roughly 0.33 percentage points, corresponding to about 11 percent of the baseline suspension rate. We obtain nearly identical estimates using other states with substantial Yondr adoption, including New York.

estimates. In the year of adoption, attendance declines slightly, by -0.193 percentage points (SE = 0.090), consistent with a short-run disruption effect. The magnitude is small, however, and estimates return to zero in subsequent periods, with no evidence of sustained changes.

Finally, Table A12 examines alternative attendance outcomes. We find no measurable effects on chronic absenteeism. In California, the only state reporting attendance by reason, we also find no effect on attendance net of disciplinary suspensions. These results suggest that the increase in disciplinary actions does not mechanically attenuate estimated effects on attendance.

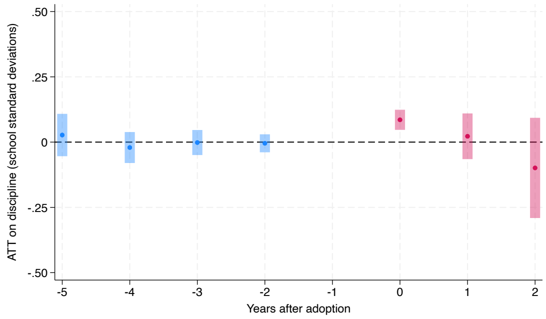
Student Survey Outcomes. Column (3) reports the ATT of Yondr adoption on SWB. The pooled estimate is -0.305 (SE = 0.201) school-level standard deviations, which rules out even modest positive effects. Figure 6 (Panel C) reports the corresponding event-study estimates. In the year of implementation, SWB declines sharply by approximately 0.603 school-level standard deviations and then recovers, becoming positive and statistically significant by $t + 2$ (approximately 0.485 standard deviations). This corresponds to roughly 0.199 and 0.16 student-level standard deviations in t and $t + 2$, respectively. For context, Allcott et al. (2020) find that deactivating Facebook for four weeks increased subjective well-being by 0.09 standard deviations.

Column (4) reports effects on classroom attention. The estimate is -0.088 (SE = 0.085), indicating no evidence of changes following adoption. Figure 6 (Panel D) reports the event-study estimates. Pre-adoption coefficients show some evidence of differential trends, suggesting that treatment effect estimates for this outcome should be interpreted with caution. With this caveat in mind, the point estimate in the year of implementation is modestly negative and not statistically significant. In $t + 1$, the estimate is slightly positive and also not statistically significant, while by $t + 2$ it turns negative and statistically significant.

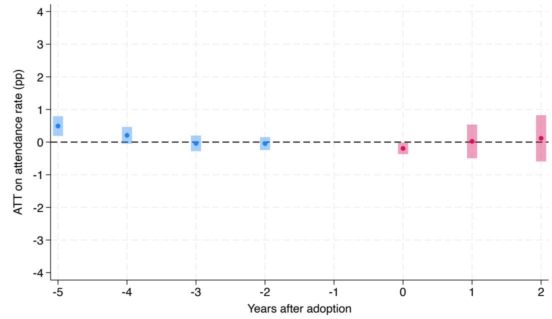
Column (5) reports the ATT on students' perceived likelihood of online bullying. The estimate is -0.039 (SE = 0.092), which is small and not statistically significant. Figure 6 (Panel E) reports the corresponding event-study estimates. Post-adoption coefficients are small and statistically insignificant, indicating no clear changes in perceived online bullying following adoption.

Heterogeneity by School Level and Cohort. We next examine heterogeneity by school level and adoption cohort across all outcomes. Figures A2 and A3, along with Table A13, report estimates separately for high schools and middle schools for discipline and attendance. Results are broadly similar across school levels, though disciplinary effects are somewhat larger and more precisely estimated in high schools.

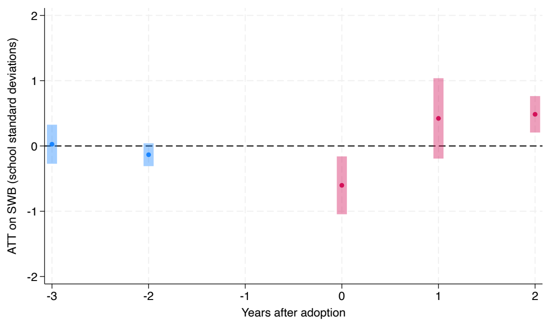
Figure 6: Event Study Estimates of Effects of Phone Pouches on Pooled High School and Middle School Discipline, Attendance, and Student Survey Outcomes



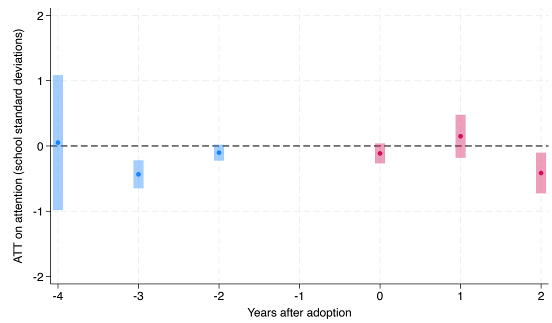
(a) Discipline Index



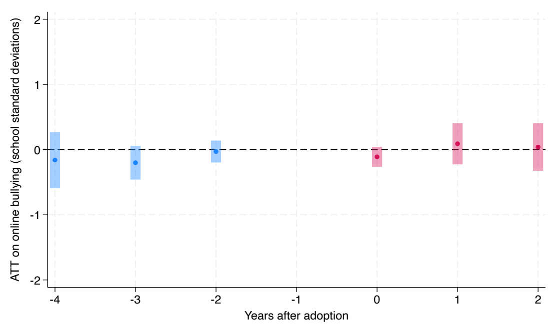
(b) Attendance



(c) Subjective Well-Being Index



(d) Classroom Attention Index



(e) Online Bullying

Notes: This figure reports event-study estimates of the effects of Yondr phone pouch adoption on pooled (across middle and high schools) discipline, attendance, subjective wellbeing, attention, and online bullying. Estimates in Panels (a) and (b) are based on the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator, whereas estimates in Panels (c), (d), and (e) use the regression adjustment implementation of the 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. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level. The shorter pre-treatment windows for the SWB, classroom attention, and online bullying outcomes reflect the later start of Panorama survey coverage.

Figures A4, A5, and A6, together with Table A13, report analogous results for student survey outcomes. For classroom attention, estimates are negative and statistically significant in high schools and close to zero in middle schools, though these differences should be interpreted cautiously given evidence of pre-trends, particularly in high schools.

Finally, Figures A7, A8, A9, A10, and A11 report event-study estimates separately by adoption cohort for each outcome. For discipline, increases at the time of adoption are generally present across all cohorts. For SWB, cohort-specific estimates are broadly consistent with the dynamic pattern observed in the pooled event study: for the 2023 cohort, we observe little change in the year of adoption followed by improvement in subsequent periods, which becomes statistically significant by $t + 2$, whereas for the 2025 cohort—where we only observe the adoption year—we find a large and statistically significant decline at t. Thus, the initial decline in SWB is driven primarily by later-adopting cohorts, while the subsequent recovery reflects earlier-adopting cohorts. These results highlight that the relatively short panel and staggered adoption make it difficult to fully disentangle dynamic treatment effects from cross-cohort heterogeneity.

6.3 Effects on Test Scores

We next examine whether restricting phone access affects student academic achievement. Table 3 reports average treatment effects across post-adoption periods and cohorts. In the pooled sample of middle and high schools (Panel A), the estimated effect on the combined test score index is close to zero (-0.004, SE = 0.010), with similarly small estimates for Math and ELA. This corresponds to approximately -0.002 student-level standard deviations, and the precision of the estimates allows us to rule out improvements larger than approximately 0.008 student-level standard deviations. Effects of this magnitude are extremely small relative to 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 ELA.

These estimates reflect the effect of Yondr adoption as implemented in practice, which reduces phone use by a substantial but incomplete amount. This is a policy-relevant treatment effect, as it captures the impact of one of the most widely adopted forms of stringent phone restriction under real-world conditions. To assess how our estimates relate to other parameters of interest, it is useful to consider how these magnitudes would scale under larger reductions in phone use. Our estimates on average test scores imply improvements no larger than approximately 0.008 student-level standard deviations under the observed policy. If one assumes that the effects on achievement scale linearly with the reduction in phone use, the bounds implied by our estimates remain small even under much larger interventions. For example, if our Yondr treatment reduces student phone use by 80 percent, as indicated by the survey evidence, the corresponding bound on the effect of a hypothetical intervention that reduced use to zero is approximately 0.01 student-level standard deviations. If the Yondr treatment reduces student phone use by 20 percent—a highly conservative bound based on our GPS evidence—the bound would be 0.04 student-level standard deviations. These calculations suggest that, even under conservative assumptions about the intensity of the treatment, the potential academic gains from restricting phone use during the school day may be modest.

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

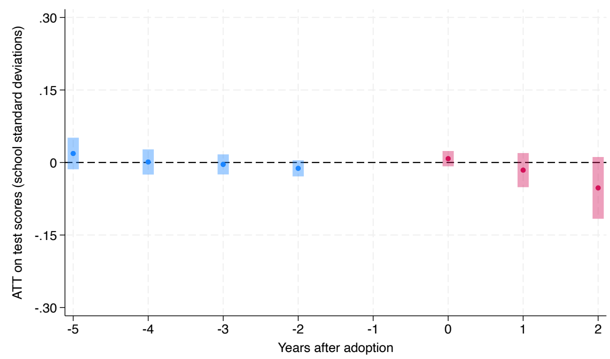
	Combined score (1)	Math score (2)	ELA score (3)
Panel A: Pooled			
ATT	−0.004	−0.003	−0.002
	(0.010)	(0.012)	(0.012)
Observations	214,556	210,951	211,259
Panel B: High School			
ATT	0.025	0.048***	0.011
	(0.017)	(0.018)	(0.022)
Observations	65,978	64,502	65,318
Panel C: Middle School			
ATT	−0.024**	−0.027*	−0.017
	(0.012)	(0.014)	(0.013)
Observations	162,974	159,048	159,987

Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption on average standardized test scores, measured in school-level standard deviations. Column (1) reports estimates for the combined math and ELA 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. Never-treated schools serve as the comparison group. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

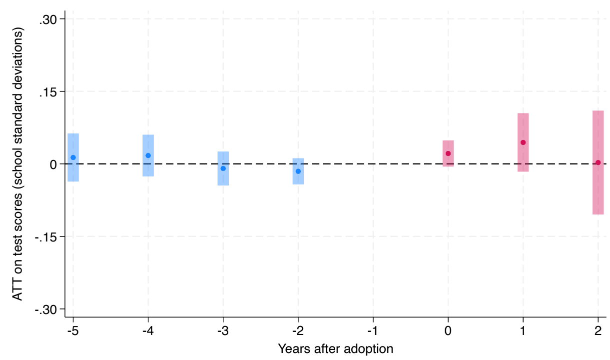
Panel A of Figure 7 reports the corresponding event-study estimates. The pre-adoption coefficients are small and statistically indistinguishable from zero. The estimate in the year of adoption is also close to zero. In subsequent years, the estimates become slightly negative but remain statistically insignificant.

However, these pooled estimates mask interesting heterogeneity across school levels. Panels B and C of Table 3 and Figure 7 report effects separately for high schools and middle

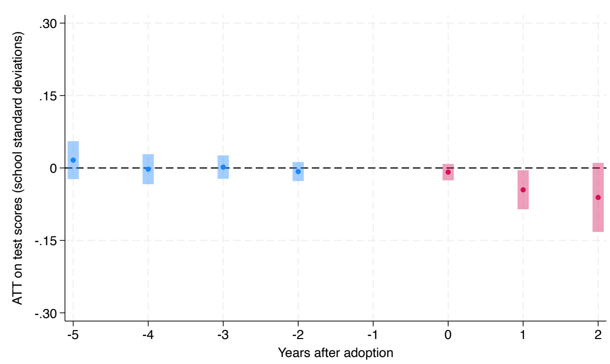
Figure 7: Event Study Estimates of Effects of Phone Pouches on Pooled, High School, and Middle School Test Scores



(a) Pooled



(b) High School



(c) Middle School

Notes: This figure reports event-study estimates of the effect of Yondr adoption on average test scores, measured in school-level standard deviations, for the pooled sample and separately for high schools and middle schools. 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. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level.

schools. In high schools, the combined score increases by 0.025 (SE = 0.017), with a statistically significant increase in math of 0.048 (SE = 0.018) and a near-zero estimate for ELA. In middle schools, the combined score declines by -0.024 (SE = 0.012), with a decline in math of -0.027 (SE = 0.014) and a smaller, statistically insignificant estimate for ELA.

These results suggest that the near-zero average effect may reflect offsetting positive effects in high schools and negative effects in middle schools. In high schools, effects are concentrated in math, with an effect of approximately 0.048 school-level standard deviations, which translates to 0.024 student-level standard deviations. Effects of this magnitude are modest relative to many other educational interventions. For comparison, this corresponds to an increase of roughly 0.9 percentile points in the test score distribution (von Hippel, 2025) and about one-fifth of the gain associated with a one standard deviation increase in teacher value added (Chetty, Friedman and Rockoff, 2014). In middle schools, effects are negative and also concentrated in math, but are even smaller—roughly half the magnitude of the high school effects.[12] This concentration of effects in math is consistent with a broader literature in the economics of education finding that interventions often generate larger or more detectable impacts in math than in ELA (e.g., Dee and Jacob, 2011; Dobbie and Fryer Jr, 2011; Chetty, Friedman and Rockoff, 2014).

We also examine event-study estimates separately by adoption cohort for the pooled sample in Figure A14. The post-adoption estimates suggest somewhat more positive effects in later cohorts, though these patterns should be interpreted cautiously since we observe a limited number of post-adoption periods. For the 2023 cohort, test scores decline modestly in the years following adoption, with effects around 0.05 school-level standard deviations. For the 2024 cohort, estimates are small and positive in both t and $t+1$, though not statistically significant. For the 2025 cohort, for which we observe only the adoption year, the estimate is small, positive, and statistically significant (0.030 school-level standard deviations).

6.4 Robustness Checks

Given the central role of academic achievement in the policy debate and its importance as a measure of human capital accumulation, we focus our most extensive robustness checks and additional heterogeneity analyses on test score outcomes. Tables A14 and A15 report a range of alternative specifications. Across these specifications, pooled estimates (across middle and high schools) remain close to zero and precisely estimated, while the pattern of positive effects in high schools and negative effects in middle schools is preserved. Table A15 further shows that this heterogeneity is robust when focusing specifically on math test scores, where effects

[12] Figures A12 and A13 present the corresponding event-study estimates for math and ELA, respectively, in the pooled sample and separately by school level.

are most pronounced: high school math effects remain positive and statistically significant, while middle school math effects remain negative, smaller in magnitude, and statistically significant in most specifications.

We first assess sensitivity to the estimation approach. Using the Callaway and Sant'Anna (2021) estimator with regression adjustment only, rather than the doubly robust implementation, yields nearly identical estimates. We next modify the composition of the control group using information from the NTS. Specifically, we consider four alternative definitions: (i) restricting to schools observed in the NTS whose reported policy is not Yondr; (ii) restricting to NTS schools that report either no policy change in the past five years or only weak policy changes, excluding those adopting strict alternatives (centralized collection, locker storage, or campus-wide bans); (iii) restricting to NTS schools that report no policy change; and (iv) excluding schools that implement strict phone policies from the control group, regardless of whether a recent change occurred.

We also examine alternative comparison groups and sample restrictions. Excluding future adopters (the 2026 cohort) from the control group, as well as restricting the sample to schools adopting between 2023 and 2026 to leverage within-adopter timing variation, yields qualitatively similar estimates. Additional specifications that restrict the treated sample (to high-confidence implementation dates or by excluding schools with canceled accounts), modify sample construction (e.g., by restricting to a balanced panel or excluding imputed test scores), or control for the number of students tested produce similar results.

We next consider additional robustness checks that address how school characteristics and sample construction may affect the pooled estimates. First, in the pooled specification, we augment our baseline controls by allowing the state-by-urbanicity indicators to vary by school level (middle versus high). Second, we address the fact that, in the test score data, a small share of school-year observations (roughly 7.1%) appear in both the middle and high school samples. In our baseline specification, we prioritize the middle school test score for these observations. As a robustness check, we instead construct a "stacked" dataset in which these observations are duplicated and treated as separate observations, allowing both middle and high school test scores to enter the pooled analysis. Both checks yield nearly identical estimates. Appendix E provides further details on these robustness checks.

Table A16 provides a complementary check by examining how the estimated ATT changes as we incrementally expand the set of covariates, culminating in the baseline specification. While the estimates decline somewhat with the introduction of demographic controls, they stabilize thereafter. Because our identification strategy relies on parallel trends conditional on these baseline characteristics, this stability suggests that the results are not driven by differential trends in observable school characteristics.

Finally, because our analysis relies on school-level outcomes, compositional changes represent a potential concern. Table A17 examines whether Yondr adoption is associated with changes in school enrollment or student demographics. We consider total enrollment, racial composition, and the share of students eligible for free or reduced-price lunch. Across these measures, we find little evidence of meaningful compositional change. Changes in enrollment and demographic composition are either small relative to baseline levels or statistically insignificant, with similar patterns across high schools and middle schools.

To further assess whether compositional shifts could account for the test score results, Figure A15 reports event-study estimates using predicted test scores based on school composition. These estimates are close to zero throughout the event window, with no evidence of systematic changes following adoption. The overall ATT is -0.000 (SE = 0.005) in the pooled sample, and similarly small and statistically insignificant across school levels. These results indicate that compositional changes are unlikely to explain the test score findings.

6.5 Subgroup Analysis

Before exploring any subgroup analysis, we pre-registered dimensions of heterogeneity that we thought were most important to test. This section presents a subgroup analysis along those dimensions. Table A18 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. Because subgroup-specific test scores are not available for high schools, these within-school subgroup estimates are limited to the middle school sample. The number of observations varies across columns because subgroup-specific test scores are not reported for some school–years with small cell sizes.

Across the pre-registered subgroups, the estimates differ across school levels. In high schools, effects are positive for both high- and low-poverty schools, and are larger for higher-poverty schools than for lower-poverty schools, though the latter estimate is not statistically significant and the two estimates are not statistically distinguishable. In middle schools, the estimates are generally negative and small in magnitude, with larger declines in lower-poverty schools than in higher-poverty schools. Effects are broadly similar across gender. Across racial and ethnic groups, estimates are most negative for White students and somewhat smaller for Black students. Finally, effects are more negative for students who are not economically disadvantaged than for economically disadvantaged students. These patterns suggest that the modest positive effects observed in high schools are relatively broad-based, though somewhat larger for high-poverty schools, while the absence of positive average effects in middle schools does not mask benefits for any particular group.

7 Conclusion

Phone 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 phone activity as measured by independent GPS data and teacher reports. These effects are large and persistent, indicating that lockable pouches meaningfully tighten phone access relative to prevailing alternatives.

We find that adopting lockable phone pouches appears to generate some short-run disruption to the school environment. Disciplinary incidents increase and student-reported subjective well-being declines in the year of implementation, consistent with adjustment costs as schools and students adapt to the new restriction. Over time, however, disciplinary impacts fade and well-being rebounds, becoming positive in subsequent years. Average effects on standardized test scores are close to zero and precisely estimated, with similarly small and null effects on attendance, classroom attention, and perceived online bullying. While the average academic effects mask interesting heterogeneity—modest positive effects in high schools and small negative effects in middle schools—the magnitudes are relatively small in both settings.

Our study leaves several questions for future research. We observe outcomes for at most three years beyond adoption, so longer-run effects remain an open question. We also focus on one specific and well-defined implementation design. Other forms of phone 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. 2026. "Smartphone bans, student outcomes and mental health." *Journal of Human Resources*.

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." *Quarterly Journal of Economics*, 141(1): 1–49.

Beland, Louis-Philippe, and Richard Murphy. 2016. "Ill communication: technology, distraction & student performance." *Labour Economics*, 41: 61–76.

Beneito, P., and Ó. Vicente-Chirivella. 2022. "Banning mobile phones in schools: evidence from regional-level policies in Spain." *Applied Economic Analysis*, 30(90): 153–175.

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 Working Paper 33584.

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.*

Dee, Thomas S., and Brian Jacob. 2011. "The impact of No Child Left Behind on student achievement." *Journal of Policy Analysis and Management*, 30(3): 418–446.

Dobbie, Will, and Roland G Fryer Jr. 2011. "Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone." *American Economic Journal: Applied Economics*, 3(3): 158–187.

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.

Haidt, Jon. 2023. "The Case for Phone-Free Schools." *After Babel*. Available at https://www.afterbabel.com/p/phone-free-schools. Accessed February 15, 2026.

Hatfield, Jenn. 2024. "72% of U.S. high school teachers say cellphone distraction is a major problem in the classroom." *Pew Research Center*. Available at https://www.pewresearch.org/short-reads/2024/06/12/72-percent-of-us-high-school-teachers-say-cellphone-distraction-is-a-major-problem-in-the-classroom/. Accessed April 17 2026.

Hess, Rick. 2026. "What's the Right Way to Limit Phones in School?" *Education Week*. Available at https://www.edweek.org/technology/opinion-whats-the-right-way-to-limit-phones-in-school/2026/02. Accessed April 17, 2026.

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*. Available at https://nymag.com/intelligencer/article/how-new-york-public-school-phone-ban-saved-high-school.html. Accessed April 17, 2026.

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*. Available at https://www.nytimes.com/2026/02/25/style/yondr-pouch-school-phone-ban.html. Accessed March 3, 2026.

Lichand, Guilherme, Luca Moreno-Louzada, Thiago da Costa, and Matthew Gentzkow. 2026. "The Educational Impacts of School Phone Bans: Evidence from Brazil." *Working Paper*.

Livingstone, Sonia. 2026. "The Case Against School Cell Phone Bans." *JAMA Pediatrics*, 180(4): 362–363.

Oster, Emily. 2026. "Did Smartphones Cause Global Test Score Declines?" *ParentData*. Available at https://parentdata.org/kids/smartphones-global-test-score-declines/. Accessed April 12, 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*. Available at https://www.edweek.org/technology/which-states-ban-or-restrict-cellphones-in-schools/2024/06. Accessed April 17, 2026.

Roth, J., P. H. Sant'Anna, A. Bilinski, and J. Poe. 2023. "What's trending in difference-in-differences? A synthesis of the recent econometrics literature." *Journal of Econometrics*, 235(2): 2218–2244.

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." *IZA Discussion Paper 18426*.

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.

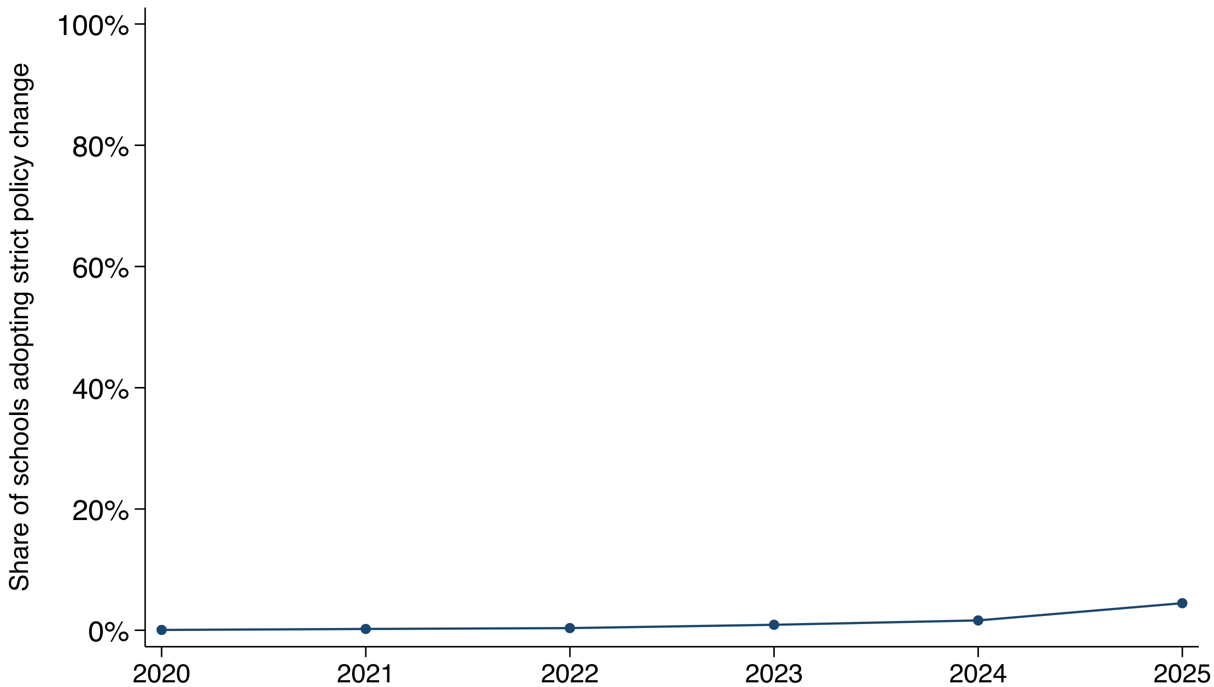
Yondr. 2026. “Phone-Free Schools Annual Trend Report: Insights, Impact, and a Look Ahead ’26.” Yondr Annual Trend Report.

Appendix

Table of Contents

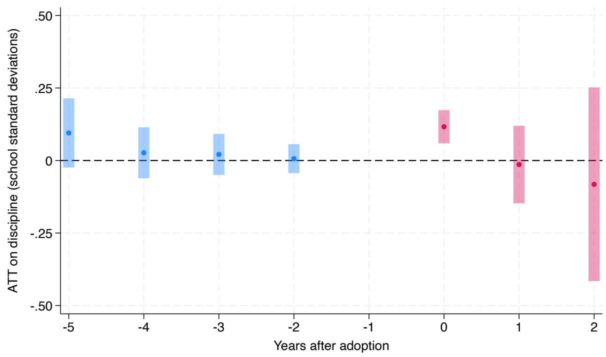
A Additional Figures

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

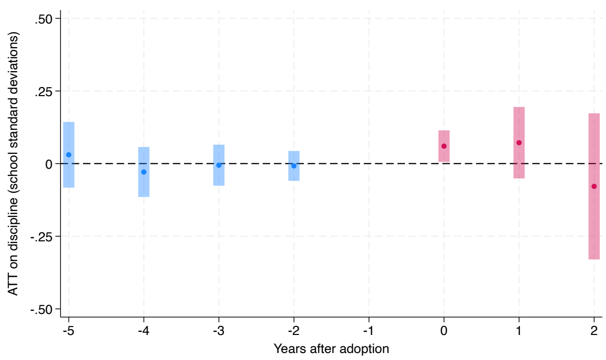


Notes: This figure uses data from the NTS to show phone policy changes among schools that do not adopt Yondr. We combine responses to Q10–Q11 (whether the school experienced a major phone policy change in the past five years and the timing of that change) with the school's current policy reported in 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. Each point reports the percentage of non-Yondr schools in the NTS 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 High School and Middle School Discipline



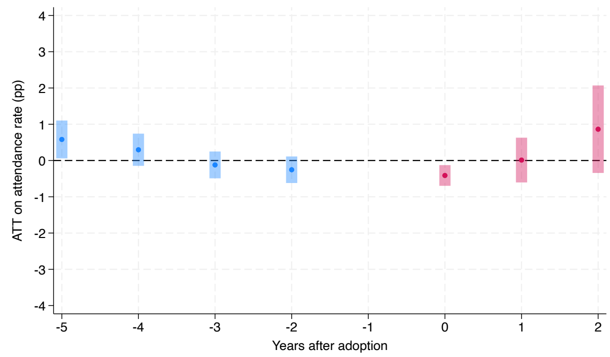
(a) High School



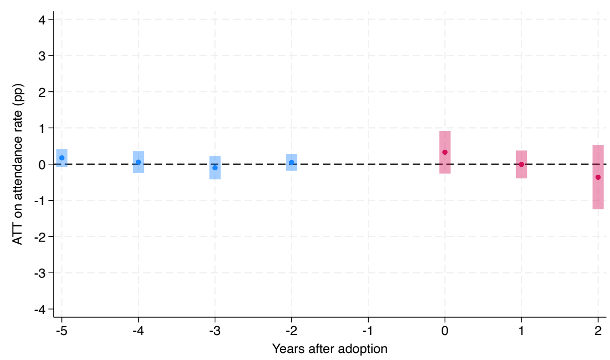
(b) Middle School

Notes: This figure reports event-study estimates of the effect of Yondr adoption on the discipline index, measured in school-level standard deviations, separately for high schools and middle schools. 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. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level.

Figure A3: Event Study Estimates of Effects of Phone Pouches on High School and Middle School Attendance Rates



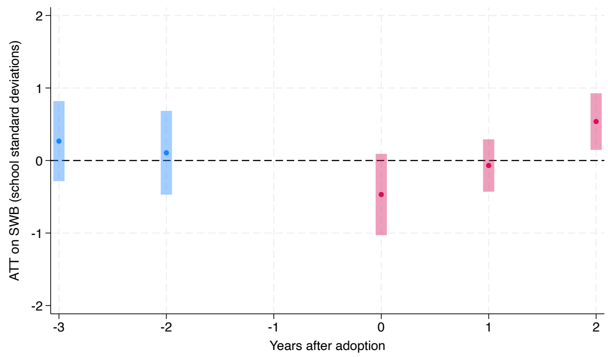
(a) High School



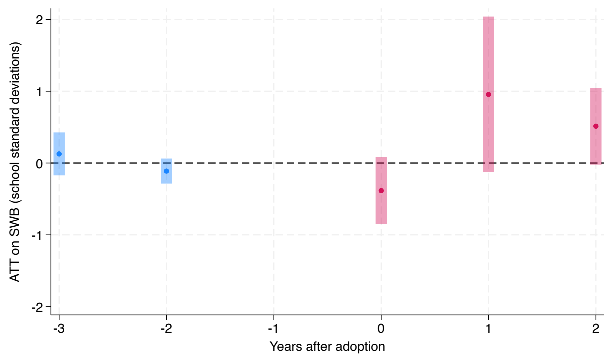
(b) Middle School

Notes: This figure reports event-study estimates of the effect of Yondr adoption on attendance rates, measured in percentage points (0-100), separately for high schools and middle schools. 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. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level.

Figure A4: Event Study Estimates of Effects of Phone Pouches on High School and Middle School Subjective Well-Being



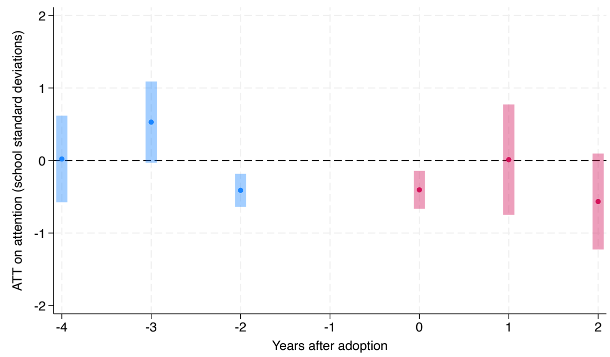
(a) High School



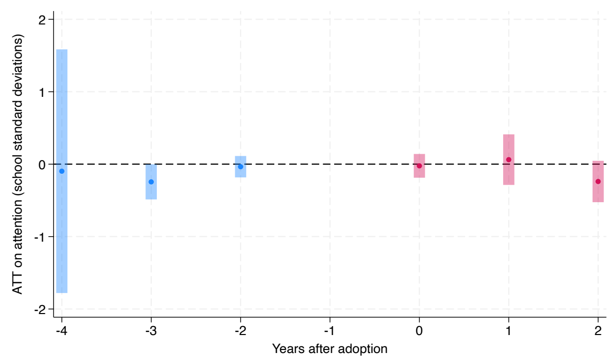
(b) Middle School

Notes: This figure reports event-study estimates of the effect of Yondr adoption on subjective well-being, measured in school standard deviations, separately for high schools and middle schools. Estimates are based on the regression adjustment implementation of the Callaway and Sant'Anna (2021) 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. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level.

Figure A5: Event Study Estimates of Effects of Phone Pouches on High School and Middle School Classroom Attention



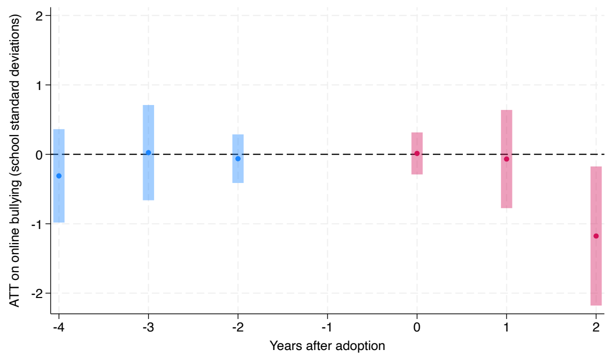
(a) High School



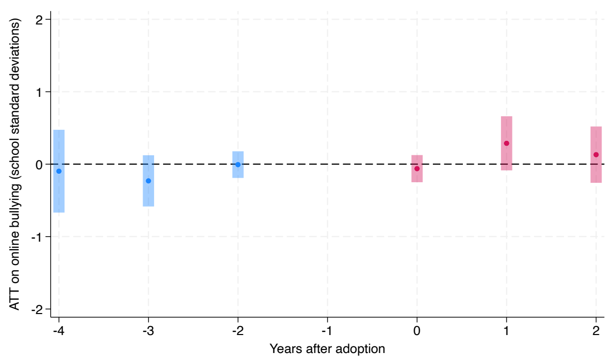
(b) Middle School

Notes: This figure reports event-study estimates of the effect of Yondr adoption on classroom attention, measured in school standard deviations, separately for high schools and middle schools. Estimates are based on the regression adjustment implementation of the Callaway and Sant'Anna (2021) 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. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level.

Figure A6: Event Study Estimates of Effects of Phone Pouches on High School and Middle School Online Bullying



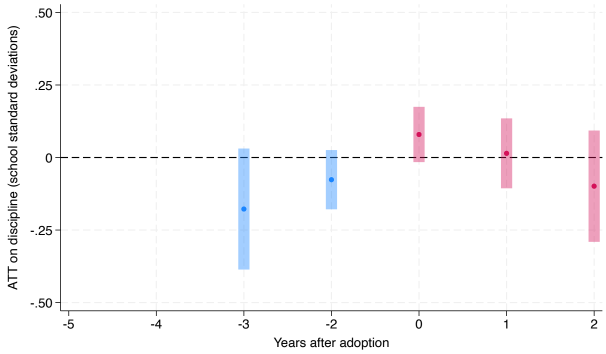
(a) High School



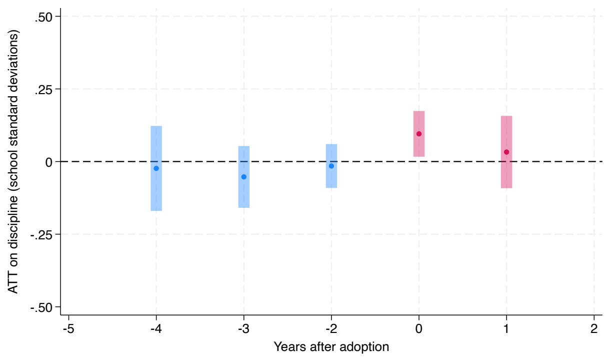
(b) Middle School

Notes: This figure reports event-study estimates of the effect of Yondr adoption on online bullying, measured in school standard deviations, separately for high schools and middle schools. Estimates are based on the regression adjustment implementation of the Callaway and Sant'Anna (2021) 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. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level.

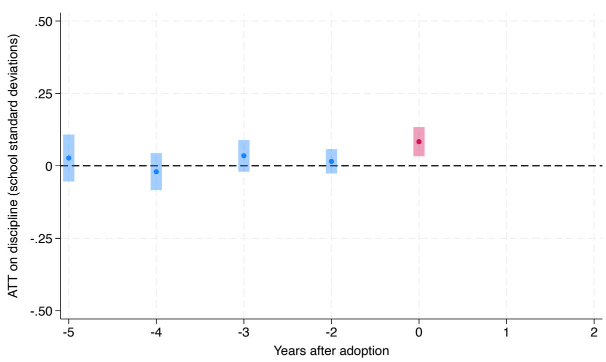
Figure A7: Event Study Estimates of Effects of Phone Pouches on Pooled High School and Middle School Discipline by Cohort



(a) G = 2023



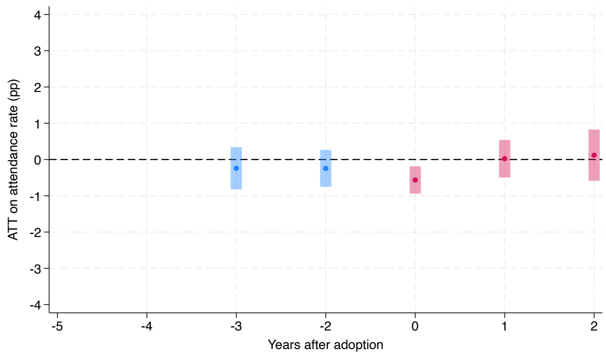
(b) G = 2024



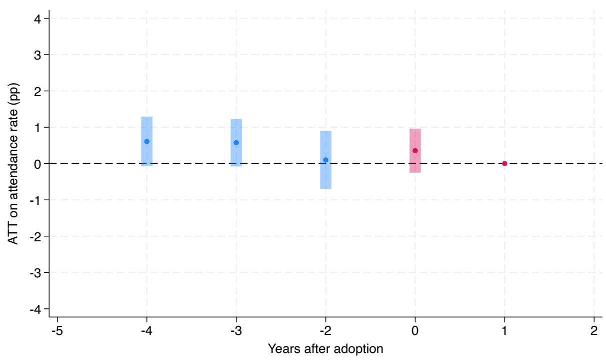
(c) G = 2025

Notes: This figure reports event-study estimates of the effect of Yondr adoption on discipline, measured in school-level standard deviations, for the pooled sample (across middle and high schools). Each panel corresponds to a cohort of schools first adopting Yondr in year g. Estimates are based on the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator. The plotted coefficients correspond to group-time average treatment effects, $\text{ATT}(g, t)$, for each cohort, normalized relative to the year immediately preceding adoption ($t = -1$). Never-treated schools serve as the comparison group. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level.

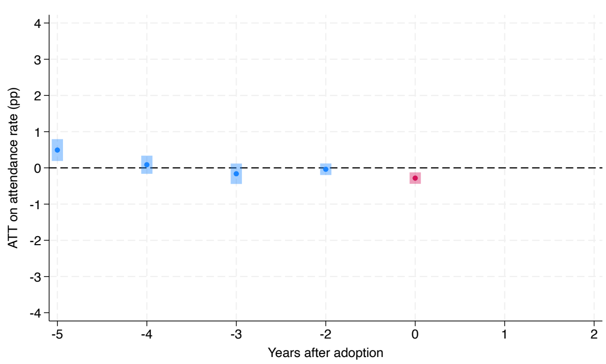
Figure A8: Event Study Estimates of Effects of Phone Pouches on Pooled High School and Middle School Attendance by Cohort



(a) G = 2023



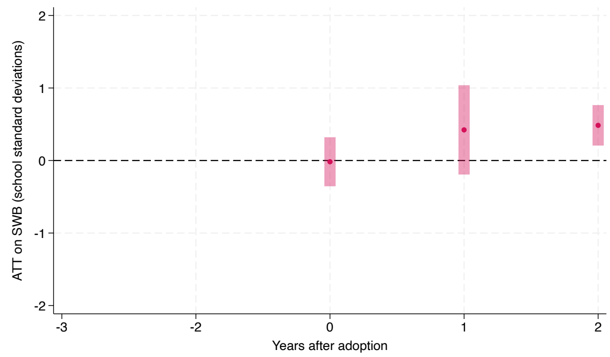
(b) G = 2024



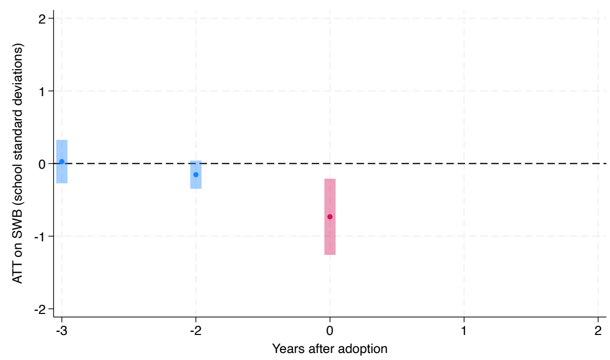
(c) G = 2025

Notes: This figure reports event-study estimates of the effect of Yondr adoption on attendance rates, measured in percentage points (0-100), for the pooled sample (across middle and high schools). Each panel corresponds to a cohort of schools first adopting Yondr in year g. Estimates are based on the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator. The plotted coefficients correspond to group-time average treatment effects, $\text{ATT}(g,t)$, for each cohort, normalized relative to the year immediately preceding adoption ($t=-1$). The estimate for the 2024 cohort in $t+1$ is not reported because the estimator does not return an $\text{ATT}(g,t)$ for this cell, plausibly due to limited support or insufficient comparison observations. Never-treated schools serve as the comparison group. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level.

Figure A9: Event Study Estimates of Effects of Phone Pouches on Pooled High School and Middle School Subjective Well-Being by Cohort



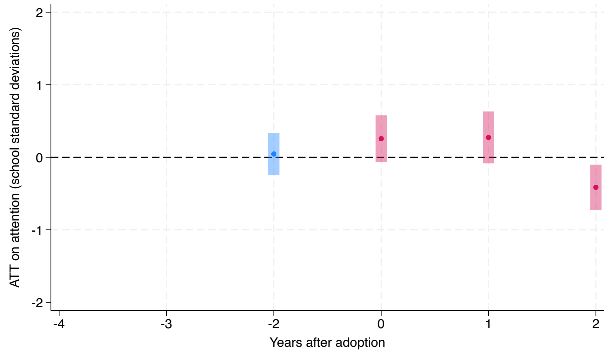
(a) G = 2023



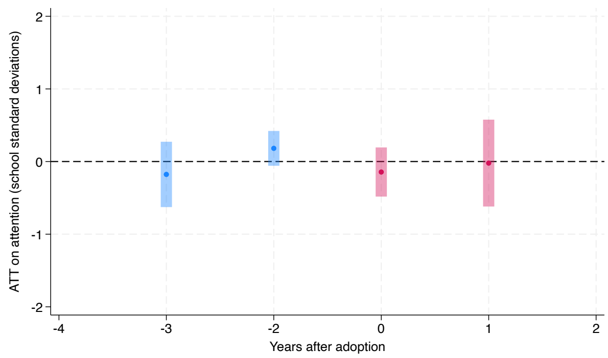
(b) G = 2025

Notes: This figure reports event-study estimates of the effect of Yondr adoption on subjective well-being, measured in school standard deviations, for the pooled sample (across middle and high schools). Each panel corresponds to a cohort of schools first adopting Yondr in year g. Estimates are based on the regression adjustment implementation of the Callaway and Sant'Anna (2021) difference-in-differences estimator. The plotted coefficients correspond to group-time average treatment effects, $\text{ATT}(g, t)$, for each cohort, normalized relative to the year immediately preceding adoption ($t = -1$). Never-treated schools serve as the comparison group. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level. The 2024 adoption cohort is omitted because no post-adoption observations are available for these schools in this sample.

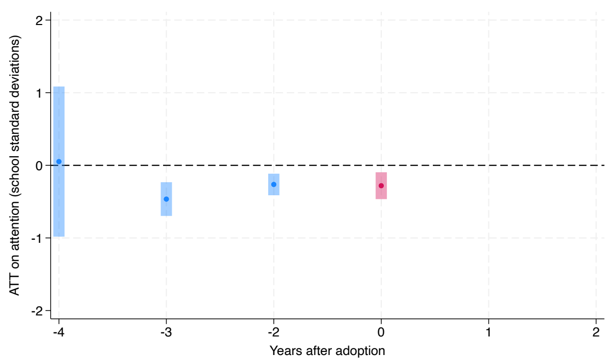
Figure A10: Event Study Estimates of Effects of Phone Pouches on Pooled High School and Middle School Classroom Attention by Cohort



(a) G = 2023



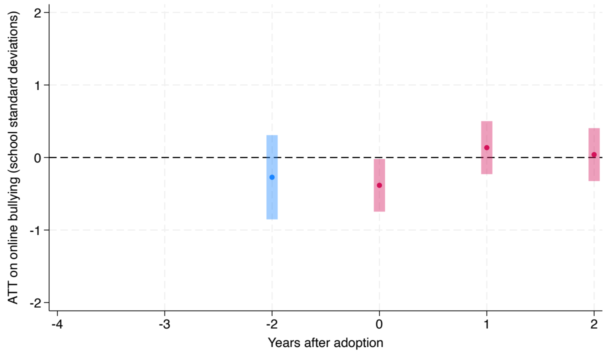
(b) G = 2024



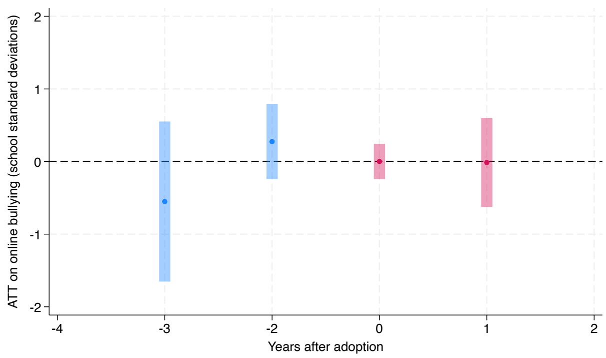
(c) G = 2025

Notes: This figure reports event-study estimates of the effect of Yondr adoption on classroom attention, measured in school standard deviations, for the pooled sample (across middle and high schools). Each panel corresponds to a cohort of schools first adopting Yondr in year g. Estimates are based on the regression adjustment implementation of the Callaway and Sant'Anna (2021) difference-in-differences estimator. The plotted coefficients correspond to group-time average treatment effects, ATT(g, t), for each cohort, normalized relative to the year immediately preceding adoption ($t = -1$). Never-treated schools serve as the comparison group. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level.

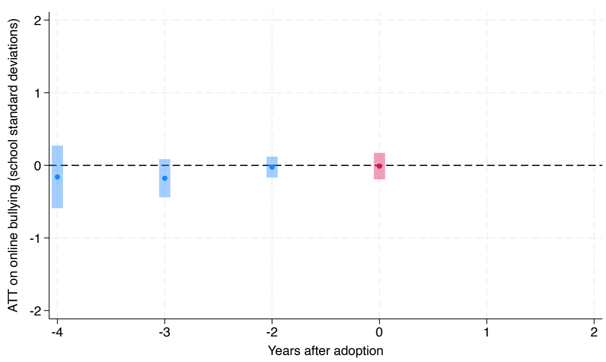
Figure A11: Event Study Estimates of Effects of Phone Pouches on Pooled High School and Middle School Online Bullying by Cohort



(a) G = 2023



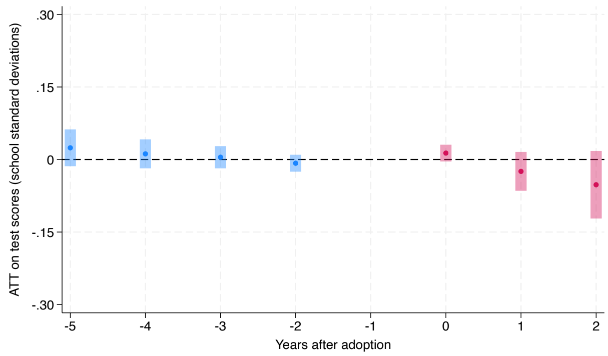
(b) G = 2024



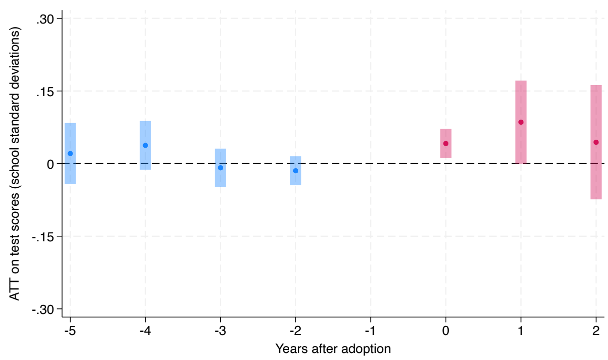
(c) G = 2025

Notes: This figure reports event-study estimates of the effect of Yondr adoption on online bullying, measured in school standard deviations, for the pooled sample (across middle and high schools). Each panel corresponds to a cohort of schools first adopting Yondr in year g. Estimates are based on the regression adjustment implementation of the Callaway and Sant'Anna (2021) difference-in-differences estimator. The plotted coefficients correspond to group-time average treatment effects, ATT(g, t), for each cohort, normalized relative to the year immediately preceding adoption ($t = -1$). Never-treated schools serve as the comparison group. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level.

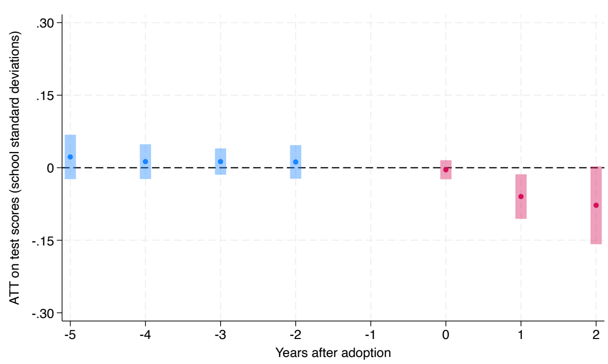
Figure A12: Event Study Estimates of Effects of Phone Pouches on Pooled, High School, and Middle School Math Test Scores



(a) Pooled



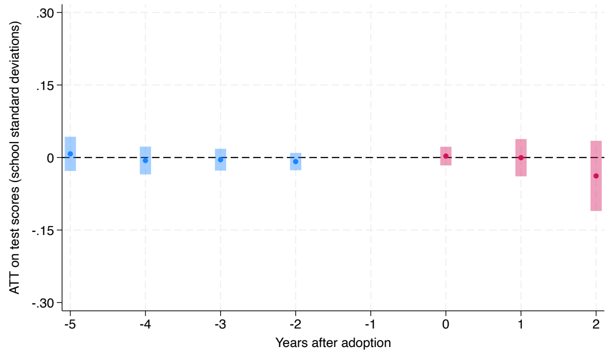
(b) High School



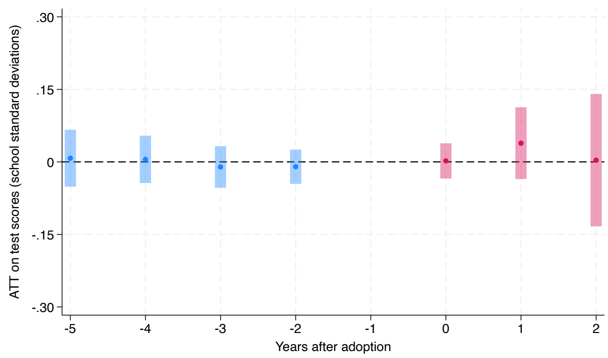
(c) Middle School

Notes: This figure reports event-study estimates of the effect of Yondr adoption on average Math test scores, measured in school-level standard deviations, for the pooled sample and separately for high schools and middle schools. 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. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level.

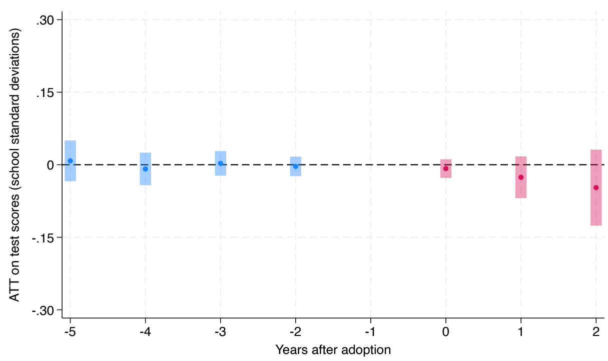
Figure A13: Event Study Estimates of Effects of Phone Pouches on Pooled, High School, and Middle School ELA Test Scores



(a) Pooled



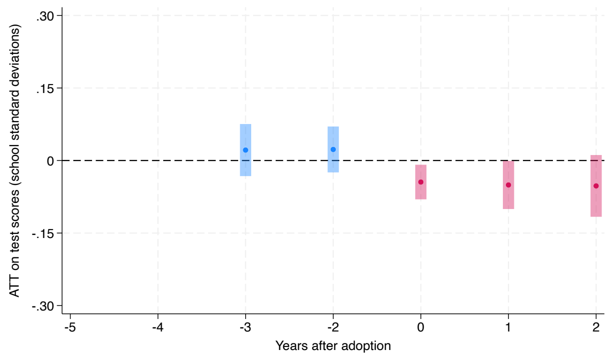
(b) High School



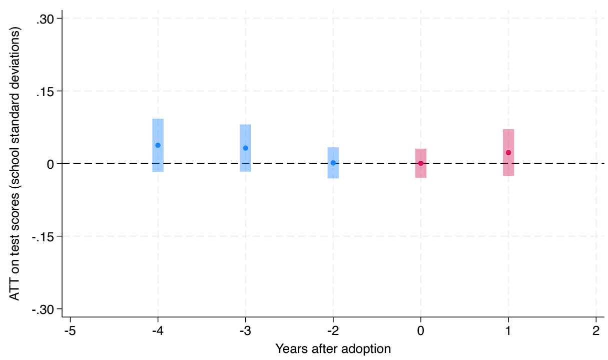
(c) Middle School

Notes: This figure reports event-study estimates of the effect of Yondr adoption on average ELA test scores, measured in school-level standard deviations, for the pooled sample and separately for high schools and middle schools. 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. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level.

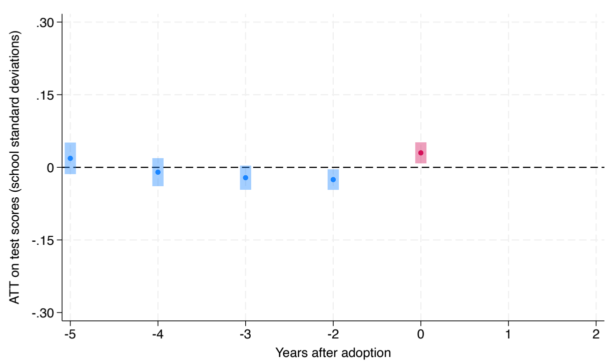
Figure A14: Event Study Estimates of Effects of Phone Pouches on Pooled High School and Middle School Test Scores by Cohort



(a) G = 2023



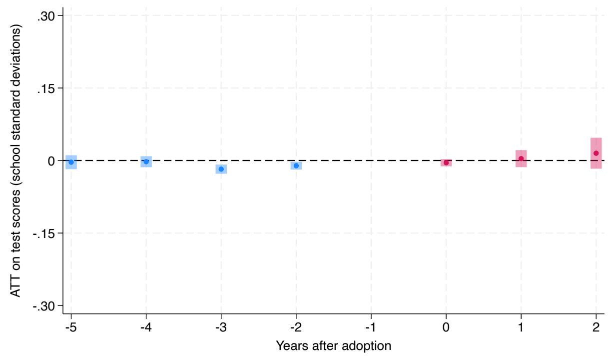
(b) G = 2024



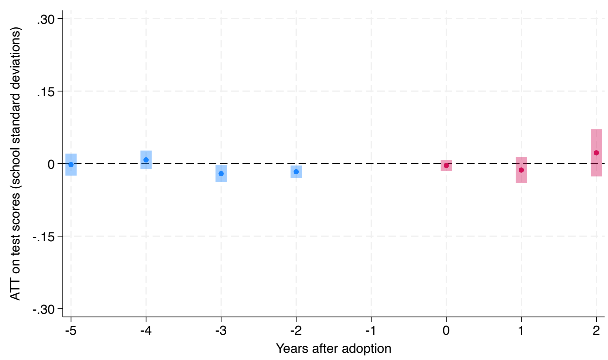
(c) G = 2025

Notes: This figure reports event-study estimates of the effect of Yondr adoption on average test scores, measured in school-level standard deviations, for the pooled sample (across middle and high schools). Each panel corresponds to a cohort of schools first adopting Yondr in year g. Estimates are based on the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator. The plotted coefficients correspond to group-time average treatment effects, ATT(g, t), for each cohort, normalized relative to the year immediately preceding adoption ($t = -1$). Never-treated schools serve as the comparison group. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level.

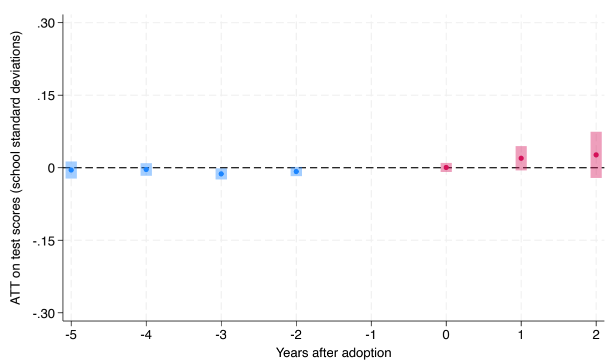
Figure A15: Event Study Estimates of Effects of Phone Pouches on Pooled, High School, and Middle School Predicted Test Scores



(a) Pooled



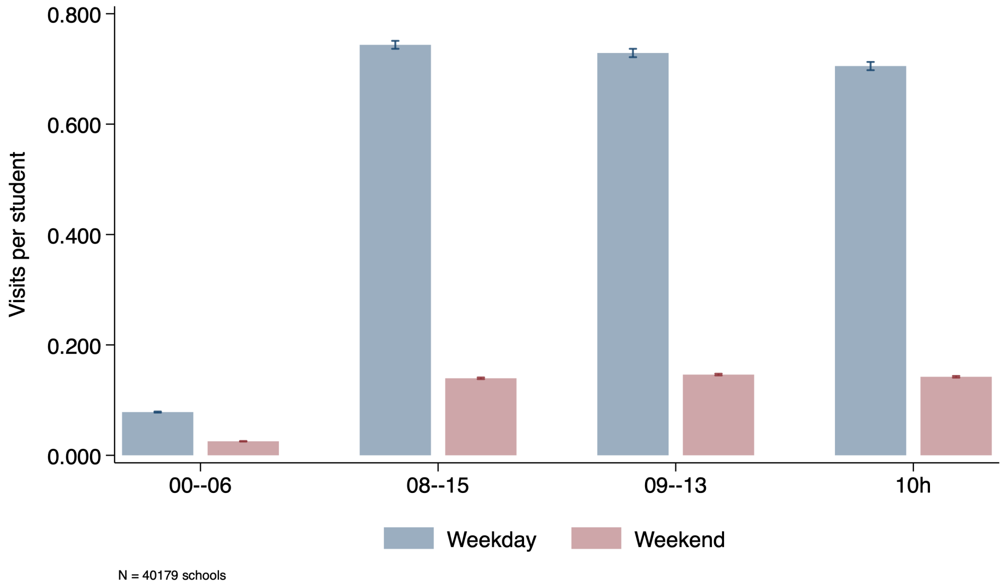
(b) High School



(c) Middle School

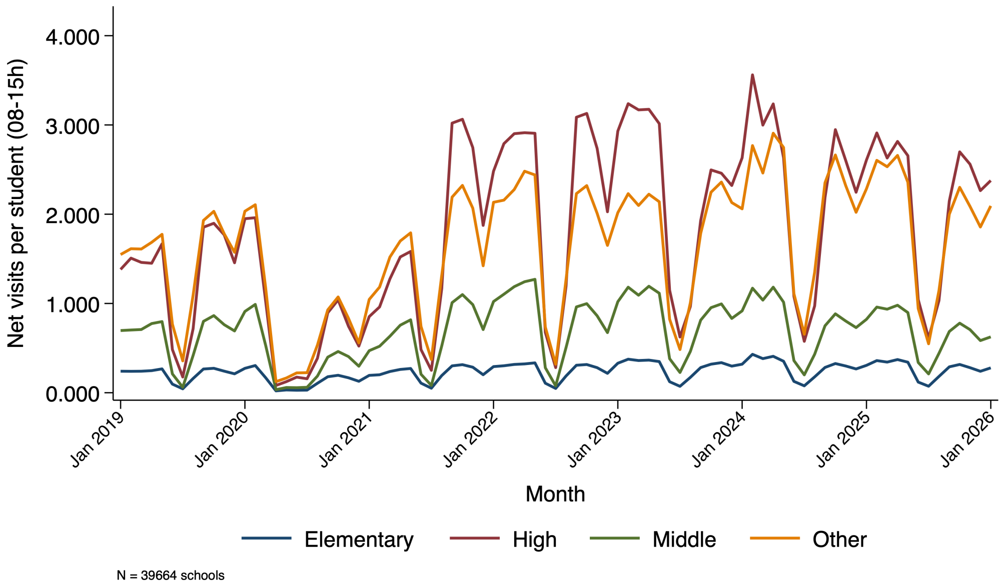
Notes: This figure reports event-study estimates of the effect of Yondr adoption on predicted test scores, measured in school-level standard deviations, for the pooled sample and separately by high schools and middle schools. The outcome is the fitted value from a regression of average test scores on the five time-varying compositional measures reported in Table A17. The overall ATT is -0.000 (SE = 0.005) in the pooled sample, -0.004 (SE = 0.007) in high schools, and 0.008 (SE = 0.007) in middle schools. Estimates are based on the Callaway and Sant'Anna (2021) improved doubly robust difference in differences estimator. The 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. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level.

Figure A16: Visits on Weekdays vs Weekends



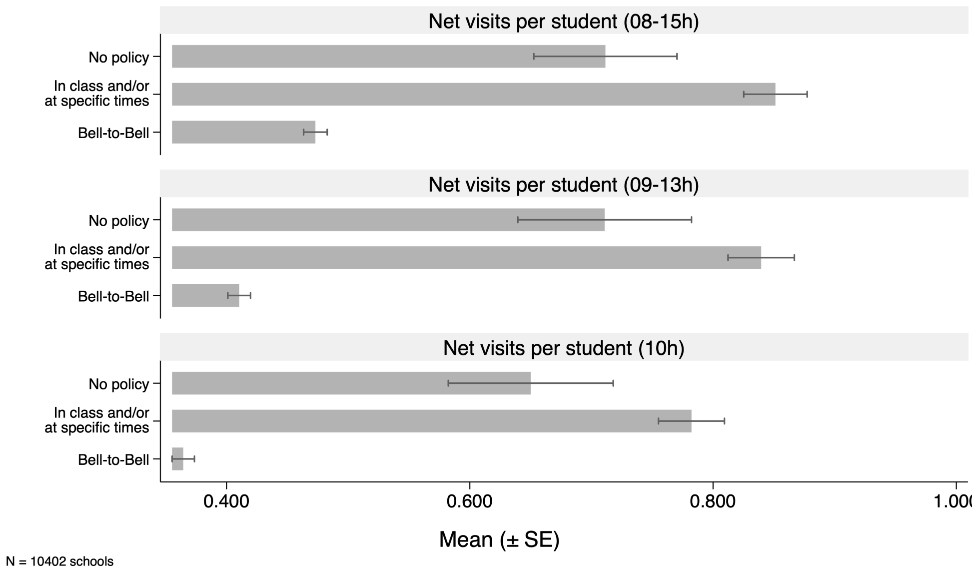
Notes: This figure compares phone visits during school-hour windows on weekdays and weekends using Advan GPS data. For each school, visits are measured as average device activity within school buildings during the indicated hour ranges, normalized by enrollment to obtain per-student values. The plotted 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 A17: Net Visits Time Series



Notes: The figure reports the average net visits by school level in the NCES using phone 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). We then compute a school-level average for each month and show the mean of the distribution of schools by level.

Figure A18: Net Visits by Phone Policy



Notes: The figure reports school-level averages constructed from the NTS and phone location data from Advan. Schools are grouped by the modal respose 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). We then compute a school-level average for each month and show 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.

B Additional Tables

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

	NCES (1)	NTS (2)	Panorama (3)
Total enrollment	595	832	756
Share White	0.53	0.57	0.35
Share Black	0.15	0.14	0.17
Share Hispanic	0.23	0.20	0.37
Urban	0.27	0.24	0.42
Suburban	0.28	0.32	0.36
Rural/Town	0.45	0.44	0.22
Charter school	0.14	0.06	0.03
Schools with grades above 9	0.36	0.52	0.35
Schools with grades below 6	0.39	0.19	0.38
Income-to-poverty ratio	300	326	324
Number of schools	57,711	20,842	7543

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, 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 Middle School Test Score Sources

State	Source	Coverage / exclusion note
Alabama	SEDA	—
Alaska	No data	No middle-school test source available.
Arizona	SEDA	—
Arkansas	SEDA	—
California	Zelma	—
Colorado	Zelma	—
Connecticut	Zelma	—
Delaware	Zelma	—
District of Columbia	SEDA	—
Florida	Zelma	—
Georgia	SEDA	—
Hawaii	SEDA	—
Idaho	SEDA	—
Illinois	SEDA	—
Indiana	SEDA	—
Iowa	SEDA	—
Kansas	SEDA	—
Kentucky	SEDA	—
Louisiana	Zelma	—
Maine	Excluded	Data for after 2020 not available
Maryland	SEDA	—
Massachusetts	Zelma	—
Michigan	Zelma	—
Minnesota	Zelma	—
Mississippi	SEDA	—
Missouri	SEDA	—
Montana	Excluded	No data for schools that adopted Yondr.
Nebraska	Zelma	—
Nevada	SEDA	—
New Hampshire	Zelma	—
New Jersey	Zelma	—
New Mexico	Excluded	Data for after 2020 not available
New York	Zelma	—
North Carolina	Zelma	—
North Dakota	SEDA	—
Ohio	SEDA	—
Oklahoma	Zelma	—
Oregon	Excluded	Data for after 2020 not available
Pennsylvania	SEDA	—
Rhode Island	Zelma	—
South Carolina	Zelma	—
South Dakota	Excluded	No data for schools that adopted Yondr.
Tennessee	SEDA	—
Texas	Zelma	—
Utah	SEDA	—
Vermont	Zelma	—
Virginia	Zelma	—
Washington	SEDA	—
West Virginia	SEDA	—
Wisconsin	Zelma	—
Wyoming	Zelma	—

Table A3: State-specific High School Test Score Sources

State	Source	Coverage / exclusion note
Alabama	No data	Data request declined.
Alaska	No data	Data request pending.
Arizona	No data	Data request pending.
Arkansas	No data	Data request pending.
California	Grade 11 EOG (Math & ELA)	—
Colorado	Grade 9–11 PSAT/SAT (Math & ELA)	—
Connecticut	Grade 11 SAT (Math & ELA)	2021 data not available.
Delaware	Grade 11 SAT (Math, ELA, Essay)	—
District of Columbia	No data	Data request declined.
Florida	HS EOC + Grade 9–10 ELA	—
Georgia	HS EOC (US Lit. & Algebra)	Algebra test contents changed in 2024.
Hawaii	Grade 11 EOG (Math & ELA)	—
Idaho	No data	Data request pending.
Illinois	Grade 11 SAT/ACT (Math & ELA)	State-mandated test was switched from SAT to ACT in 2025.
Indiana	Grade 10 EOG / Grade 11 SAT	Required test changed from ISTEP to SAT in 2022.
Iowa	Grade 9–11 EOG (Math & ELA)	Assessment has been administered since 2019.
Kansas	No data	Data request pending.
Kentucky	No data	Data request declined.
Louisiana	No data	Data request pending.
Maine	No data	Data request pending.
Maryland	No data	Data request pending.
Massachusetts	Grade 10 EOG (Math & ELA)	2018 data not available.
Michigan	Grade 11 SAT (Math & ELA)	—
Minnesota	Grade 10–11 EOG (Reading & Math)	—
Mississippi	No data	Data request pending.
Missouri	No data	Data request pending.
Montana	No data	Data request pending.
Nebraska	Grade 11 ACT (Math & ELA)	—
Nevada	No data	Data request declined.
New Hampshire	Grade 11 SAT (Math & Reading)	Missing 2025.
New Jersey	HS EOC + Grade 9 ELA	Assessment administered since 2019; 2021 data not available; the average scores include a small number of Grade 7 and 8 students.
New Mexico	No data	Data request pending.
New York	HS EOC (varied subjects)	—
North Carolina	HS EOC (Math & English)	—
North Dakota	No data	Data request declined.
Ohio	HS EOC (Algebra, Geometry, ELA)	—
Oklahoma	No data	Data request declined.
Oregon	No data	Data request pending.
Pennsylvania	No data	Data request pending.
Rhode Island	Grade 11 SAT (ELA & Math)	—
South Carolina	HS EOC (Algebra I & Eng II)	—
South Dakota	No data	The research team elected not to incur costs to acquire data.
Tennessee	No data	Data request pending.
Texas	HS EOC (Algebra, Eng I, Eng II)	—
Utah	No data	Data request declined.
Vermont	Grade 9 EOG (Math & ELA)	2018 data missing and requested.
Virginia	HS EOC (Algebra, Writing, Reading)	—
Washington	No data	Data request pending.
West Virginia	No data	Data request declined.
Wisconsin	Grade 11 ACT (Math & ELA)	—
Wyoming	Grade 9–10 EOG (Math & ELA)	—

Table A4: State-specific Attendance Rate Definitions and Availability

State	Used	Definition used	Coverage / exclusion note
Alabama	Yes	Aggregated days attended / aggregated days enrolled	Missing 2018-21 data; requested.
Alaska	Yes	Pending confirmation	Missing 2018-21 data; requested.
Arizona	No	—	Data request declined.
Arkansas	Yes	Aggregated days attended / aggregated days enrolled	—
California	Yes	Total days absent for any reason / (cumulative enrollment × average instructional days)	—
Colorado	Yes	Aggregated days attended / aggregated days enrolled	2018-19 data not available.
Connecticut	Yes	Aggregated days attended / aggregated days in membership	—
Delaware	Yes	Aggregated days attended / aggregated days enrolled	—
District of Columbia	No	—	Data request declined.
Florida	Yes	Aggregated days attended / aggregated days enrolled	—
Georgia	Yes	Aggregated days attended / aggregated days enrolled	—
Hawaii	Yes	1 - Aggregated days absence / aggregated days in membership	—
Idaho	No	—	The research team elected not to incur costs to acquire data.
Illinois	Yes	Aggregated days attended / aggregated days of potential attendance days	—
Indiana	Yes	Aggregated days attended / aggregated days enrolled	2022 data not available.
Iowa	Yes	Aggregated days attended / aggregated days enrolled	—
Kansas	Yes	Average daily attendance /average daily membership	—
Kentucky	Yes	Aggregated days attended / aggregated days enrolled	2021-22 data not available.
Louisiana	Yes	Pending confirmation	—
Maine	Yes	Aggregated days attended / aggregated days enrolled	—
Maryland	Yes	Aggregated days attended / aggregated days enrolled	—
Massachusetts	Yes	Pending confirmation	—
Michigan	Yes	Aggregated days attended / aggregated days enrolled	—
Minnesota	No	—	Data not available at school level.
Mississippi	No	—	Data request pending.
Missouri	Yes	Aggregated hours attended / aggregated hours possible	—
Montana	Yes	Aggregated days attended / aggregated days enrolled	—
Nebraska	Yes	Aggregated days attended / aggregated days enrolled	—
Nevada	Yes	Pending confirmation	—
New Hampshire	No	—	Data request declined.
New Jersey	No	—	Data request pending.
New Mexico	No	—	Data request pending.
New York	Yes	Aggregated days attended / aggregated days enrolled	Missing 2025 data.
North Carolina	No	Average daily attendance computed for each grade level, then averaged across grade levels	Only 2018 data available; requested.
North Dakota	Yes	Average daily attendance / average daily membership	—
Ohio	Yes	Aggregated hours attended / aggregated hours possible to attend	—
Oklahoma	No	—	Excluded since the measure is defined differently from other states.
Oregon	No	—	Data request pending.
Pennsylvania	Yes	Average daily attendance / average daily membership	Missing 2025 data.
Rhode Island	No	—	Data not available at school level.
South Carolina	No	—	Data not available at school level.
South Dakota	No	—	Data not available at school level.
Tennessee	No	—	Data request pending.
Texas	Yes	Aggregated days attended / aggregated days in membership	—
Utah	No	Aggregated days attended / aggregated days enrolled	Excluded due to few recent adopters.
Vermont	No	—	Data request pending.
Virginia	Yes	Aggregated days attended / aggregated days enrolled	—
Washington	No	—	Data not produced by the state.
West Virginia	No	—	Data request pending.
Wisconsin	Yes	Aggregated days attended / aggregated days enrolled	Missing 2025 data.
Wyoming	No	—	Data request pending.

Table A5: State-specific Chronic Absenteeism Definitions and Availability

State	Used	Definition used	Coverage / exclusion note
Alabama	Yes	Absent for 18 or more days for any reason	—
Alaska	Yes	Absent for 10% or more for any reason	Missing 2018-21 data; requested.
Arizona	No	—	Data request declined.
Arkansas	No	—	Data request pending.
California	Yes	Absent for 10% or more for any reason	—
Colorado	Yes	Absent for 10% or more for any reason	2018-23 data not available; requested.
Connecticut	Yes	Absent for 10% or more for any reason	—
Delaware	Yes	Absent for 10% or more for any reason	—
District of Columbia	No	—	Data request declined.
Florida	Yes	Absent for 10% or more for any reason	Missing 2025 data.
Georgia	Yes	Absent for 10% or more for any reason	—
Hawaii	No	—	Data request pending.
Idaho	No	—	The research team elected not to incur costs to acquire data.
Illinois	Yes	Absent for 10% or more for any reason	—
Indiana	Yes	Absent for 10% or more for any reason	—
Iowa	Yes	Absent for 10% or more for any reason	—
Kansas	No	—	Data request pending..
Kentucky	Yes	Absent for 10% or more for any reason	—
Louisiana	Yes	Absent for 10% or more for any reason	2018 data not available.
Maine	Yes	Absent for 10 % or more for any reason	—
Maryland	Yes	Absent for 10% or more for any reason	—
Massachusetts	Yes	Absent for 10% or more for any reason	—
Michigan	Yes	Absent for 10% or more for any reason	—
Minnesota	No	—	Data request pending.
Mississippi	Yes	Absent for 10% or more for any reason	—
Missouri	Yes	1 - % of students that attend at least 90 % of their hours.	—
Montana	Yes	Absent for 10% or more for any reason	Percentage computed using enrollment counts.
Nebraska	Yes	Absent for 10% or more for any reason	—
Nevada	Yes	Absent for 10% or more for any reason	—
New Hampshire	No	—	Data request declined.
New Jersey	Yes	Absent for 10% or more for any reason	Missing 2025 data.
New Mexico	No	—	Data request pending.
New York	Yes	Absent for 10% or more for any reason	Missing 2025 data.
North Carolina	Yes	Absent for 10% or more for any reason	Missing 2025 data.
North Dakota	Yes	Absent for 10% or more for any reason	—
Ohio	Yes	Absent for 10% or more for any reason	—
Oklahoma	Yes	Absent for 10% or more for any reason	2021 data not available.
Oregon	Yes	Absent for 10% or more for any reason	—
Pennsylvania	Yes	Absent for 10% or more for any reason	Missing 2025 data.
Rhode Island	Yes	Absent for 10% or more for any reason	Missing 2018-21 data; requested.
South Carolina	Yes	Absent for 10% or more for any reason	—
South Dakota	Yes	Absent for 10% or more for any reason	2022 data not available.
Tennessee	Yes	Absent for 10% or more for any reason	—
Texas	Yes	Absent for 10% or more for any reason	Missing 2025 data; 2018 and 19 data not available.
Utah	Yes	Absent for 10% or more for any reason	—
Vermont	No	—	Data request pending.
Virginia	Yes	Absent for 10% or more for any reason	—
Washington	No	—	Data request pending.
West Virginia	No	—	Data request pending.
Wisconsin	Yes	Absent for 10% or more for any reason	Missing 2025 data.
Wyoming	No	—	Data request declined.

Table A6: State-specific Disciplinary Measure Definitions and Availability

State	**Used**	**Definition used**	**Coverage / exclusion note**
Alabama	No	—	Data request declined.
Alaska	No	—	Data request declined.
Arizona	No	—	Available only for students with disabilities.
Arkansas	Yes	% students placed in-school suspension + % students placed out-of-school suspension	—
California	Yes	% students placed in any suspensions	—
Colorado	No	—	Data request declined.
Connecticut	Yes	% students placed in any suspensions	—
Delaware	Yes	One of the following: % students placed in-school suspension + % students placed out-of-school suspension without CDAP placement; % students placed in-school suspension; or % students placed out-of-school suspension without CDAP placement	—
District of Columbia	No	—	Data request declined.
Florida	Yes	One of the following: % students placed in-school suspension + % students placed out-of-school suspension; % students placed in-school suspension; or % students placed out-of-school suspension	Percentage computed using Fall enrollment counts.
Georgia	Yes	% students placed in-school suspension + % students placed out-of-school suspension	—
Hawaii	No	—	Pending state confirmation of suspension definitions.
Idaho	No	—	The research team elected not to incur costs to acquire data.
Illinois	Yes	% students with any disciplinary measures taken	Missing 2018-22 data and requested; percentage computed using October enrollment counts.
Indiana	Yes	One of the following: % students placed in-school suspension + % students placed out-of-school suspension; % students placed in-school suspension; or % students placed out-of-school suspension	—
Iowa	Yes	% students placed in any suspensions	Percentage computed using enrollment counts.
Kansas	No	—	Data request pending.
Kentucky	No	—	Data request pending.
Louisiana	Yes	% students placed in-school suspension + % students placed out-of-school suspension	—
Maine	No	—	Data request pending.
Maryland	No	—	Data request pending.
Massachusetts	Yes	% students with any disciplinary measures taken	—
Michigan	No	—	Data not available at school level.
Minnesota	Yes	% students placed in any suspensions	—
Mississippi	Yes	One of the following: % students placed in-school suspension + % students placed out-of-school suspension; % students placed in-school suspension; or % students placed out-of-school suspension	Missing 2025 data.
Missouri	No	—	Majority of observations suppressed.
Montana	No	—	Data request pending.
Nebraska	No	—	—
Nevada	No	—	Data request declined.
New Hampshire	No	—	Excluded due to the insufficient number of Yondr adopters.
New Jersey	Yes	One of the following: % students placed in any suspensions; % students placed in-school suspension; or % students placed out-of-school suspension	Missing 2025 data.
New Mexico	No	—	Data request declined.
New York	Yes	% students placed out-of-school suspension	Missing 2025 data; percentage computed using enrollment counts.
North Carolina	No	—	Data request pending.
North Dakota	No	—	Data not available at school level.
Ohio	Yes	% students with any disciplinary measures taken	—
Oklahoma	No	—	Data request declined.
Oregon	No	—	Data request pending.
Pennsylvania	Yes	% students placed in any suspensions	Percentage computed using October enrollment counts.
Rhode Island	Yes	% students placed out-of-school suspension	Missing 2018-21 data; requested.
South Carolina	No	—	Pending data correction by the state.
South Dakota	No	—	The research team elected not to incur costs to acquire data.
Tennessee	Yes	2018–2019: % students placed in-school suspension + % students placed out-of-school suspension; 2020–2025: % students with any disciplinary measures taken	—
Texas	Yes	One of the following: % students with any disciplinary measures taken; % students placed in-school suspension + % students placed out-of-school suspension; % students placed in-school suspension; or % students placed out-of-school suspension	Percentage computed using enrollment counts.
Utah	No	—	Data request declined.
Vermont	No	—	Majority of observations suppressed.
Virginia	Yes	2018–2019: Incident counts of in-school + out-of-school suspension; 2022-25: One of % students placed in-school suspension + % students placed out-of-school suspension; % students placed in-school suspension; or % students placed out-of-school suspension	Missing 2025 data; 2021 data not available percentage computed using enrollment counts.
Washington	No	—	Data request pending.
West Virginia	No	—	Data request pending.
Wisconsin	Yes	% students placed out-of-school suspension	Missing 2025 data; percentage computed using enrollment counts.
Wyoming	No	—	Data not available at school level.

Table A7: Gallup Survey Demographcis

Category	Variable / Level	Mean (SD)
Parent	Age	46.9 (8.3)
	Male	596 (38%)
Parent Gender	Female	962 (61%)
	Non-binary	7 (0.4%)
	White	954 (61%)
	Black	254 (16%)
Parent Race	Hispanic	227 (15%)
	Asian	94 (6.0%)
	Other	31 (2.0%)
	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%)
Household Income	$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)
	Male	815 (52%)
Child Gender	Female	725 (46%)
	Nonbinary	25 (1.6%)
	White	824 (53%)
	Black	284 (18%)
Child Race	Hispanic	319 (20%)
	Asian	116 (7.4%)
	Other	22 (1.4%)
	5th Grade	1 (<0.1%)
	6th Grade	34 (2.2%)
	7th Grade	163 (10%)
	8th Grade	256 (16%)
Child Grade	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%)
	Public school	1,125 (75%)
School Type	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 A8: Prior Phone Policies Among Yondr Adopters

Prior policy	Percent
No show	62.01
No policy	23.16
Class collection	7.36
Locker	6.17
Centralized collection	0.97
No phone	0.32
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.

Table A9: Comparison of All Yondr Adopters versus Yondr Adopters in NTS

	Non-adopters (1)	Adopters (2)	Adopters in NTS (3)
Total enrollment	591	681	821
Share White	0.54	0.35	0.44
Share Black	0.14	0.27	0.22
Share Hispanic	0.23	0.29	0.24
Urban	0.25	0.49	0.42
Suburban	0.28	0.21	0.25
Rural/Town	0.46	0.30	0.33
Charter school	0.14	0.13	0.08
Schools with grades above 9	0.35	0.49	0.56
Schools with grades below 6	0.39	0.27	0.16
Income-to-poverty ratio	301	280	314
Number of schools	54,736	2975	1025

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 A10: Difference-in-Differences Estimates of Effects of Phone Pouches on Pooled, High School, and Middle School Phone GPS Activty

	Net visits 10h (1)	Net visits 09-13h (2)	Net visits 08-15h (3)
Panel A: Pooled			
ATT	-0.189^{***}	-0.182^{***}	-0.182^{***}
	(0.054)	(0.050)	(0.067)
Observations	85,928	88,455	89,726
Schools	18579	18591	18608
Panel B: Middle schools			
ATT	-0.140^{*}	-0.122^{*}	-0.103
	(0.076)	(0.068)	(0.064)
Observations	62,397	64,699	65,860
Schools	13675	13686	13700
Panel C: High schools			
ATT	-0.187^{**}	-0.243^{***}	-0.194^{***}
	(0.080)	(0.079)	(0.073)
Observations	23,531	23,756	23,866
Schools	4904	4905	4908

Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption on phone activity 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 per student, 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), divided by enrollment. Never-treated schools serve as the comparison group. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. 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 Phone GPS Activty, by Sample

	Net visits 10h (1)	Net visits 09-13h (2)	Net visits 08-15h (3)
Panel A: Middle School Test Score Sample			
ATT	-0.147^{*}	-0.180^{**}	-0.162^{**}
	(0.080)	(0.073)	(0.068)
Observations	54,522	56,009	56,749
Schools	13593	13607	13628
Panel B: High School Test Score Sample			
ATT	-0.244^{**}	-0.285^{***}	-0.202^{**}
	(0.111)	(0.100)	(0.086)
Observations	14,289	14,390	14,437
Schools	3123	3123	3125
Panel C: Attendance Sample			
ATT	-0.187^{***}	-0.172^{***}	-0.151^{***}
	(0.062)	(0.057)	(0.054)
Observations	59,859	61,521	62,300
Schools	13800	13807	13812
Panel D: Discipline Sample			
ATT	-0.249^{***}	-0.333^{***}	-0.180^{***}
	(0.069)	(0.087)	(0.059)
Observations	47,762	49,102	49,730
Schools	11485	11497	11506

Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption on phone activity 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 per student, 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), divided by enrollment. Never-treated schools serve as the comparison group. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level. Each panel subsets the sample to schools used in the analyses for that specific outcome. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A12: 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	0.197	−0.100	−0.086
	(0.242)	(0.099)	(0.099)
Mean of dependent variable	22.41	93.98	94.04
Observations	224,322	34,174	34,174

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 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. Never-treated schools serve as the comparison group. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A13: Difference-in-Differences Estimates of Effects of Phone Pouches on Pooled, High School, and Middle School Discipline, Attendance, and Student Survey Outcomes

	Discipline index (1)	Attendance rate (2)	SWB index (3)	Attention index (4)	Online bullying (5)
Panel A: Pooled					
ATT	0.059**	−0.138	−0.305	−0.088	−0.039
	(0.024)	(0.099)	(0.201)	(0.085)	(0.092)
Mean of dependent variable	-	92.62	-	-	-
Observations	153,968	201,483	3382	9247	5648
Panel B: High School					
ATT	0.073**	−0.244	−0.156	−0.348**	−0.067
	(0.037)	(0.161)	(0.204)	(0.169)	(0.191)
Mean of dependent variable	-	91.54	-	-	-
Observations	58,666	76,651	1041	2606	1680
Panel C: Middle School					
ATT	0.053	0.107	−0.177	−0.030	0.069
	(0.034)	(0.207)	(0.239)	(0.090)	(0.116)
Mean of dependent variable	-	93.28	-	-	-
Observations	95,302	124,325	2332	6618	3937

Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption. Columns (1) and (2) use the Callaway and Sant'Anna (2021) improved doubly robust difference-in-differences estimator, whereas Columns (3)–(5) use the regression adjustment implementation of the estimator. Column (1) reports effects on a standardized school-level discipline index constructed from suspension and discipline measures and standardized within state-by-year cells. Column (2) reports effects on the attendance rate. 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 (2) are in percentage points, and the estimates in all other columns are in school-level standard deviations. Never-treated schools serve as the comparison group. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A14: Robustness Checks for Effects of Phone Pouches on Test Scores

	Pooled		High School		Middle School	
Specification	ATT	N	ATT	N	ATT	N
1. Baseline	−0.004	214,556	0.025	65,978	−0.024**	162,974
	(0.010)		(0.017)		(0.012)	
2. Regression adjustment only	−0.006	214,556	0.026	65,978	−0.025**	162,974
	(0.010)		(0.017)		(0.012)	
3. NTS-matched controls (non-Yondr)	0.009	84,580	0.042**	37,960	−0.016	52,425
	(0.012)		(0.019)		(0.015)	
4. Exclude strict policy change from control	0.003	71,321	0.041**	33,548	−0.029*	42,271
	(0.013)		(0.021)		(0.015)	
5. Exclude any policy change from control	−0.001	30,252	0.026	10,669	−0.023	21,518
	(0.017)		(0.028)		(0.017)	
6. Exclude any strict policy from control	0.006	64,742	0.055**	32,806	−0.038**	36,212
	(0.015)		(0.022)		(0.016)	
7. Not yet treated as controls	0.013	13,485	0.022	5586	0.002	9462
	(0.015)		(0.026)		(0.015)	
8. Exclude 2026 adopters	−0.005	210,389	0.027	63,813	−0.026**	159,534
	(0.010)		(0.018)		(0.012)	
9. High confidence schools only	−0.005	212,675	0.026	65,162	−0.024**	161,723
	(0.010)		(0.018)		(0.012)	
10. Exclude canceled adopters	−0.003	213,291	0.029*	65,468	−0.025**	162,129
	(0.011)		(0.017)		(0.013)	
11. Balanced panel	0.002	136,716	0.035*	54,000	−0.021	90,348
	(0.012)		(0.021)		(0.014)	
12. Controlling for number of students tested	−0.003	214,556	0.027	65,978	−0.024**	162,974
	(0.010)		(0.017)		(0.012)	
13. No imputed test scores	−0.001	207,650	0.030*	63,948	−0.021*	156,768
	(0.010)		(0.017)		(0.012)	
14. Control for state × urbanicity × grade levels	−0.008	214,556	-	-	-	-
	(0.010)		-		-	
15. Stacked combined schools	−0.000	229,009	-	-	-	-
	(0.010)		-		-	

Notes: The table reports estimates of the average treatment effect on the treated (ATT) of Yondr adoption on standardized test scores. The baseline specification uses the Callaway and Sant'Anna (2021) doubly robust difference-in-differences estimator with an unbalanced panel and never-treated schools as controls. The second row reports estimates using the Callaway and Sant'Anna (2021) estimator with regression adjustment only. Rows 3–6 modify the control group using information from the NTS. Row 3 restricts to schools observed in the NTS whose reported policy is not Yondr. Row 4 further restricts to schools reporting no policy change or only weak policy changes. Row 5 restricts to schools reporting no policy change. Row 6 excludes schools that implement strict alternative phone policies. Rows 7–8 modify the comparison group based on treatment timing. Row 7 uses only not-yet-treated schools as controls, and row 8 excludes future adopters (the 2026 cohort) from the control group. Rows 9–10 restrict the treated sample by limiting to high-confidence adopters and excluding schools with canceled accounts. Rows 11–13 restrict to a balanced panel, control for the number of students tested, and exclude imputed test scores. Row 14 augments the baseline controls by allowing the state-by-urbanicity indicators to vary by school level (middle versus high). Row 15 constructs a "stacked" dataset in which the relatively small share of schools that appear in both the middle school and high school test score sample are duplicated and treated as separate school-by-year-by-level observations, allowing both middle and high school test scores to enter the pooled analysis. Rows 3–7 are estimated using the regression-adjustment implementation of the Callaway and Sant'Anna (2021) estimator rather than the doubly robust version used in our primary specifications. With the substantially smaller sample, the doubly robust estimator fails to compute some cohort-by-year ATTs, whereas the outcome-regression-only implementation is well behaved. As shown in the table, the two approaches yield nearly identical estimates in settings where both can be implemented. 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 A15: Robustness Checks for Effects of Phone Pouches on Math Test Scores

	Pooled		High School		Middle School	
Specification	ATT	N	ATT	N	ATT	N
1. Baseline	−0.003	210,951	0.048***	64,502	−0.027*	159,048
	(0.012)		(0.018)		(0.014)	
2. Regression adjustment only	−0.002	210,951	0.047**	64,585	−0.027**	159,775
	(0.011)		(0.018)		(0.013)	
3. NTS-matched controls (non-Yondr)	0.011	83,091	0.063***	37,277	−0.012	51,106
	(0.013)		(0.021)		(0.017)	
4. Exclude strict policy change from control	0.005	70,112	0.061***	32,920	−0.029	41,335
	(0.014)		(0.022)		(0.018)	
5. Exclude any policy change from control	0.002	29,718	0.058*	10,439	−0.026	21,066
	(0.018)		(0.031)		(0.020)	
6. Exclude any strict policy from control	0.003	63,589	0.063***	32,184	−0.041*	35,337
	(0.016)		(0.023)		(0.022)	
7. Not yet treated as controls	0.005	13,164	0.047*	5409	−0.005	9204
	(0.015)		(0.026)		(0.017)	
8. Exclude 2026 adopters	−0.004	206,862	0.049**	62,489	−0.028**	156,770
	(0.012)		(0.019)		(0.014)	
9. High confidence schools only	−0.003	209,096	0.043**	63,649	−0.027*	158,529
	(0.012)		(0.020)		(0.014)	
10. Exclude canceled adopters	−0.002	209,741	0.056***	64,098	−0.027*	158,971
	(0.012)		(0.020)		(0.014)	
11. Balanced panel	0.003	135,638	0.063***	52,851	−0.029*	90,144
	(0.013)		(0.021)		(0.016)	
12. Controlling for number of students tested	−0.002	210,951	0.047**	64,453	−0.027**	159,775
	(0.011)		(0.019)		(0.014)	
13. Control for state × urbanicity × grade levels	−0.006	210,951	-	-	-	-
	(0.012)		-		-	
14. Stacked combined schools	0.003	224,449	-	-	-	-
	(0.012)		-		-	

Notes: The table reports estimates of the average treatment effect on the treated (ATT) of Yondr adoption on Mathematics test scores. The baseline specification uses the Callaway and Sant'Anna (2021) doubly robust difference-in-differences estimator with an unbalanced panel and never-treated schools as controls. The second row reports estimates using the Callaway and Sant'Anna (2021) estimator with regression adjustment only. Rows 3–6 modify the control group using information from the NTS. Row 3 restricts to schools observed in the NTS whose reported policy is not Yondr. Row 4 further restricts to schools reporting no policy change or only weak policy changes. Row 5 restricts to schools reporting no policy change. Row 6 excludes schools that implement strict alternative phone policies. Rows 7–8 modify the comparison group based on treatment timing. Row 7 uses only not-yet-treated schools as controls, and row 8 excludes future adopters (the 2026 cohort) from the control group. Rows 9–10 restrict the treated sample by limiting to high-confidence adopters and excluding schools with canceled accounts. Rows 11–12 address additional sample and measurement considerations by restricting to a balanced panel and controlling for the number of students tested. Rows 13–14 introduce additional robustness checks for the pooled specification. Row 13 augments the baseline controls by allowing the state-by-urbanicity indicators to vary by school level (middle versus high). Row 14 constructs a "stacked" dataset in which the relatively small share of schools that appear in both the middle school and high school test score sample are duplicated and treated as separate school-by-year-by-level observations, allowing both middle and high school test scores to enter the pooled analysis. Rows 3–7 are estimated using the regression-adjustment implementation of the Callaway and Sant'Anna (2021) estimator rather than the doubly robust version used in our primary specifications. With the substantially smaller sample, the doubly robust estimator fails to compute some cohort-by-year ATTs, whereas the outcome-regression-only implementation is well behaved. As shown in the table, the two approaches yield nearly identical estimates in settings where both can be implemented. 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 A16: Difference-in-Differences Estimates with Incremental Baseline Controls

	Minimum controls (1)	Demographic controls (2)	Grade level controls (3)	Charter status (4)
Panel A: Pooled				
ATT	0.007	−0.005	−0.004	−0.004
	(0.010)	(0.010)	(0.010)	(0.010)
Observations	214,556	214,556	214,556	214,556
Panel B: High School Only				
ATT	0.042**	0.026	0.024	0.025
	(0.016)	(0.017)	(0.017)	(0.017)
Observations	65,978	65,978	65,978	65,978
Panel C: Middle School Only				
ATT	−0.013	−0.025**	−0.023**	−0.024**
	(0.012)	(0.012)	(0.012)	(0.012)
Observations	162,974	162,974	162,974	162,974

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 A17: Tests of Compositional Effects

	Total enrollment (1)	Share White (2)	Share Black (3)	Share Hispanic (4)	Share FRPL (5)
Panel A: Pooled					
ATT	-6.996^{**}	0.004^{***}	0.006	0.006^{**}	-0.002
	(2.870)	(0.001)	(0.005)	(0.003)	(0.002)
Mean of dependent variable	662	0.536	0.163	0.267	0.550
Observations	214,276	210,489	202,436	210,253	190,451
Panel B: High School Only					
ATT	-7.650^{**}	0.005^{**}	0.004	0.004	0.002
	(3.432)	(0.002)	(0.009)	(0.004)	(0.004)
Mean of dependent variable	922	0.558	0.153	0.236	0.514
Observations	65,978	65,042	62,997	64,645	58,539
Panel C: Middle School Only					
ATT	-5.539	0.003^{**}	-0.001	0.006^{*}	-0.007^{**}
	(3.631)	(0.001)	(0.005)	(0.003)	(0.003)
Mean of dependent variable	549	0.534	0.163	0.273	0.562
Observations	162,966	159,548	152,890	159,466	144,515

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) improved doubly robust 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. Never-treated schools serve as the comparison group. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A18: Heterogeneous Effects of Phone Pouches on Pooled, High School and 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)
Panel A: Pooled									
ATT	0.006	-0.025	-	-	-	-	-	-	-
	(0.013)	(0.018)							
Observations	108,018	106,538	-	-	-	-	-	-	-
Panel B: High School Only									
ATT	0.043^{**}	0.022	-	-	-	-	-	-	-
	(0.021)	(0.026)							
Observations	33,950	31,903	-	-	-	-	-	-	-
Panel C: Middle School Only									
ATT	-0.012	-0.049^{**}	-0.021	-0.020	-0.006	-0.044^{*}	-0.032^{*}	-0.010	-0.042^{**}
	(0.014)	(0.022)	(0.014)	(0.014)	(0.023)	(0.026)	(0.018)	(0.016)	(0.021)
Observations	82,745	80,229	123,447	122,889	50,016	105,742	82,913	130,054	112,746

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. Never-treated schools serve as the comparison group. All specifications use an unbalanced panel. Covariates and weighting follow the main empirical specification. Standard errors are clustered at the school level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A19: 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 high- and highest-confidence schools.

Table A20: Construction of Analysis Samples Across Outcomes

	MS test score (1)	HS test score (2)	Attendance (3)	Chronic abs. (4)	Discipline (5)
NCES Panel	806,853	806,853	806,853	806,853	806,853
Initial outcome panel	183,658	95,532	489,862	539,742	365,798
Restrict to middle/high school grades	178,886	95,295	297,919	326,878	233,953
Drop 2020 and 2021	178,886	81,770	230,903	254,446	178,734
Drop pre-2023 adopters	177,257	80,034	228,050	251,443	176,374
Drop contaminated controls	173,186	78,353	223,045	245,978	171,573
Drop low-confidence schools	172,378	77,743	221,790	244,661	170,681
Keep states with 2023–2025 adoption	172,378	75,761	220,776	244,661	170,681
Keep only regular schools	169,163	65,971	203,780	227,774	159,630

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 NCES defines "Regular Schools" as a public school providing instruction and education services that does not focus primarily on special education, career/technical education, alternative education, or on any of the particular themes associated with magnet/special program-emphasis schools.

Table A21: Construction of Panorama Analysis Samples

	Subjective well-being (1)	Classroom attention (2)	Online bullying (3)
NCES Panel (2019–2025)	907,430	907,430	907,430
Drop 2020 and 2021	504,993	504,993	504,993
Restrict to middle/high school grades	289,144	289,144	289,144
Require any Panorama index	21,317	21,317	21,317
Require outcome-specific index	5917	14,804	9718
Drop pre-2023 adopters	5834	14,668	9600
Drop contaminated controls	5732	14,162	9318
Drop low-confidence schools	5709	14,122	9286
Keep states with 2023–2025 adoption	4059	11,087	6705
Restrict to regular schools	3742	10399	6258

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 A19 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. Advan relies on opt-in location sharing through a panel of mobile applications. Thus, the dataset should not be interpreted as observing every phone in the population; instead, it 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.

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.[13]

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.

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, 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

[13] For examples of previous work that has used Advan, see Cook et al. (2025); Figlio and Ozek (2025).

filtering. At the same time, it is not plausible to assume the data exclude minors entirely; the dataset reflects 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.

Data processing. We use 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.

Validation exercises. To assess whether the constructed measures contain meaningful signal, we present three sets of validation exercises. First, we examine whether observed activity is concentrated at times when students are plausibly present and phones are plausibly in use. Figure A16 summarizes aggregate activity over time and shows 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 captures school-related device presence and activity rather than purely mechanical background location signals.

A second validation check, shown in Figure A17, 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 predictions: older students are more likely to own phones, carry them throughout the school day, and generate usage-related location events.

A final validation check links GPS outcomes to independent, survey-based measures of phone use and policy strictness from NTS. Figure A18 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 phone 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 phone activity. For the 8:00–15:00 measure, the correlation is 0.147 ($p = < 0.001$); for the 9:00–13:00 measure, the correlation is 0.161 ($p = < 0.001$); and for the 10:00 single-hour measure, the correlation is 0.161 ($p = < 0.001$). Together, these results provide external validation for the GPS-based measures, showing that they covary with independently measured indicators of phone use in expected ways.

Additional results. Finally, we present additional evidence on the effect of Yondr adoption on GPS-measured phone activity. The main text focuses on the outcome defined using the 10:00–11:00 school-hours window and the netting procedure described above. The results in Tables A10 and A11 show that the main findings are not sensitive to reasonable alternative choices. Across definitions, including other time-windows for school-hours and other samples of schools that match our main analyses (for test scores, attendance, and discipline outcomes), we find the same qualitative pattern of a sustained decline in phone activity after adoption.[14]

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.

[14] We cannot conduct an analogous robustness exercise for the Panorama sample because, as described in Section 3, Panorama matched school identifiers to survey responses prior to anonymization, preventing linkage to additional school-level samples.

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.

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 Details on Gallup Questions and Data

This section presents more detail about the Gallup data used in the analysis.

Overview. Gallup maintains a probability-based panel representative of the U.S. population, from which it draws samples for several studies. The data used in this paper is part of a broader initiative we conducted jointly with Gallup which invited all Gen Z (born between 1997 and 2012) members of the Gallup Panel, as well as members identified as parents of children ages 12-18, to participate in four survey waves conducted between 2025 and 2026. The data used in this paper consist of parent/child dyad responses collected during Wave 3, fielded between February 10 and February 17, 2026. Parents were recruited via email and SMS with a promised postpaid incentive of $2 for themselves and an $8 incentive for their child (standard values for Gallup's panels), and completed the web survey online.

A total of 1,801 eligible parent/child dyads completed this survey wave: 236 who had not responded to the previous two waves and 1,565 who had. This corresponds to an overall response rate of 35.45% (9.80% among new respondents; 66.01% among returning respondents). We focus on 1,399 children who reported studying in public or charter schools; we exclude 159 who attend private schools, 165 who are homeschooled, and 78 who did not report what kind of school they attend.

Sample weighting. Gallup provides two separate survey weights: for youth and 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 ages 12–29 or of parents of youth ages 12–18.

Question wording. 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

School phone policies

W37. Officially, when is the personal use of phones (e.g., for texting, social media) restricted?

- Students are prohibited from using their phones during the entire school day, including during lunch and between classes.
- Students can use phones sometimes, like during lunch or between classes.
- There is no school-wide restriction. Individual teachers and staff decide when students can use phones for their class.

W38. Officially, where are students allowed to keep their phones?

- Students must leave their phones at home.
- Students must deposit their phones in one place, like the main office, at the beginning of the school day.
- Students must keep their phones in their lockers all day.
- Students must keep their phones in a lockable pouch (like Yonder).
- Students must deposit their phones at the beginning of every class.
- Students can keep their phones in backpacks, pockets, etc., but phones must be out of sight ('no-show').
- There is no school-wide policy. Individual teachers and staff decide where students can keep phones.

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 A20 summarizes the construction of each analysis sample and reports how each sample restriction affects the number of school–year observations.

Middle-School Test Score Analysis Sample

The initial middle-school test score panel combines data from the EDC and SEDA and includes 183,658 school–year observations with non-missing English Language Arts or mathematics test scores averaged across grades 6 through 8. The initial sample covers 39,133 unique public schools 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—those 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. Finally, we retain only regular schools in the sample.[15]

After applying these restrictions, the middle school test score analysis sample consists of 169,163 school–year observations, representing 35,323 unique schools across 44 states and the District of Columbia.[16] The baseline specifications throughout the paper 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. 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 15,058 schools, representing approximately 43 percent of the baseline test score sample. Estimated effects in the balanced-panel specifications are similar in magnitude and direction to the baseline results.

High-School Test Score Analysis Sample

The initial high-school test score panel is constructed from state-specific school-level high-school assessment files, which we compiled from publicly available data on state websites or

[15]The NCES defines "Regular Schools" as a public school providing instruction and education services that does not focus primarily on special education, career/technical education, alternative education, or on any of the particular themes associated with magnet/special program-emphasis schools.

[16]The states included in the panel are listed in Table A2.

from data requests made to state departments. It includes 95,532 school–year observations with non-missing mathematics or English Language Arts test scores. The sample covers 16,500 unique public schools across 26 states over the 2018, 2019, and 2021 through 2025 academic years.

Depending on the state, these score measures come from end-of-course (EOC) exams, statewide SAT or ACT administrations, or end-of-grade (EOG) statewide exams. We retain the measure that best represents the state's primary high-school testing system among those with decent data coverage. We first standardize scores within state, year, and exam. When multiple mathematics or ELA tests remain for the same school-year cell, we combine them using averages weighted by the number of students tested. For example, some school-year cells retain separate Algebra I, Algebra II, and Geometry EOC exams, which are all mapped into the math subject group for each grade. If the number tested is not reported directly, we recover it from reported tested ranges using the midpoint of the range or, when only participation rates are available, from participation rates multiplied by grade-level enrollment. This procedure yields one mathematics score and one ELA score per school–year. We restrict the panel to schools serving at least one high-school grade (9 or above).

We next apply the same common sample restrictions used elsewhere in the analysis. After applying these restrictions, the main high-school test score analysis sample consists of 65,971 school–year observations, representing 12,978 unique schools across 24 states.

Attendance Analysis Sample

The initial attendance panel includes 489,862 school–year observations, covering 73,818 unique public schools.[17] We first restrict the sample to schools serving at least one middle or high school grade. 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 and to regular schools only. 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 203,780

[17] 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). The states included in both the initial attendance panel and the final attendance analysis sample are listed in Table A4. Coverage varies by state and year.

school–year observations, representing 39,706 unique schools. 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.

Disciplinary Outcomes Analysis Sample

The initial disciplinary panel includes 365,798 school–year observations, covering 60,418 unique public schools across 22 states with at least one school-level disciplinary measure reported during the academic year.[18] 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, 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 32 percent of observations use measure (i), 61 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 then apply the same grade-level and treatment-related sample restrictions used in the test score and attendance analyses. After applying all restrictions and excluding observations with missing standardized discipline measures, the final disciplinary analysis sample consists of 159,630 school–year observations, representing 33,147 unique schools across the 22 states listed in Table A6. 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.

[18] 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. The states included are listed in Table A6.

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 subjective 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 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 and respondents. 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 subjective 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 62,289 school–year observations with non-missing survey data, representing 16,285 unique schools across 46 states. Across all school–year observations in the NCES panel during this period, 6.9 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 3000 in 2019 to over 11,000 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 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 phone restrictions is concern about student subjective 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 subjective 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 phones 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 phone 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 907,430 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 402,437 school–year observations.

We next restrict the sample to schools serving at least one middle or high school grade, where phone 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. After applying all restrictions, the final subjective well-being analysis sample consists of 3742 school–year observations. The corresponding classroom attention analysis sample is larger, reflecting broader item coverage, and consists of 10,399 school–year observations. The online bullying analysis sample consists of 6258 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 A21 summarizes the construction of the Panorama analysis samples alongside the other outcomes studied in the paper.

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 4,308 observations out of the 169,163 school–year observations in the final middle school test score analysis sample, and 1,469 out of 65,971 for high schools.

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:

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

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:

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

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

E Robustness Checks for Test Score Estimates

Tables A14 and A15 examine the sensitivity of our estimated treatment effects to alternative estimators, control group definitions, treated group restrictions, and outcome constructions. Across all specifications, the pooled estimates remain close to zero and precisely estimated, while the pattern of positive effects in high schools and negative effects in middle schools is preserved, particularly for Mathematics.

Alternative estimators. We first assess sensitivity to the estimation approach. In addition to the baseline doubly robust estimator, we implement the Callaway and Sant'Anna (2021) estimator using regression adjustment only, omitting the inverse probability tilting component. 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 merge the NTS data to our test score panel, retaining all schools in the panel and attaching survey responses where available. We then implement four alternative control group definitions.

First, we restrict the control group to schools that appear in the NTS and whose modal reported policy is not Yondr. This specification serves as a benchmark, allowing us to isolate the effect of restricting the sample to NTS schools from the additional restrictions imposed in the subsequent specifications. Second, we further restrict to non-Yondr NTS schools that report either no policy change in the past five years or only weak policy changes, excluding schools that adopt strict alternative policies (centralized collection, locker storage, or campus-wide bans). Third, we restrict to non-Yondr NTS schools that report no policy change, regardless of the nature of the policy. Finally, we exclude schools that implement strict phone policies from the control group, regardless of whether a recent change occurred.

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 768 in the test score sample) is very 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. 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 middle school test score 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.

Pooled specification and sample construction. We also examine two additional specifications that address how school characteristics and sample construction affect the pooled estimates. First, we augment the baseline controls by allowing the state-by-urbanicity indicators to vary by school level (middle versus high), thereby permitting more flexible differences across these settings. Second, we address the fact that a relatively small share of schools (approximately 7.1%) appear in both the middle school and high school test score samples. In our baseline pooled specification, such schools are assigned to the middle school sample, given that middle school test scores are more consistently available across states and years. As a robustness check, we instead construct a "stacked" dataset in which these observations are duplicated and treated as separate school-by-year-by-level observations, allowing both middle and high school test scores to enter the pooled analysis.

Across all specifications, the estimated effects remain small and close to zero in the pooled sample, with no evidence of meaningful impacts of Yondr adoption on test scores. At the same time, the contrasting patterns across school levels—modest positive effects in high schools and smaller negative effects in middle schools—remain stable across specifications, particularly when focusing on Mathematics (Table A15).