

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, Revised September 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 Maya Brown-Hunt, 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 OSF. 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>

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Hunt Allcott, E. Jason Baron, Thomas Dee, Angela L. Duckworth, Matthew Gentzkow, and Brian Jacob. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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, Revised September 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 a large-scale teacher survey, 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 we find suggestive evidence that student subjective well-being falls, consistent with short-term disruption. However, disciplinary effects fade, while the estimated effects on well-being appear to recover and become positive in later years. 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 and no clear evidence of improvements in self-reported classroom attention or perceived online bullying.

Hunt Allcott
Stanford University
and NBER
allcott@stanford.edu

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

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

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

Thomas Dee
Stanford University
Graduate School of Education
and NBER
tdee@stanford.edu

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

An online appendix is available at <http://www.nber.org/data-appendix/w35132>

I 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 restriction that proponents of limits on phones have advocated for (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. Our primary specification uses the regression adjustment implementation of this estimator. 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 stan-

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 a new national survey we fielded between April 2025 and February 2026 that provides additional information on policy implementation and the views of teachers. 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 by using our teacher survey to characterize the phone policies of Yondr-adopting and comparison schools. Understanding this policy 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. There is substantial variation in school phone policies. Some schools restrict phone use 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 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 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 phone 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 15 percent increase in suspension rates (in-school or out-of-school)—but this effect fades in subsequent years. We also find suggestive evidence of a dynamic response in student-reported subjective well-being: 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, Braghieri, Eichmeyer and Gentzkow (2020) find that deactivating Facebook for four weeks increased subjective well-being by 0.09 standard deviations. These estimates are suggestive of short-term disruption as Yondr pouches are adopted, followed by adaptation and subsequent improvements in well-being. However, as we discuss below, we interpret this pattern cautiously, as the relatively short panel limits our ability to fully distinguish dynamic treatment 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 0.007 student-level standard deviations. These averages reflect small and precisely estimated effects across subjects, but mask interesting heterogeneity across grade levels. In general, high schools see modest positive effects while middle schools see small negative effects, although this difference is not statistically significant. These differences are larger and statistically significant for math, where adoption in high schools increases student achievement by 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 0.189 percentage points. Relative to a mean attendance rate of 93 percent, this corresponds to a change of roughly 0.2 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 and online bullying, however, some evidence of differential pre-trends limits the conclusions we can draw for these outcomes.

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, attendance, discipline, and student survey measures are widely used in the education literature, they may not capture 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.

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

¹Consistent 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).

effects of similar bans in Sweden. Lichand, Moreno-Louzada, da Costa and Gentzkow (2026) find that a district-wide ban in Rio de Janeiro led to improvements in test scores. Beneito and Vicente-Chirivella (2022) find 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.

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 additional outcomes that are 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 digital devices and social media in adult populations. Experimental and quasi-experimental studies show that reducing access to social media platforms can improve subjective well-being (Allcott, Braghieri, Eichmeyer and Gentzkow, 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 labor market outcomes, with substantial peer spillovers (Barwick, Chen, Fu and Li, 2026). This literature mainly 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 influence may operate differently.

II Background

II.A 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. 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 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.

II.B 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 improves outcomes across these domains. We study standardized test scores and attendance as measures of academic performance and school engagement, student-reported classroom attention as a measure of in-class focus and engagement, 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 state.

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.

With regard to disciplinary incidents and school climate, limiting phone access may reduce 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, Braghieri, Eichmeyer and Gentzkow, 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.

III Data

A key challenge for national research on school phone policies is the absence of standardized records documenting policy adoption and implementation. In addition, many outcomes central to the debate over school phone restrictions are not captured in administrative data, while traditional outcomes such as test scores, attendance, and discipline are typically reported through separate state-level systems that differ in measurement and availability.

To address these challenges, we combine administrative adoption data from Yondr with GPS-based phone location data, a national teacher survey, and harmonized school-level administrative datasets. Together, these sources allow us to identify Yondr adoption, quantify changes in student phone use, characterize the broader policy environment, and construct consistent measures of both administrative and student-reported outcomes. All datasets are merged using standardized National Center for Education Statistics (NCES) school identifiers. Appendix C provides additional details.

III.A 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 4608 U.S. public schools that purchased pouches between 2014 and the end of our study period. These records include school identifiers and implementation timing.

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.

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. Coverage of minors may vary across contexts because of parental consent requirements, OS-level privacy protections, and because app and provider systems do not consistently capture user age in a way that would support age-based filtering. At the same time, the underlying panel is expected to include both minors and adults, as it reflects opt-in users of partner apps, which often do not record user age.

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). 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. 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. Appendix C.2 discusses these caveats and our data processing procedures in more detail and presents several validation exercises showing that, despite these limitations, the data contain meaningful signal.

Nationwide Teacher Survey (NTS). 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 document school cell phone policies and student phone use. The survey was fielded nationally through partnerships with state education leaders and direct outreach to educators. Since its national release on April 8, 2025, the survey has received nearly 108,000 responses from educators in all 50 states, representing roughly 50 percent of U.S. middle and high schools.

The NTS collects information on school phone policies, including when phones are restricted, where students keep their phones, and how policies are enforced. It also collects teacher reports on student phone use and teachers’ evaluations of their school’s phone policy. Finally, the survey documents recent policy changes by asking whether schools changed their phone policy during the past five years and, if so, when the change occurred, what the prior policy was, and whether other major school-level changes occurred at roughly the same time. Appendix C.3 reproduces the survey questions used in the analysis.

Table A.1 compares baseline characteristics of schools in NTS to the full NCES population of middle and high schools. Relative to the NCES population, schools represented in the NTS are larger on average and more likely to serve high school grades. NTS schools are also somewhat more suburban, less likely to be charter schools, and serve student populations that are from higher-income neighborhoods.

Common Core of Data. We also use the Common Core of Data (CCD), maintained by the NCES, which provides annual school-level administrative data for all public schools in the United States. The CCD serves 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. Second, we use CCD data to construct baseline school-level covariates for our difference-in-differences analyses. These include total enrollment, enrollment by race and 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 K–8 schools from schools serving only middle school grades, as well as middle schools from combined middle–high schools. To measure school-level poverty, we use income-to-poverty ratios from the EDGE School Neighborhood Poverty Estimates. Finally, we use CCD data to examine whether Yondr adoption is associated with changes in student composition.

Standardized Test Score Data. 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 com-

parable achievement data across the country and over time.²

We begin with data from the EDC. 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 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*.

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. See Table A.2 for details on which source was used for each state.

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.

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 end-of-grade exams, 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. We therefore compiled high school assessment data ourselves, relying on publicly available files from state departments of education when available and requesting data directly from states otherwise. Ultimately, we obtained high school assessment data from 26 states, representing 64 percent of high school students nationwide.

We focus on school-level average math and ELA assessment scores. Table A.3 reports

²The EDC’s State Assessment Data Repository is sometimes informally referred to as “Zelma,” the name of its AI interface for querying the data.

the assessments administered in each state over our sample period. Regardless of the specific exam, these assessments were required for all students in the relevant grade or course. We standardize school-level average scores by state-year-exam and then construct school-level achievement measures by taking weighted averages across exams within math and 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.

School-Level Attendance Data. We also assemble school-level attendance data from state education agencies. We focus on two attendance measures: attendance rates and chronic absenteeism. Attendance rates measure the percentage of enrolled student-days on which students are present, while chronic absenteeism measures the share of students missing at least a state-defined threshold of instructional days, most commonly 10 percent. Because attendance rates are continuous and have been used in prior research on school phone bans (e.g., Figlio and Ozek (2025)), we pre-specified attendance rates as our primary outcome and chronic absenteeism as a secondary measure.

In nearly all states, out-of-school suspensions are coded as absences, so attendance rates may reflect both instructional attendance and disciplinary exclusions. California is an exception, reporting disciplinary absences separately. We therefore construct an additional attendance measure excluding disciplinary absences and report the corresponding estimates as a robustness check. Attendance-rate data are available for 31 states between 2018 and 2025, while chronic absenteeism data are available for 38 states. Tables A.4 and A.5 show data availability and definitions for each state.

School-Level Disciplinary Outcomes. We similarly assemble school-level disciplinary data from state education agencies. 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.

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.4 details the construction of the discipline index and Table A.6 shows data availability and definitions for each state. The final analytic sample includes 22 states with discipline data between 2018 and 2025.

Panorama Student Survey Data. To examine outcomes not typically available in standard administrative data, 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 across states, schools, and years.

To protect client confidentiality, Panorama does not disclose school or respondent identities. Instead, Panorama provided anonymized school-by-year aggregates through a secure matching process. For each matched school-year, we observe the mean response to each survey item across all responding students. 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 survey items and survey content varies across schools and years, we pre-specified a small set of primary outcomes before estimating treatment effects. This approach was designed to reduce concerns about specification searching and selective reporting. We focus on three outcomes: subjective well-being (SWB), classroom attention, and students’ perceived likelihood of online bullying. These outcomes reflect central channels through which phone restrictions are hypothesized to affect students. We first code all survey items so that higher values correspond to more desirable outcomes and standardize them within survey year. For SWB and classroom attention, we combine the multiple survey items corresponding to each construct into indices by taking the unweighted average of the standardized items and then re-standardize the resulting indices by year. In contrast, the online bullying outcome is based on a single standardized survey item. All three outcomes are therefore expressed in school-level standard deviation units, facilitating comparison across outcomes.

Table A.1 compares baseline characteristics of schools with non-missing Panorama data for at least one of the three indices to the full NCES population, showing that, relative to all NCES schools, Panorama participation is more common among larger schools, schools located in urban areas, schools serving more racially diverse student populations, and schools

in higher-income neighborhoods. Appendix C.5 provides more specific details about the Panorama data, including additional documentation of the outcome selection process.

III.B Outcome-Specific Analysis Samples

Because outcomes are drawn from distinct data sources with different coverage across states, years, and schools, each outcome is analyzed using the largest possible sample in which that outcome is consistently observed, subject to a common set of sample restrictions. Across all outcomes, we restrict the sample to regular schools serving at least one middle or high school grade (6–8 and 9–12, respectively). We focus on academic years 2018 through 2025, excluding the pandemic-affected years 2020 and 2021. We further exclude schools that adopted Yondr before the 2023 academic year, schools in districts that may have adopted Yondr but for which school-level adoption timing cannot be reliably determined, and schools with the lowest-confidence implementation dates in the Yondr administrative data. Finally, we restrict the sample to states with at least one Yondr adoption between the 2023 and 2025 academic years so that identification relies on within-state variation in adoption timing. We then apply outcome-specific restrictions, requiring non-missing observations for the relevant outcome in each school-year. Table A.7 reports the number of Yondr-adopting schools retained for each outcome, together with the corresponding share of potential adopters. Appendix C provides additional details on the sample construction for each outcome. Figure A.1 presents the raw mean outcomes by adoption cohort and calendar year for all primary outcomes.

III.C 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.

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

$$Y_{it} = \sum_{k \neq -1} \beta_k \cdot \mathbb{1}[\text{event_time}_{it} = k] + X_i' \gamma_t + \alpha_i + \delta_t + \varepsilon_{it}, \quad (1)$$

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 $\times$ urbanicity. These covariates enter the model interacted with year fixed effects, allowing their associations with outcomes to vary flexibly over time while remaining anchored to pre-treatment values.

While the TWFE event-study specification is intuitive and widely used, recent work shows that it can yield biased estimates in settings with staggered treatment adoption and heterogeneous treatment effects (e.g., see Roth, Sant’Anna, Bilinski and Poe (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.³ Our baseline specification uses the regression adjustment implementation of this estimator. 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 $ATT(g, t)$, 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 $ATT(g, t)$'s are estimated for different values of g and t , we aggregate them to summarize treatment effects in two main ways. First, we report an overall ATT, which averages treatment effects across all post-adoption years and cohorts. Second, we construct event-study estimates that trace the evolution of treatment effects relative to the year immediately preceding adoption, allowing treatment effects to vary flexibly over time. Because outcomes are measured at the school level, later event-time estimates may reflect a combination of greater cumulative student exposure to the policy, changing student cohorts, and changes in implementation over time.

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.

IV.A Characterizing Yondr Treatment and Control

Treatment. Figure I illustrates the timing of Yondr adoption across schools. While a small number of schools adopted Yondr prior to the COVID-19 pandemic, adoption increased sharply beginning in the 2023 school year, with the largest number of schools adopting in 2025 and 2026. To focus on the period in which adoption accelerates and to allow for longer pre-treatment trend analysis, our main analyses restrict attention to schools that adopt Yondr in 2023, 2024, or 2025. Schools that adopted Yondr between 2018 and 2022 are excluded from the analysis. Schools scheduled to adopt Yondr in 2026 are grouped with never-treated schools and serve as part of the control group throughout the analysis. In robustness checks, we show that our results are not sensitive to how these schools are handled, including specifications that exclude 2026 adopters entirely.

³We 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.

[Figure I about here.]

Table I reports summary statistics for schools in the pooled (middle and high school) test score sample, comparing Yondr adopters to never-treated schools. All covariates are measured in the first year each school appears in the sample. Relative to never adopters, Yondr-adopting schools are larger, more likely to be located in urban areas, and serve student populations with 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 differences underscore the non-random selection of schools into Yondr adoption and motivate the inclusion of baseline covariates in our difference-in-differences specification.

[Table I about here.]

Table A.8 summarizes the phone policy environments that preceded Yondr adoption, based on teacher reports from the NTS.⁴ 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 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.

Control. Figure A.2 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*

⁴The 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 A.9). 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.

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). In 2023, roughly 1 percent of control schools adopted a strict alternative policy; in 2024, 2 percent; and in 2025, 5 percent. 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.

V Difference-in-Differences Estimates

V.A 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 A.10 reports estimates of the ATT of Yondr adoption averaged across post-adoption periods and cohorts. 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 0.18 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.⁵ 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 A.11).

Figure II presents event-study estimates of the ATT of Yondr adoption on the natural log

⁵One possible explanation is that phone use is more prevalent in high schools than in middle schools. Using the NTS, we estimate that, among schools reporting a phone policy change, teachers report that approximately 48 percent of middle school students and 66 percent of high school students used their phones in class for personal reasons prior to the adoption of the restriction.

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 0.16 log points. The decline grows in subsequent years, reaching 0.2 log points in $t + 1$ and 0.3 log points by $t + 2$.

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 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.

[Figure II about here.]

We next turn to survey evidence from the NTS. Teachers report whether their school experienced a major change in its official cell phone policy during the past five years. If so, the survey asks parallel questions about current conditions and conditions before 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 schools reporting no policy change, retrospective measures are unavailable, so we set pre- and post-policy outcomes equal to current reported values. The resulting comparison therefore mirrors our research design.

Figure III 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, while reported in-class student phone use for personal reasons falls from 61 percent to 13 percent, a decline of roughly 48 percentage points. Reported phone use between classes also declines sharply. Non-Yondr schools also exhibit improvements in teacher satisfaction and reductions in reported phone use, consistent with some schools adopting less restrictive alternatives such as “no-show” rules or classroom-level

collection. Even so, the changes associated with Yondr adoption are substantially larger across all measures. These comparisons are descriptive and rely on retrospective teacher reports. Nevertheless, the survey evidence is consistent with the GPS-based results and suggests that Yondr adoption meaningfully reduces in-school phone use, even if the exact magnitude of the reduction cannot be read directly from any single measure.

[Figure III about here.]

V.B Effects on Discipline, Attendance, and 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 III, 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.

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, so that event times $t - 5, \dots, t - 1$ correspond to successive pre-adoption years. This relabeling is expositional and does not affect the estimated post-adoption treatment effects.

Discipline and Attendance. Column (1) of Table II reports the ATT of Yondr adoption on the school-level disciplinary index. Adoption increases the disciplinary index by 0.057 (SE = 0.023) school-level standard deviations, corresponding to approximately 0.02 student-level standard deviations (see Section III.C 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 0.32 percentage points, corresponding to about an 11 percent increase relative to baseline levels.⁶ Figure IV (Panel A)

⁶This 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

reports event-study estimates for the disciplinary index. At adoption, disciplinary incidents rise sharply, by approximately 0.078 (SE = 0.019) school-level standard deviations, corresponding to roughly 0.03 student-level standard deviations or about a 15 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 II about here.]

Column (2) reports effects on the attendance rate. The estimate of -0.003 percentage points (SE = 0.098) allows us to rule out even modest improvements relative to the mean attendance rate of 93 percent. Figure IV (Panel B) reports the corresponding event-study estimates. The estimates remain close to zero throughout the event window, with no evidence of meaningful changes following adoption. Finally, Table A.12 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.

[Figure IV about here.]

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 IV (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, Braghieri, Eichmeyer and Gentzkow (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 IV (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

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 0.057 implies an increase of roughly 0.32 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.

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 IV (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.

A potential concern is that, because Panorama surveys are administered at the discretion of schools and districts, Yondr adoption could affect the availability of survey outcomes. For example, schools may begin administering Panorama surveys to monitor the rollout of Yondr. Figure A.3 examines this possibility by estimating the effects of Yondr adoption on the availability of each Panorama outcome. We find no evidence that Yondr adoption affects the availability of subjective well-being observations, nor do we find evidence of differential pre-treatment trends in their availability. In contrast, the availability of classroom attention and online bullying outcomes exhibits some evidence of differential pre-trends, making it difficult to assess whether Yondr adoption affects the availability of these outcomes. These findings reinforce our cautious interpretation of the classroom attention estimates and suggest that the online bullying estimates should also be interpreted with caution.

Heterogeneity by School Level and Cohort. We next examine heterogeneity by school level and adoption cohort across all outcomes. Figures A.4 and A.5, along with Table A.13, 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.

Figures A.6, A.7, and A.8, together with Table A.13, 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 A.9, A.10, A.11, A.12, and A.13 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, the 2024 adoption cohort is omitted because no schools in that cohort have both pre- and post-adoption SWB observations in the final estimation sample. The cohort-specific estimates are otherwise 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 the 2025 cohort, while the subsequent recovery comes from the 2023 cohort. These results highlight that the relatively short panel and staggered adoption make it difficult to fully disentangle dynamic treatment effects from cross-cohort heterogeneity.

V.C Effects on Test Scores

We next examine whether restricting phone access affects student academic achievement. Table III 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 (-0.006 , $SE = 0.010$), with similarly small estimates for Math and ELA. This corresponds to approximately -0.003 student-level standard deviations, and the precision of the estimates allows us to rule out improvements larger than approximately 0.007 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.

[Table III about here.]

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 0.007 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 0.009 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 0.035 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.

Panel A of Figure V reports the corresponding event-study estimates. 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 III and Figure V report effects separately for high schools and middle schools. In high schools, the combined score increases by 0.026 (SE = 0.017), with a statistically significant increase in math of 0.047 (SE = 0.018) and a near-zero estimate for ELA. In middle schools, the combined score declines by -0.025 (SE = 0.012), with a decline in math of -0.026 (SE = 0.013) and a smaller, statistically insignificant estimate for ELA.

[Figure V about here.]

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 0.047 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.⁷ 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 A.16. 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 statistically insignificant in both t and $t + 1$. For the 2025 cohort, for which we observe only the adoption year, the estimate is small, positive, and statistically significant (0.029 school-level standard deviations).

⁷Figures A.14 and A.15 present the corresponding event-study estimates for math and ELA, respectively.

V.D 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 heterogeneity analyses on test score outcomes. Tables A.14 and A.15 report estimates from a range of alternative specifications. Across these specifications, the pooled estimates (combining middle and high schools) generally 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 A.15 further shows that this heterogeneity is robust when focusing specifically on math test scores, where the estimated effects are largest. Appendix E provides additional details on each of these checks.

We first assess sensitivity to the estimation approach. Using the doubly robust implementation of the Callaway and Sant’Anna (2021) estimator, rather than the regression adjustment implementation, yields nearly identical estimates. We next modify the composition of the comparison group to address the possibility that changes in phone policies among comparison schools attenuate the estimated effects of Yondr adoption. First, using information from the NTS, 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 phone 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 phone policy change in the past five years; and (iv) excluding schools that implement strict phone policies from the comparison group, regardless of whether a recent policy change occurred. We also exclude comparison schools located in states that adopted statewide bell-to-bell phone restrictions during our sample period. Because these laws prohibit student phone use throughout the school day, they could similarly reduce the contrast between treated and comparison schools.

Additional specifications exclude future adopters (the 2026 cohort) from the comparison group, restrict the sample to schools adopting between 2023 and 2026 to leverage only within-adopter timing variation, 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), control for the number of students tested, include baseline test scores in the regression adjustment component of the difference-in-differences estimator, and cluster standard errors at the district level.

We also consider two robustness checks that are specific to the pooled estimates. First, we augment our baseline controls by allowing the state-by-urbanicity indicators to vary by school level (middle versus high school). Second, 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 the middle and high school test scores to enter the pooled analysis. Both checks yield nearly identical estimates.

Table A.16 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 A.17 examines whether Yondr adoption is associated with changes in school enrollment or student demographics, including 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. Consistent with this evidence, Figure A.17 reports event-study estimates using predicted test scores based on school composition. These estimates remain close to zero throughout the event window, with no evidence of systematic changes following adoption.

V.E Subgroup Analysis

Before exploring subgroup heterogeneity, we pre-registered the dimensions along which we would examine treatment effect variation. This section presents results for those pre-specified subgroups. We consider two types of heterogeneity. Table A.18 reports school-level heterogeneity by poverty status, comparing schools above and below the within-state median of the neighborhood income-to-poverty ratio. Table A.19 reports heterogeneity across within-school student subgroups in middle schools, including gender, race and ethnicity, and economic disadvantage. These within-school, subgroup-specific outcomes are not available for high schools, so this analysis is limited to the middle school sample.

Across the pre-registered subgroups, we find limited evidence of systematic heterogeneity. In high schools, effects are positive for both high- and low-poverty schools and somewhat larger for higher-poverty schools, although the difference between the two groups is not statistically distinguishable. In middle schools, estimates are generally negative and small in magnitude for both high- and low-poverty schools. Within middle schools, estimates

are negative across all student subgroups and generally small in magnitude. Effects are similar for male and female students. Across racial and ethnic groups, point estimates vary but the differences are not statistically significant. Estimates are somewhat more negative for students who are not economically disadvantaged than for economically disadvantaged students, although these differences are not statistically distinguishable.

Finally, we report two additional heterogeneity analyses that were not pre-registered and should therefore be interpreted with caution (see Appendix F for details). First, using Yondr’s internal measure of implementation quality (Table A.20), we find no consistent relationship between implementation ratings and treatment effects. Second, we examine heterogeneity across schools based on baseline characteristics (Table A.21), including school size and baseline achievement. We find no evidence of positive effects on pooled test scores for any subgroup. In middle schools, negative effects are more pronounced among higher-achieving schools. In high schools, effects are more positive in larger schools. However, these patterns are not consistently observed across specifications and should be interpreted as suggestive, particularly given the number of subgroup comparisons considered.

Although we find limited evidence of systematic heterogeneity, one possible explanation for the variation we do observe is that some school subgroups experience larger reductions in student phone use than others. Figure A.18 compares subgroup-specific treatment effects on GPS-measured phone activity with the corresponding effects on standardized test scores. Each point represents the estimated ATT for the same subgroup in the GPS and test score data, allowing us to assess whether larger reductions in phone activity are associated with larger changes in test scores. We estimate these effects for all school subgroups considered in this section, as described in the figure notes.

We find little evidence of such a relationship. School subgroups experiencing larger reductions in GPS-measured phone activity do not systematically exhibit larger improvements in test scores. However, we emphasize that this comparison should be interpreted with caution. It is intended to be descriptive rather than causal. The exercise assesses whether heterogeneity in reductions in phone use correlates with changes in student outcomes, but it cannot distinguish whether any such relationship arises because larger reductions in phone use *cause* larger effects on outcomes or because treated subgroups experiencing larger reductions in phone use have *other characteristics* that also affect student outcomes.

VI 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 un-

dermine 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 evidence that adopting lockable phone pouches may generate some short-run disruption to the school environment. Disciplinary incidents increase in the year of implementation, and we find suggestive evidence that student-reported subjective well-being also declines, consistent with adjustment costs as schools and students adapt to the new restriction. Over time, disciplinary impacts fade, while subjective well-being appears to recover and become positive in subsequent years. Average effects on standardized test scores and attendance are close to zero and precisely estimated. While these average effects mask modest positive effects in high schools and small negative effects in middle schools, the estimated magnitudes are relatively small in both settings. We also do not find evidence of improvements in classroom attention or perceived online bullying, although differential pre-trends limit the conclusions that can be drawn from these measures.

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.

Author Affiliations

Hunt Allcott
Stanford University and NBER
allcott@stanford.edu

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

Thomas Dee
Stanford University and NBER
tdee@stanford.edu

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

Matthew Gentzkow
Stanford University and NBER
gentzkow@stanford.edu

Brian Jacob
University of Michigan and NBER
bajacob@umich.edu

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*.
- Daw, Jamie R, and Laura A Hatfield. 2018. “Matching and regression to the mean in difference-in-differences analysis.” *Health services research*, 53(6): 4138–4156.
- 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.
- 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.” National Bureau of Economic Research Working Paper 35233.
- 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.” *Education Week*. 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.

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.

Table I: Summary Statistics by Yondr Adoption Status

	(1)	(2)	(3)
	Never adopters	Yondr adopters	Diff. p-value
Total enrollment	630	698	0.000
Share White	0.52	0.28	0.000
Share Black	0.15	0.30	0.000
Share Hispanic	0.25	0.34	0.000
Urban	0.27	0.58	0.000
Suburban	0.31	0.22	0.000
Rural/Town	0.42	0.20	0.000
Charter school	0.14	0.13	0.217
Schools with grades above 9	0.32	0.38	0.000
Schools with grades below 6	0.43	0.31	0.000
Income-to-poverty ratio	308	272	0.000
Avg. test score index	0.06	-0.43	0.000
Avg. Math score	0.07	-0.43	0.000
Avg. English score	0.07	-0.41	0.000
Number of schools	42,919	1778	

Notes: This table reports summary statistics for schools in the pooled 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 for the differences between (1) and (2), using robust standard errors. All covariates are measured in the first year each school appears in the sample and precede treatment for all adopting schools. For never-adopting schools, this year is 2018 for 86% of schools and the first observed year thereafter for the remainder. Baseline test scores are reported for descriptive purposes but are not included in the regression specification. Income-to-poverty ratios enter the regression specification as state-specific quartiles.

Table II: 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	0.057** (0.023)	-0.003 (0.098)	-0.305 (0.201)	-0.088 (0.085)	-0.039 (0.092)
Mean of dependent variable	-	92.62	-	-	-
Observations	155,712	201,708	3382	9247	5648

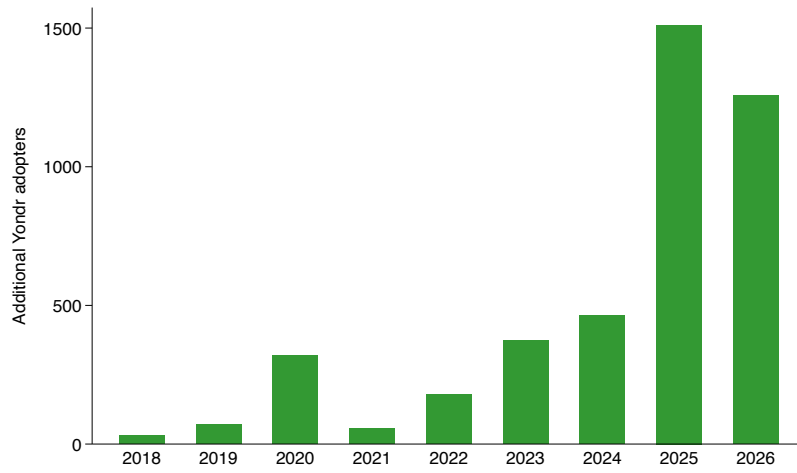
Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption. Estimates are based on the regression adjustment implementation of the Callaway and Sant’Anna (2021) difference-in-differences estimator. Column (1) reports effects on a standardized school-level discipline 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 well-being 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 follow the main empirical specification. Standard errors are clustered at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table III: 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)
<i>Panel A: Pooled</i>			
ATT	−0.006 (0.010)	−0.002 (0.011)	−0.006 (0.011)
Observations	214,556	210,951	211,259
<i>Panel B: High School</i>			
ATT	0.026 (0.017)	0.047** (0.018)	0.012 (0.021)
Observations	65,978	64,585	65,318
<i>Panel C: Middle School</i>			
ATT	−0.025** (0.012)	−0.026** (0.013)	−0.019 (0.013)
Observations	162,974	159,775	159,987

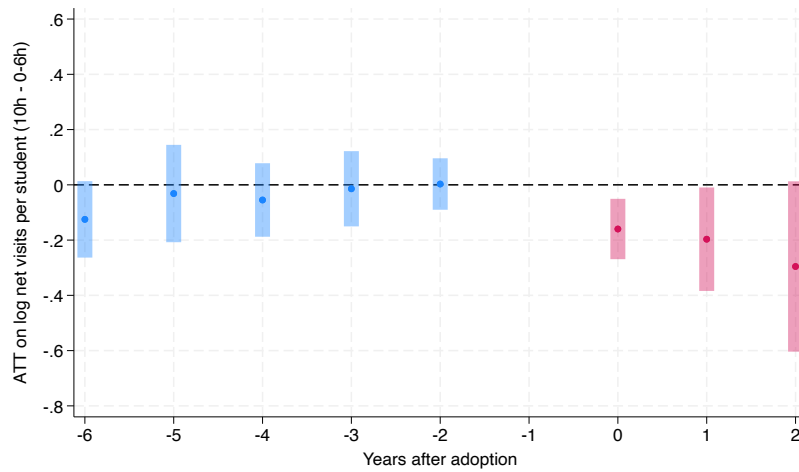
Notes: Each column reports the average treatment effect on the treated (ATT) of Yondr adoption on standardized test scores, measured in school-level standard deviations. Estimates are based on the regression adjustment implementation of the Callaway and Sant’Anna (2021) difference-in-differences estimator. Column (1) reports estimates for the combined math and ELA test score. 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 follow the main empirical specification. Standard errors are clustered at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure I: Yondr Adoption Over Time



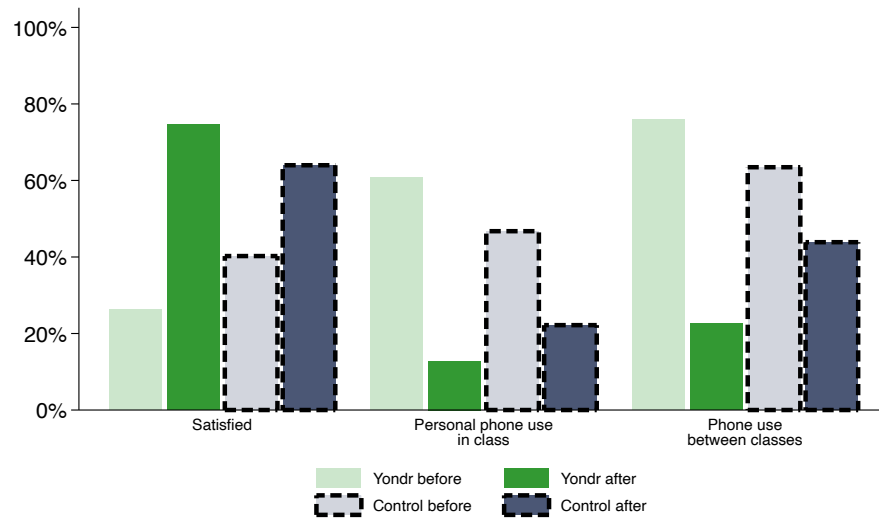
Notes: The figure shows the number of schools adopting Yondr lockable pouches by adoption year, based on administrative data from Yondr. Adoption year is defined as the spring of the academic year in which a school is first observed implementing Yondr. Each school appears once, in its initial year of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

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

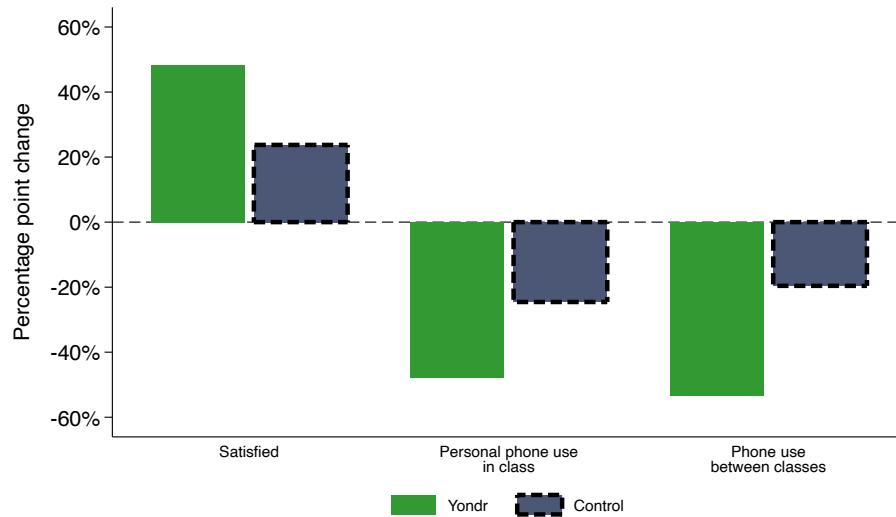


Notes: This figure reports event-study estimates of the effect of Yondr adoption on pooled (across middle and high schools) phone use, measured using GPS location data from Advan. Estimates are based on the regression adjustment implementation of the Callaway and Sant’Anna (2021) difference-in-differences 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 ($t = -1$). 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 follow the main empirical specification.

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



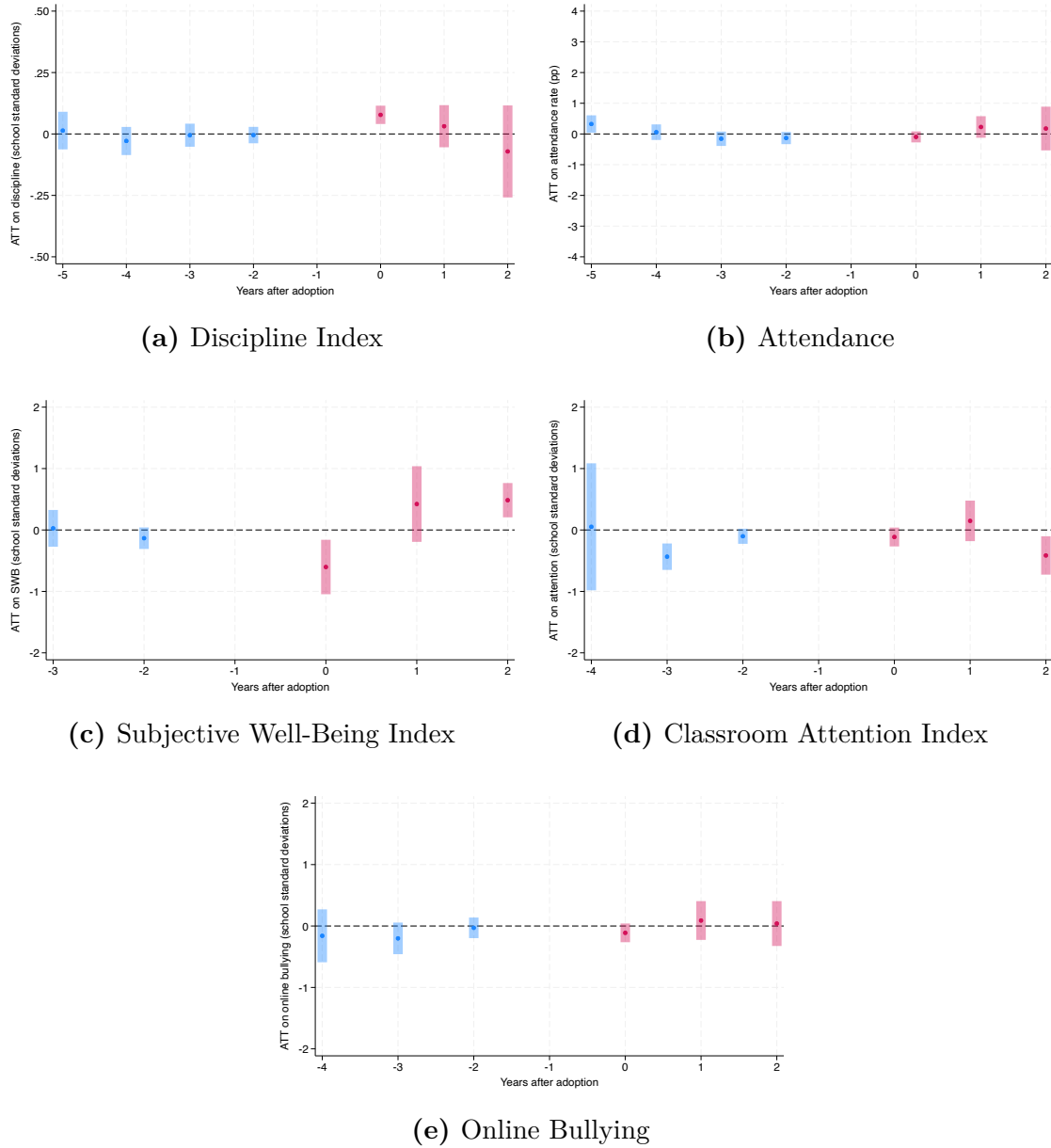
(a) Levels Before and After Policy Change



(b) Change Relative to Pre-Policy Baseline

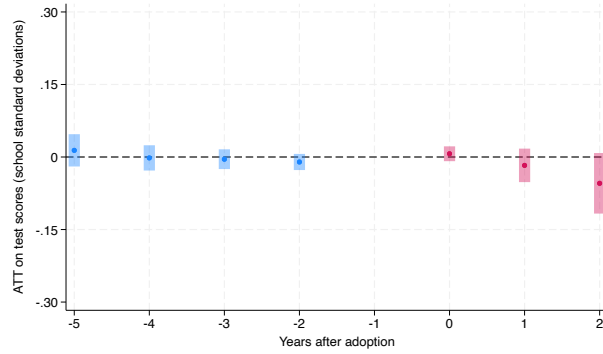
Notes: This figure uses data from the NTS to compare teacher-reported outcomes before and after reported phone policy changes. The sample is restricted to middle and high schools. Schools are classified using the modal response to Q3 (“Where are students allowed to keep their phones?”), with Yondr schools 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.

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

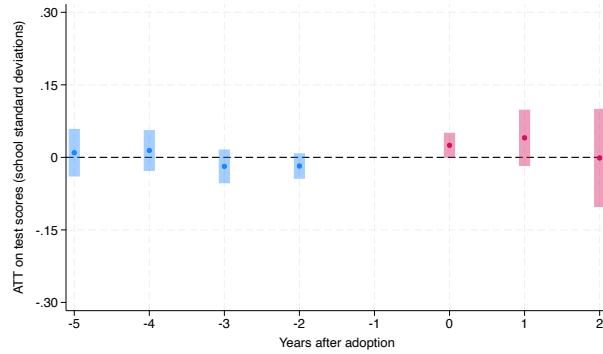


Notes: This figure reports event-study estimates of the effects of Yondr phone pouch adoption on pooled (across middle and high schools) discipline, attendance, subjective well-being, classroom attention, and online bullying. Estimates are based on the regression adjustment implementation of the Callaway 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 follow the main empirical specification. Standard errors are clustered at the school level. The shorter pre-treatment windows for the subjective well-being, classroom attention, and online bullying outcomes reflect the later start of Panorama survey coverage. Academic years 2020 and 2021 are excluded due to pandemic-related data unavailability; the 2018 and 2019 school years are relabeled as 2020 and 2021 so that pre-adoption event times are consecutive.

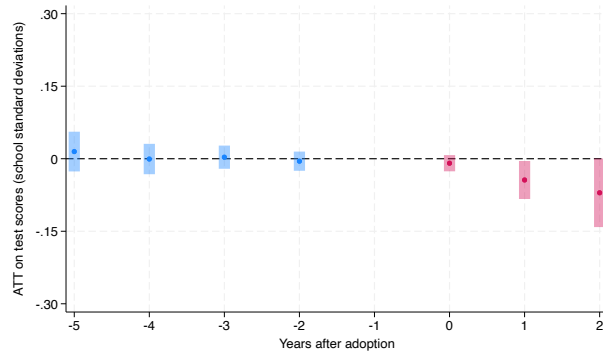
Figure V: Event Study Estimates of Effects of Phone Pouches on Pooled, High School, and Middle School Test Scores



(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 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 follow the main empirical specification. Standard errors are clustered at the school level. Academic years 2020 and 2021 are excluded due to pandemic-related data unavailability; the 2018 and 2019 school years are relabeled as 2020 and 2021 so that pre-adoption event times are consecutive.