NBER WORKING PAPER SERIES

# INSTITUTIONAL VARIATION IN CHILD MALTREATMENT REPORTING

E. Jason Baron
Micah Y. Baum
Joseph J. Doyle Jr.
Maria D. Fitzpatrick
Brian Jacob
Joseph P. Ryan

Working Paper 35619
http://www.nber.org/papers/w35619

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

We thank Anna Aizer, David Chan, Janet Currie, Marie-Pascale Grimon, Peter Hull, Chris Mills, Kate Musen, Katherine Rittenhouse, and seminar participants at the FLX Economics of Education Conference, Georgia Tech, NBER Children and Families Spring 2026 Meeting, Notre Dame, Opportunity Insights, the University of Chicago, and the University of Michigan for valuable feedback. The project received approval from the University of Michigan's Institutional Review Board: HUM00104615. This research used data structured and maintained by the MERI-Michigan Education Data Center (MEDC). MEDC data are modified for analysis purposes using rules governed by MEDC and are not identical to those data collected and maintained by the Michigan Department of Education and/or Michigan's Center for Educational Performance and Information. Micah Baum gratefully acknowledges support from the National Science Foundation Graduate Research Fellowship Program (DGE1841052) and the Institute for Education Sciences (R305B150012). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not reflect the views of any other entity or the National Bureau of Economic Research.

Institutional Variation in Child Maltreatment Reporting
E. Jason Baron, Micah Y. Baum, Joseph J. Doyle Jr., Maria D. Fitzpatrick, Brian Jacob, and Joseph P. Ryan
NBER Working Paper No. 35619
August 2026
JEL No. D81, H75, I28, I38, J13

**ABSTRACT**

Involvement with the child welfare system varies dramatically across places. While some of this variation may reflect differences in underlying maltreatment risk, it may also result from differences in institutional responses to similar circumstances. To disentangle these explanations, we study Michigan public school data linked to Child Protective Services reports from 2017–2024. School-level reporting rates range from virtually none to more than half of students reported between kindergarten and fifth grade. We estimate school-specific reporting tendencies using value-added methods and document substantial variation in reporting behavior after accounting for differences in student risk. Using student movers, we show that these estimates are forecast-unbiased. Our estimates imply that school-specific reporting practices account for roughly 40 percent of the variation in reporting rates across schools. We then estimate a structural model that decomposes reporting practices into diagnostic skill and preferences. Differences in skill explain more than twice as much variation as differences in preferences, and higher-reporting schools tend to be more skilled at identifying high-risk cases. Counterfactual simulations imply that improving detection skill would yield significant welfare gains, while recent policy proposals aimed at reducing or equalizing reporting rates generally lower welfare by increasing the number of high-risk cases that go unreported.

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

Micah Y. Baum
University of Michigan
Gerald R. Ford School of Public Policy
Department of Economics
mybaum@umich.edu

Joseph J. Doyle Jr.
Massachusetts Institute of Technology
MIT Sloan School of Management
and NBER
jjdoyle@mit.edu

Maria D. Fitzpatrick
Cornell University
and NBER
maria.d.fitzpatrick@cornell.edu

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

Joseph P. Ryan
University of Michigan
joryan@umich.edu

# 1 Introduction

Involvement with the child welfare system is a common experience for children in the United States and carries substantial consequences for children and families. By age 18, more than one third of children are subject to a formal child welfare investigation for alleged maltreatment, and roughly five percent enter foster care (Doyle, Emanuel and Eyzaguirre, 2025; Kim et al., 2017; Yi, Edwards and Wildeman, 2020). A growing body of research shows that contact with the child welfare system can have lasting effects on children's later-life outcomes, including criminal justice involvement and earnings (Bald et al., 2022). Prior work has also documented substantial variation in child welfare involvement across racial and socioeconomic groups, and across places. More than half of Black children are investigated at some point during childhood, and up to ten percent enter foster care. Across states, cumulative childhood investigation risk ranges from 14 percent to 63 percent, while the risk of confirmed maltreatment ranges from 3 percent to 27 percent (Yi et al., 2023). This variation is equally stark at finer geographic levels (Fong, 2019). In Michigan, the setting for this study, annual CPS reporting rates range from about 2 percent in census tracts at the 10th percentile of the distribution to more than 15 percent at the 90th percentile.[1]

These large disparities have raised concerns that some families may be exposed to unnecessary surveillance and intervention (Dettlaff and Boyd, 2020; Edwards et al., 2021; Fong, 2020; Raz, 2020). To determine the proper policy response, however, a central question is whether these differences primarily reflect underlying maltreatment risk or differences in how institutions and mandated reporters respond to similar circumstances. Distinguishing between these explanations is essential for designing and evaluating the growing number of reforms across states aimed at reducing disparities in child welfare system involvement (New York State Legislature, 2024; Safe & Sound, 2025).

In this paper, we focus on child maltreatment reporting, the initial step of the child welfare process. The stakes of these reporting decisions are high. Failing to report a case of true maltreatment can leave a child in a harmful environment, with potentially severe and lasting consequences. Moreover, professionals such as teachers and physicians are "mandated reporters" who can face criminal penalties for failing to report suspected maltreatment. At the same time, reporting a case that does not warrant intervention can subject families to intrusive scrutiny, including formal investigations and the threat of having their children

[1] We assign each child to the census tract of residence recorded in Michigan education records and calculate annual tract-level reporting rates as the share of children residing in the tract who are reported to CPS during the school year.

removed. Because these decisions affect a large share of children and involve substantial risks on both sides, understanding how and why reporting varies across institutions is critical not only for evaluating child welfare policy, but also for understanding how frontline institutions shape children's and parents' exposure to high-stakes interventions.

We conceptualize reporting as the outcome of both demand-side factors – driven by children's underlying risk – and supply-side factors – driven by reporting behavior, discretion, and organizational practices. Disentangling the relative roles of demand- and supply-side factors in child maltreatment reporting is difficult for two related reasons. First, doing so requires data that track not only children who are reported to CPS, but also the population who could have been reported but were not. Most administrative data are drawn from CPS records and therefore condition on reporting itself. Second, children are not randomly assigned to the environments in which reporting decisions occur. As a result, observed differences in reporting rates may reflect selection on underlying maltreatment risk rather than differences in reporters' behavior, discretion, or institutional practices.

We address the first challenge by focusing on child maltreatment reporting in public schools. We use administrative, student-level data from Michigan public schools – both traditional and charter – linked to child welfare reports made to the Department of Health and Human Services' CPS. Our sample includes the universe of reports between the 2017–18 and 2023–24 school years, including reports that are not screened in for investigation. Importantly, CPS intake screeners record the reporter's role, allowing us to distinguish calls made by school staff from those made by other professionals or family members.

Public schools provide a useful setting for studying child maltreatment reporting for several reasons. First, schools account for roughly three-quarters of the variation in CPS reporting rates across census tracts.[2] Second, a substantial share of all maltreatment reports originate with school staff (17 percent in our data), the largest among all mandated reporters. Third, the school setting allows us to combine rich administrative data from education and child welfare systems to control flexibly for student risk.

We begin by documenting substantial variation across schools in maltreatment reporting. In the median elementary school, approximately 28 percent of students are reported to CPS between kindergarten and fifth grade. At the 10th percentile of the school distribution, fewer

[2] We compare the $R^2$ from regressions where the dependent variable is an indicator for whether a K–5 student was reported to CPS in year $t$ (by any reporter type) and the predictors are census tract indicators. We estimate these regressions using the student-year panel that forms our main estimation sample, described in Section 3.C. After residualizing the outcome with respect to school fixed effects, the tract-level $R^2$ falls by roughly 80 percent, indicating that most tract-level variation in reporting reflects differences across schools rather than differences across tracts within schools.

than 11 percent of students are reported, while at the 90th percentile, more than 47 percent are reported.

To disentangle student risk from institutional tendencies, we adapt value-added methods developed primarily to estimate the effects of schools on test scores and other noncognitive outcomes (e.g., Angrist, Hull and Walters, 2023; Angrist et al., 2016, 2026; Deming, 2014; Jackson et al., 2020). Specifically, we construct residualized measures of CPS reports that control for a rich set of student, school, and neighborhood characteristics and apply empirical Bayes shrinkage to account for estimation error. Our setting is well suited to this approach because the administrative data allow us to flexibly control for students' prior exposure to the child welfare system, as well as their educational characteristics and trajectories.

We focus on schools as a whole rather than teachers for two reasons. First, the CPS data identify the role of the person who formally initiated the report to CPS (e.g., a teacher or police officer), but in many schools this may be a social worker or administrator rather than the staff member who first raised the concern. Second, anecdotal evidence suggests that school-wide staff such as principals, nurses, and social workers play an important role in determining a school's approach to maltreatment reporting. Using a large "connected set" of teachers for whom reporting tendencies are identified by moves across grades or schools, we validate this claim empirically and show that there is little variation in reporting tendencies across teachers at the same school.

We assess the validity of our estimates using tests analogous to those employed in mover-based designs that separate supply- and demand-side factors in other institutional contexts (e.g., Chetty, Friedman and Rockoff (2014), Finkelstein, Gentzkow and Williams (2016)). Specifically, we show that when students switch to schools with higher reporting tendencies, reports made by school staff increase sharply, with no evidence of differential pre-trends. In contrast, reports made outside of school (e.g., reports that occurred during the preceding summer) do not change. Together, these patterns indicate that our measure of school reporting tendencies is forecast-unbiased and isolates meaningful differences in school reporting practices rather than differences in student risk, allowing us to quantify the relative contributions of supply- and demand-side factors.

The distribution of school reporting tendencies is economically meaningful. A one standard deviation increase in "reports-added" raises a student's probability of being reported to CPS by approximately 57 percent in a given school year. Combining the estimated dispersion in reports-added with the overall dispersion in school reporting rates implies that approximately 40 percent of the variation in reporting across schools reflects supply-side differences in reporting behavior rather than differences in students' underlying risk. These results suggest that institutional practices are a major determinant of children's exposure to child welfare reporting.

We also examine which institutional features are associated with school reporting tendencies. Higher reporting is systematically related to observable characteristics of school staffing, resources, and student composition. For example, high-reporting schools are significantly more likely to employ a school nurse, consistent with greater capacity to identify and respond to concerns about student well-being. At the same time, reporting tendencies are negatively correlated with traditional measures of school quality: schools with higher reports-added tend to have lower attendance and test score value added. These patterns suggest that schools' approaches to identifying and responding to student needs do not necessarily align with traditional measures of academic effectiveness.

There are two main reasons why institutions may have higher reports-added. Higher-reporting schools may be more willing to report cases they view as borderline, or they may be more effective at identifying children whose cases would warrant CPS intervention. These reasons have fundamentally different implications for welfare and policy. If variation reflects differences in reporting "preferences," then standardizing reporting practices could reduce unnecessary intervention while preserving detection of serious cases. However, if schools differ in their ability to identify children at risk, then uniform reporting rules may reduce welfare by constraining more accurate schools. Distinguishing between these explanations is challenging because observed reporting behavior alone does not reveal whether schools differ in how they assess risk or in how they act on that assessment. The same reporting rate could arise from different combinations of diagnostic skill and reporting preferences. As we discuss below, this distinction is central to ongoing debates over the design of child maltreatment reporting policies.

Before turning to a structural framework that allows us to separate these mechanisms, we provide two pieces of reduced-form evidence suggesting that differences in reporting behavior are unlikely to reflect reporting preferences alone. First, after accounting for differences in underlying student risk, the relationship between schools' reporting rates and the share of reported cases that ultimately warrant CPS intervention is essentially flat. If schools differed only in their willingness to report borderline cases, higher reporting rates would generally be associated with lower-quality reports. Instead, higher-reporting schools generate reports that are, on average, no less likely to warrant CPS intervention, suggesting that they may also be more effective at identifying high-risk children. Second, we show that the vast majority of schools are dominated by another school that both identifies more cases that ultimately warrant CPS intervention and generates fewer unnecessary reports – a pattern that is difficult to reconcile with a model in which schools differ only in reporting preferences.

While these reduced-form patterns suggest that differences in skill are important in our setting, they do not reveal how much of the observed variation in reporting behavior reflects skill versus preferences. To answer this question, we draw on a literature that studies how differences in diagnostic skill and preferences shape high-stakes decisions (e.g., Abaluck et al. (2016), Currie and MacLeod (2017), Chan, Gentzkow and Yu (2022); see Currie, MacLeod and Musen (2026) for a review). Our setting is closely related to that studied by Chan, Gentzkow and Yu (2022), who examine variation in radiologists' diagnostic decisions. We adapt their framework and model maltreatment reporting as arising from a noisy signal about underlying risk and a threshold-based reporting rule. In the model, schools differ both in their ability to distinguish between children whose cases do and do not warrant CPS intervention, and in how they trade off the costs of false negatives and false positives. We refer to these dimensions as skill and preferences, respectively. Using data on schools' reporting rates and CPS responses to those reports – which provide information about the quality of the reports – we estimate school-level measures of both skill and reporting preferences.[3]

We find substantial heterogeneity across schools in both skill and reporting preferences. Differences in skill account for more than twice as much variation in risk-adjusted reporting behavior as differences in preferences. Moreover, skill and preferences are positively correlated: schools that are more skilled at identifying cases that warrant CPS intervention are also more willing to report cases given the same perceived level of risk. As a result, higher reporting rates do not simply reflect a greater willingness to report marginal cases. Instead, higher-reporting schools also appear to be more effective at identifying children whose cases warrant further intervention.

These distinctions have direct implications for policy. Reporting decisions involve a tradeoff between failing to identify children whose cases warrant CPS intervention and generating unnecessary CPS intervention. Following Chan, Gentzkow and Yu (2022), we define welfare as a weighted sum of these two types of classification errors, where the weights reflect the relative social costs of missed cases and unnecessary reports. Specifically, we treat the underlying state as whether a child's case would warrant CPS intervention if reported – measured in our baseline specification by whether the report would be screened in for investigation, and alternatively by whether it would ultimately be substantiated.

[3] We also implement the approach developed by Currie and MacLeod (2017) and Currie, MacLeod and Musen (2026) and obtain qualitatively similar conclusions.

This welfare measure evaluates schools relative to their statutory role as mandated reporters. Accordingly, it assesses the quality of reporting decisions based on whether schools identify cases that warrant CPS intervention while avoiding unnecessary reports. It is therefore a partial welfare criterion, abstracting from the downstream effects of child welfare interventions and the broader effects of attending schools with different reporting practices. We use the model estimates to evaluate counterfactual policies that target either reporting behavior or schools' ability to identify high-risk children. In our baseline specification, we use the welfare weights implied by schools' observed reporting decisions, but we also examine how the results vary across a broad range of alternative social planner preferences.

Our counterfactual analysis yields three main findings with direct implications for current policy debates. First, policies aimed at improving reporters' ability to identify high-risk cases – including recent proposals in several states and jurisdictions to enhance reporter awareness, education, training, and decision-support tools – generate significant welfare gains, increasing detection with only modest increases in false positives (e.g., see Casey Family Programs (2025)). Second, recent policies that seek to reduce or equalize reporting rates – for example, reforms in New York restricting anonymous reporting (New York State Legislature, 2024; Sarkar, 2024) and reforms in California, Colorado, and other states encouraging school personnel to reduce reports to CPS (Child Welfare Information Gateway, 2023; Jones, 2024; Safe & Sound, 2025) – generally reduce welfare by increasing the number of high-risk cases that go unreported. Finally, these conclusions are robust across a broad range of plausible welfare weights: reducing reporting improves welfare only when society places a very low cost on missed cases, whereas improving reporters' ability to distinguish high-risk from low-risk cases increases welfare under a broad range of plausible welfare weights.

**Related Literature.** This paper relates to three strands of research. Conceptually, our approach parallels a broader literature on institutional variation in the provision of services. Our framework is closely related to research, originating in health economics (e.g., Finkelstein, Gentzkow and Williams (2016), Badinski et al. (2023), Doyle and Staiger (2025)), that decomposes variation in service utilization into demand-side factors driven by individual characteristics and supply-side factors driven by institutional practices and organizational culture. By bringing these ideas to the study of child maltreatment, we provide new evidence on how institutional environments shape children's exposure to the child welfare system.

Our findings also contribute to the literature on whether disparities in child welfare system involvement are driven by differences in supply-side factors, such as reporting tendencies, biases, beliefs, or organizational culture, or by differences in underlying risk (e.g., Baron et al. (2024); Drake et al. (2023); Grimon and Mills (2025); Rittenhouse, Putnam-Hornstein and Vaithianathan (2022)). Reporting is a particularly important margin because it serves as the primary gateway to the child welfare system. Children cannot be investigated, receive services, or enter foster care unless someone first reports concerns to CPS. Moreover, prior work suggests that much of the racial disproportionality observed in downstream child welfare outcomes (e.g., foster care placement) originates at this initial point of contact with the child welfare system rather than in later decisions within the system (Baron, Goldstein and Ryan, 2023). Understanding the sources of variation in reporting behavior is therefore central to understanding broader disparities in child welfare system involvement.

This study is also related to the growing literature demonstrating how schools affect children's lives in many ways beyond academic instruction. Prior research shows that schools differ substantially in their disciplinary practices, their effects on arrest and teen pregnancy, and other non-academic outcomes, with these impacts often weakly correlated with schools' effects on test scores (Beuermann et al., 2023; Chetty, Friedman and Rockoff, 2014; Cullen, Jacob and Levitt, 2006; Jackson, 2018; Jackson et al., 2020; Petek and Pope, 2023; Rose, Schellenberg and Shem-Tov, 2025). We show that schools also differ systematically in how they engage with child welfare systems – a common and high-stakes point of institutional contact for children and families – and that this variation is driven in large part by supply-side institutional practices.

The remainder of the paper proceeds as follows. Section 2 provides background on the CPS investigation process and the role of school staff as mandated reporters. Section 3 describes the data and documents substantial variation in reporting across schools. Section 4 outlines our empirical framework for estimating school reporting tendencies. Section 5 presents estimates of school reporting tendencies, quantifies the relative contributions of supply- and demand-side factors, and presents validation evidence based on student movers. Section 6 describes and estimates a structural model that decomposes reporting behavior into differences in skill and preferences and uses the model to evaluate the welfare implications of alternative policies. Section 7 concludes.

# 2 Background

## 2.A The CPS Investigation Process

A family's involvement with the child welfare system typically begins when an individual files a report of suspected child maltreatment with CPS. Reports may be made by mandated reporters – such as educators, physicians, and law enforcement officers – as well as by family members and other community members. Once a report is filed, CPS determines whether to take further action, which may include an investigation, the provision of services or, in some cases, temporary or permanent placement of a child outside the home. While the core structure of this process is broadly similar across the United States, specific procedures and organizational features can vary across states.

Figure A1 summarizes the CPS process in Michigan, the context of this study. When individuals suspect child abuse or neglect, they contact a statewide CPS hotline. Calls are routed through a common phone number to one of two centralized intake centers, located in Detroit and Grand Rapids. Within each center, calls are assigned quasi-randomly to screeners who are on queue at the time, with no discretionary reassignment. Screeners then determine whether the allegations meet the statutory criteria for investigation. Cases that are screened in are referred to local CPS offices for further investigation.

Intake screeners collect information about the child and the alleged maltreatment and determine whether the report meets the statutory threshold for investigation. Calls last approximately 15 minutes on average, and roughly 40 percent are screened in. During the call, screeners gather information provided by the reporter, including the nature of the allegation, the identity and type of reporter (e.g., teacher, medical professional, family member), and information about the child and household. Screeners also have access to administrative records through the state's case management system, including prior CPS contact for the child and family, as well as the findings of past investigations and any foster care involvement. In addition, they may access other administrative systems to verify addresses, identify household members, and observe participation in public assistance programs. Screening decisions are guided by statutory criteria and agency guidelines that have generally required determining whether the reported conduct, if true, would meet the legal definition of abuse or neglect, and whether the circumstances indicate sufficient concern about the child's safety to warrant an investigation.

Reports that are screened out receive no further CPS action. Reports that are screened in are referred to a local county child welfare office for investigation, typically within 24 hours, and investigations must be completed within 30 days. Each county operates its own local office, and cases are assigned to investigators on a rotational basis. During the investigation, caseworkers gather information from multiple sources, which may include interviews with the child, caregivers, teachers, medical professionals, and other relevant individuals, as well as reviews of medical, educational, and prior CPS records. Investigators may also conduct home visits and assess the child's living environment and immediate safety needs.

Following the investigation, the caseworker determines whether there is sufficient evidence to substantiate the alleged maltreatment. Approximately one quarter of investigations result in substantiated findings, and unsubstantiated investigations typically conclude CPS involvement. Because substantiation decisions are based on information collected during a formal investigation rather than solely on the contents of the initial report, they reflect a broader information set than is available to intake screeners. Conditional on substantiation, the caseworker further evaluates the child's ongoing risk in the home. A subset of cases are classified as high risk and proceed to court involvement. Approximately three percent of all investigations result in foster care placement, generally reflecting cases in which children face imminent risk of maltreatment.

## 2.B Reporting by School Staff

School staff are mandated reporters, meaning they are legally required to report suspected child maltreatment to CPS.[4] Our data, described below, identify the role of the reporter, distinguishing reports filed by educational personnel from those filed by law enforcement, medical professionals, family members, and others. School-based staff constitute the largest group of mandated reporters, accounting for 17 percent of all reports (Panel A, Figure A2); only family members submit reports more frequently. Moreover, prior research shows that educators tend to report cases that would otherwise go unreported (Benson, Fitzpatrick and Bondurant, 2025). Within schools, teachers and school counselors each account for approximately one-third of school-based reports.

Panel B of Figure A2 shows that reports filed by school staff are screened out at higher rates than reports filed by law enforcement, though only modestly more often than reports filed by medical professionals. Sixty-one percent of education-based reports are screened out at intake, compared with 47 percent of reports filed by law enforcement and 58 percent of reports filed by medical professionals. Similarly, a smaller share of school-based reports are ultimately substantiated: 7 percent, compared with 13 percent of reports filed by medical professionals and 28 percent of those filed by law enforcement.

Despite being legally mandated, reporting decisions by school staff involve substantial discretion. Mandated reporters are required to report when they have "reasonable suspicion" that a child is being abused or neglected, a standard that does not require definitive proof but instead relies on observations, child disclosures, or other indicators of harm. Importantly,

[4] In Michigan, other mandated reporters include medical professionals, law enforcement officers, family court and welfare employees, clergy, and social workers.

school staff are not expected to investigate allegations or determine whether maltreatment has occurred. Rather, their role is to identify situations that may warrant further assessment by CPS. In practice, educators must decide whether observed behavior, injuries, or other contextual signals are sufficiently concerning to justify a report. These signals are often incomplete and difficult to interpret, and school staff typically have limited information about how CPS will evaluate a given case. Reporting decisions therefore require educators to assess risk under considerable uncertainty and determine whether a child's circumstances warrant referral to the child welfare system.

## 2.C Measuring Cases Warranting CPS Intervention

A central challenge in studying the quality of reporting decisions in Section 6 is defining the outcome that reporters are attempting to predict. School staff do not directly observe whether a child is experiencing maltreatment, nor are they responsible for determining whether maltreatment has occurred. Rather, mandated reporters are required to identify situations that may warrant further assessment by CPS. The CPS process described above therefore provides two useful measures of whether a reported concern ultimately warrants child welfare intervention.

The first is whether a report is screened in for investigation, which is the outcome most directly related to the reporting decision itself. The second is whether allegations are substantiated following investigation. Substantiation is determined later in the child welfare process by different decision-makers after the collection of substantially more information about the child and family. As a result, substantiation reflects a broader information set than that available at intake and provides a useful complementary outcome measure. We therefore examine both outcomes in our analysis.

A potential concern with using these outcomes is that they are themselves determined by CPS decision-makers. In particular, screeners may develop persistent beliefs about particular reporting sources that influence subsequent decisions independent of the content of a given report. Several features of Michigan's CPS process may limit the scope for such reputation effects. As mentioned above, reports are routed through two centralized intake centers that share a common hotline number and are assigned to intake workers based on queue position, creating quasi-random assignment of reports to screeners. Given the large number of elementary schools generating reports statewide (approximately 2,000), repeated interactions between particular schools and screeners are likely to be rare. Reports that are screened in are subsequently assigned to investigators through rotational assignment within local CPS offices, further limiting the extent to which school-specific reputations can systematically influence investigation outcomes.[5]

# 3 Data and Motivating Statistics

## 3.A Data Sources

We use administrative student-level data from Michigan public schools – both traditional and charter – linked to records of child welfare reports made to the Michigan Department of Health and Human Services CPS. The CPS data include all calls made between the 2017–18 and 2023–24 school years, including reports that are not screened in for investigation, and identify the role of the reporter, allowing us to distinguish reports filed by school-based staff from those filed by other reporters.

We focus on elementary school students for both conceptual and empirical reasons. First, initial school-age contact with the child welfare system typically occurs in elementary school: among children investigated by CPS between kindergarten and 12th grade, 77 percent experience their first investigation in K–5.[6] Second, focusing on elementary school provides an important identification advantage. Because many children have contact with CPS before entering school, we can use CPS records from before kindergarten entry – including data from years prior to 2017 – to construct predetermined measures of underlying maltreatment risk. Although these earlier records include only reports that were screened in for investigation and do not identify the reporter's role, they allow us to flexibly control for children's prior involvement with the child welfare system before exposure to the school reporting environments we study. Throughout the paper, we measure CPS involvement using only reports in which the student is identified as the alleged victim. We exclude reports in which the student is recorded only as another child in the household (e.g., a sibling of the alleged victim), so both current and prior CPS involvement reflects allegations concerning the student rather than another household member.

[5]A related concern is that schools may differ in how effectively they communicate concerns to CPS. For example, reports from some schools may contain more detail, context, or supporting information, increasing the likelihood that a case is screened in or ultimately substantiated. Such differences would affect the interpretation of our estimates, as the resulting variation could reflect both schools' ability to identify concerning cases and their ability to communicate information about those cases to CPS. At the same time, substantiation decisions are made after investigators collect substantial additional information through interviews with children, caregivers, and other relevant parties. As a result, similarities between findings based on screening and substantiation outcomes could be less likely to arise solely from differences in how initial reports are documented.

[6]To calculate this statistic, we restrict the sample to children who entered kindergarten in 2012 or earlier to ensure we observe their entire schooling trajectory.

## 3.B Matching Across Data Systems

CPS records are linked to Michigan public school administrative data by the Michigan Education Data Center (MEDC) using a probabilistic record linkage algorithm. Because the two datasets do not share a common unique identifier, the linkage relies on students' first and last names and dates of birth, using statewide K–12 enrollment records as the base population. The matching procedure assigns a posterior probability to each potential match, reflecting the likelihood that two records correspond to the same individual. MEDC applies post-processing rules and manual review to remove likely false positives, particularly in cases where names are similar but birth dates differ. Overall, 86.7 percent of CPS records for children in the birth cohorts included in our analysis that pass basic validation are successfully matched to a Michigan public school student record. This match rate is consistent with expectations, as CPS records include children who never enroll in Michigan public schools (e.g., those who attend private schools or are homeschooled).

## 3.C Analysis Sample

Our analysis sample includes all students observed at least once between kindergarten and fifth grade in a Michigan public school between the 2017–18 and 2023–24 school years. Because the COVID-19 pandemic substantially altered schooling environments and reporting behavior (Baron, Goldstein and Wallace, 2020), we exclude reports made between March 2020 and June 2021 when estimating school reporting tendencies. This exclusion removes the 2020–21 school year and limits the 2019–20 school year to the pre-pandemic period. The resulting sample contains 3,642,261 student-year observations covering 1,232,220 unique children. These students are involved in 872,071 CPS reports while enrolled in grades K–5, along with an additional 619,748 reports occurring outside the elementary school years, which we use to measure prior exposure to the child welfare system.

Table 1 describes the sample used to estimate the value-added models that identify school reporting tendencies. Eleven percent of students experienced at least one CPS report prior to entering kindergarten. By the end of fifth grade, 21 percent had experienced a report, and 9 percent had experienced at least one substantiated investigation. These rates vary substantially across groups. By the end of fifth grade, 31 percent of Black students had been the subject of a CPS report and 13 percent had experienced a substantiated investigation, compared with 19 percent and 8 percent, respectively, among White students. Students who are reported are also disproportionately economically disadvantaged, more likely to receive special education services, and more likely to be chronically absent.

Table A1 describes the school and neighborhood environments of students in our sample. Students who are ever reported to CPS attend lower-performing schools, with lower average standardized test scores in both reading and math, and are substantially more likely to attend Title I schools. They also live in more disadvantaged neighborhoods, characterized by lower household incomes, higher unemployment rates, and lower marriage rates.

### 3.D Variation Across Schools in Reporting

The share of students ever reported to CPS during elementary school varies widely across schools. In Figure 1, we assign each student to the school they attended in kindergarten and calculate, for each school, the fraction of students who receive a CPS report between kindergarten and fifth grade.[7] At the median school, 28 percent of students receive at least one CPS report during elementary school, compared with 11 percent at the 10th percentile and 47 percent at the 90th percentile of the school distribution. Similar variation is evident when we restrict attention to reports filed by school staff (Panel A of Figure A3). At the median school, 9 percent of students receive a school-based CPS report between kindergarten and fifth grade, compared with 3 percent and 20 percent at the 10th and 90th percentiles, respectively.

These differences are also evident across demographic groups. At schools at the 90th percentile of the reporting distribution, 44 percent of White students, 57 percent of Black students, and 66 percent of economically disadvantaged students are reported to CPS between kindergarten and fifth grade. These disparities are evident for reports filed by both school-based and non-school-based reporters (Panels A and B of Figure A3).

## 4 Identifying School Reporting Tendencies

Our goal is to obtain unbiased estimates of each school's reporting tendency, defined as the propensity to report students to CPS holding fixed students' underlying risk. We refer to these estimates as *reports-added*. The ideal experiment we seek to approximate is one in which students are randomly assigned to schools, allowing us to ask whether a given child would be more likely to be reported if they attended school A rather than school B.

A naive approach would measure a school's reporting tendency using, for example, its average number of reports per student. However, as documented above, higher- and lower-reporting schools tend to serve populations of students who differ in many ways. As a

[7]We restrict the sample to students observed from kindergarten through fifth grade, so they must have entered kindergarten in 2017 or 2018. As a result, overall reporting rates differ from those reported in Table 1, which can include students observed for only part of their elementary school years during the sample period.

result, simple school-level reporting rates conflate differences in student risk with differences in school reporting behavior and cannot separately identify variation in reporting practices across schools.

We adapt methods from the school value-added literature (e.g., Angrist, Hull and Walters (2023); Angrist et al. (2016, 2026); Deming (2014); Jackson et al. (2020); Jacob, Lefgren and Sims (2010)) to isolate school-level reporting behavior. We focus on estimating school-level rather than teacher-level reporting tendencies for several reasons. First, although the CPS data identify the reporter's role (e.g., teacher, counselor, administrator), they do not identify the individual who made the report – e.g., the exact teacher. Second, the person who formally files the report may not be the staff member who first raised the concern. For example, a teacher may refer a case to a counselor or administrator who subsequently contacts CPS. Third, our data do not contain student-teacher linkages, which would preclude estimation of standard teacher value-added models even if reporter identities were observed. Finally, conversations with school staff and CPS workers suggest that reporting practices are often shaped by school-wide factors, including staff training, reporting protocols, and guidance provided by counselors or administrators. These considerations suggest that the school, rather than the individual teacher, may be the appropriate unit of analysis.

Although our primary analysis focuses on schools, we can also assess the extent of within-school variation in reporting behavior. Following Biasi (2021), we exploit school-grade variation and identify teacher-level reports-added measures using teachers who switch schools or grades.[8] We find little evidence that reporting tendencies persist within teachers over time, and the dispersion in teacher reports-added is substantially smaller than the dispersion in school reports-added.[9] Taken together, these findings suggest that persistent differences in reporting behavior arise primarily across schools rather than across teachers within schools.

A key concern with value-added models is that available controls may not fully capture

[8]The Michigan education data do not contain direct teacher-student links, only teacher-grade links. We therefore estimate Equation 1 below including school-by-year rather than school fixed effects and calculate average residual reports at the school-grade-year level. We assign these mean residuals to each teacher working in the corresponding school-grade-year cell and then shrink teacher-level reports-added estimates using the empirical Bayes procedure described below. As in Biasi (2021), teacher reports-added is identified by teachers who move across grades or schools. For example, if teachers A and B both teach third grade in year $t$, and teacher B moves to fourth grade in year $t+1$, then the reports-added measures for teachers A and B become separately identified. More generally, reports-added is identified for any teacher who is connected to other teachers through moves across grades or schools. The median school-grade team contains only three teachers, and teacher mobility is common: 23 percent of teacher-year observations involve either a school or grade switch. Consequently, the teacher mobility network is highly connected, allowing us to estimate teacher-level reports-added measures for approximately 80 percent of teachers.

[9]The estimated within-teacher autocorrelation is $-0.10$, and the estimated standard deviation of teacher reports-added is 0.0021, compared with 0.0119 for school reports-added.

unobserved student "risk," leaving the resulting estimates vulnerable to omitted-variable bias. Our setting offers several advantages in this regard. First, the linked data allow us to control for a set of student-, school-, and neighborhood-level characteristics that are likely to be related to underlying maltreatment risk. In particular, we observe children's involvement with the child welfare system *prior* to entering public school, as well as neighborhood-level reporting rates from non-school-based reporters. These measures capture important sources of variation in underlying maltreatment risk that are typically unavailable in administrative education data. Second, the institutional context may make sorting on school reporting practices less likely than sorting on other dimensions of school quality. Families frequently consider observable characteristics such as academic performance, student demographics, and proximity when choosing schools (Abdulkadiroğlu et al., 2020; Bayer, Ferreira and McMillan, 2007; Campos and Kearns, 2024). In contrast, because child maltreatment reporting is generally confidential, differences in reporting practices are less visible and therefore may be less likely to influence school choice. While this observation is certainly not sufficient to rule out sorting on reporting behavior, it suggests that such sorting could be less pronounced than sorting on more readily observable school characteristics. Finally, we assess the validity of our approach using a series of tests described below.

We estimate static measures of school-level child welfare "reports-added" following the approach and notation in Kane and Staiger (2008). For student $i$ attending school $j$ in year $t$, we model the latent propensity to receive a school-based child welfare report, $R^*_{ijt}$, as a function of a vector of student-, school-, and neighborhood-level covariates $X_{ijt}$; a school-specific reporting tendency, $\mu_j$; a school-year shock, $\theta_{jt}$; and an idiosyncratic student-year shock, $\varepsilon_{ijt}$:

$$R^*_{ijt} = X'_{ijt}\beta + \mu_j + \theta_{jt} + \varepsilon_{ijt}. \tag{1}$$

Our object of interest is $\mu_j$, which captures school $j$'s contribution to reporting behavior, holding fixed observable determinants of student risk.

We first construct residualized measures of child welfare reporting, $\hat{R}_{ijt}$, for each student-year by partialling out observed covariates:

$$\hat{R}_{ijt} \equiv \hat{\mu}_j + \hat{\theta}_{jt} + \hat{\varepsilon}_{ijt} = R^*_{ijt} - X'_{ijt}\hat{\beta}. \tag{2}$$

Table A2 lists the covariates included in $X_{ijt}$. We control for a set of student characteristics, including race/ethnicity, gender, economic disadvantage, English Language Learner status, and special education status, as well as demographic characteristics of the student's home Census tract. We also include leave-out school-level averages of these student and neighborhood characteristics to capture peer composition. Finally, we include county-by-year fixed effects to account for variation in reporting and investigation practices across geography and CPS county offices over time. Importantly, we control for students' prior exposure to the child welfare system by including lagged counts of CPS reports and substantiated investigations received by student $i$ in previous years, including those occurring prior to kindergarten entry. These measures include reports filed by a broad set of reporters, including school personnel, family and community members, law enforcement, and medical professionals, and serve as proxies for baseline maltreatment risk.

We further include a control for the leave-out average number of *non–school-based* reports among students attending the same school in the prior year, defined as the average number of reports per student filed by non-school reporters over the preceding 12 months (including the summer). This measure captures underlying student risk and community reporting behavior that is plausibly distinct from school reporting practices. Using lagged non-school-based reports also mitigates concerns about contemporaneous shocks that may jointly affect school- and non-school-based reporting. We do *not* control for a student's own non-school-based reports or the leave-out average number of non-school-based reports during the contemporaneous school year, as school- and non-school-based reports may be jointly determined and therefore represent competing risks.

We also do *not* include school-level covariates that might be correlated with any "supply-side" determinants of reporting. For example, we do not include measures of support staff (e.g., nurses, social workers) who work in a particular school or characteristics of the educators in the school (e.g., race or gender of teachers or administrators). We later explore whether these factors are correlated with our measures of reports-added.

The regressions include school fixed effects to account for correlations between unobserved, time-invariant school characteristics and reporting outcomes.[10] However, we exclude these fixed effects when constructing the residuals in Equation 2, so that variation across schools is retained in the residualized outcome. As a result, $\hat{R}_{ijt}$ decomposes into a persistent school-level component $\mu_j$, a transitory school-year shock $\theta_{jt}$, and an idiosyncratic student-year shock $\varepsilon_{ijt}$.

If a student has been reported to CPS previously by staff in the same school, educators may respond differently to subsequent concerns involving that same student. Including later reports for the same student in the same school could therefore complicate the interpretation

[10]Including school fixed effects allows us to estimate the coefficients on observed predictors conditional on unobserved, time-invariant school factors that might otherwise bias these estimates.

of school reporting tendencies. Accordingly, we estimate models excluding all student-year observations following a student's first CPS report at a given school, akin to measuring the hazard of being reported at least once by the school.[11]

We then construct a school's reports-added by shrinking school average residuals towards zero, the "grand mean" of the distribution. We assume that the unobserved components of reporting, $\{\mu_j, \theta_{jt}, \varepsilon_{ijt}\}$, are uncorrelated, meaning we can decompose the total variance in residual reports into the sum of each variance term: $\mathrm{Var}(\hat{R}_{ijt}) = \hat{\sigma}_\mu^2 + \hat{\sigma}_\theta^2 + \hat{\sigma}_\varepsilon^2$. We estimate $\hat{\sigma}_\mu^2$ using the covariance of average school residuals across adjacent years. The year-to-year correlation in residual reports per school is approximately 0.56, indicating substantial persistence in school reporting behavior over time. We estimate $\hat{\sigma}_\varepsilon^2$ using within-school-year variation in residual reports. $\hat{\sigma}_\theta^2$ is then the total variance minus $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_\mu^2$.

We define the school's average residual reports as a precision-weighted average of school-year residuals: $\bar{\mu}_j = \sum_t w_{j,t}\mu_{j,t}$. The precision weight for each school-year, $w_{j,t}$, is based on the inverse sampling variance for school $j$ in year $t$:

$$w_{j,t} = \frac{h_{j,t}}{\sum_t h_{j,t}} \quad \text{with} \quad h_{j,t} = \frac{1}{\hat{\sigma}_\theta^2 + (\hat{\sigma}_\varepsilon^2 / n_{j,t})} \tag{3}$$

where $n_{j,t}$ is the school's enrollment in year $t$. Given these weights, the variance of $\bar{\mu}_j$ can be decomposed into the variance attributable to true differences in reporting tendencies across schools and the within-school variance due to sampling error:

$$\mathrm{Var}(\bar{\mu}_j) = \underset{\{\text{signal}\}}{\hat{\sigma}_\mu^2} + \underset{\{\text{noise (within-school)}\}}{\mathrm{Var}(\bar{\mu}_j|\mu_j)} \tag{4}$$

where $\mathrm{Var}(\bar{\mu}_j|\mu_j) = \left(\sum_t h_{jt}\right)^{-1}$.[12] From this, we obtain the shrunken empirical Bayes estimate of the school's reporting tendency, which we denote $RA_j$:

$$RA_j = \bar{\mu}_j \left( \frac{\hat{\sigma}_\mu^2}{\mathrm{Var}(\bar{\mu}_j)} \right) \tag{5}$$

Our measure of reliability, $\frac{\hat{\sigma}_\mu^2}{\mathrm{Var}(\bar{\mu}_j)}$, captures the share of the total across-school variance in average residual reports attributable to variance in true school reporting tendencies. Because $h_{j,t}$ is increasing in $n_{j,t}$, this means that estimates for larger schools are shrunk less, since $\bar{\mu}_j$ is a more precise signal of their true reporting behavior.

[11] If the student switches schools, we include observations from the new school.

[12] Given $\mu_{j,t} = \mu_j + \upsilon_{j,t}$ where $\upsilon_{j,t} \sim (0, \hat{\sigma}_\theta^2 + (\hat{\sigma}_\varepsilon^2/n_{j,t}))$, then assuming that $\upsilon_{j,t}$ is uncorrelated across years, and denoting $H \equiv \sum_t h_{jt}$, $\mathrm{Var}(\bar{\mu}_j|\mu_j) = \frac{1}{H^2}\sum_t h_{j,t}^2 \mathrm{Var}(\mu_{j,t}) = \frac{1}{H^2}\sum_t (\hat{\sigma}_\theta^2 + (\hat{\sigma}_\varepsilon^2/n_{j,t})) = \frac{1}{H^2}\sum_t h_{j,t}^2 \frac{1}{h_{j,t}} = \frac{1}{H^2}\sum_t h_{j,t} = \frac{H}{H^2} = \frac{1}{H}$.

# 5 Estimates of School Reporting Tendencies

Figure 2 presents the distribution of estimated school-level reporting value added ("reports-added") under three specifications. Panel (a) reports estimates based on school staff reports with no additional controls. Panel (b) adds student, school, and neighborhood characteristics, and Panel (c) further incorporates controls for students' prior exposure to the child welfare system through lagged reports. Note that in each case the reporting value-added is "shrunk" and de-meaned in the way described above.

Across specifications, the distribution of reports-added exhibits substantial dispersion, indicating meaningful heterogeneity in reporting tendencies across schools. At the same time, the inclusion of controls shifts schools' relative positions: the correlation between estimates with no controls and those from the fully controlled specification is 0.56. This highlights the importance of accounting for differences in underlying student risk when measuring institutional reporting behavior.

In our preferred specification, which includes the full set of controls, the standard deviation of reports-added is 0.0119. In other words, a one standard deviation increase in reports-added raises a student's probability of receiving a CPS report from school staff by approximately 1.2 percentage points in a given year – an increase of about 57 percent relative to the sample mean annual reporting rate of 2.1 percent. This magnitude indicates that differences in school reporting tendencies are economically meaningful.

**Validation of Reports-Added Estimates.** We validate our reports-added estimates by examining whether students' reporting outcomes change when they move between schools with different reporting tendencies. Intuitively, if reports-added captures causal differences in school reporting behavior, then students who move to higher-reporting schools should experience an increase in school-based CPS reports at the time of the move. Moreover, if school moves are unrelated to the relative reporting rates across the origin and destination schools, then we would expect no differential trends prior to the move.

To test this, we estimate event studies that compare the evolution of reporting outcomes for students who switch from origin school $j(0)$ to destination school $j(1)$ in year $t$ in grade $g$:

$$\widehat{R}_{i,g,t} = \sum_{k \in \{-4,\ldots,3\}} \beta_k \, \text{event}^k_{i,t} \times \Delta\widehat{RA}_{j(1),j(0)} + \Delta\widehat{RA}_{j(1),j(0)} + \gamma_g + \gamma_t + \varepsilon_{i,g,t}. \tag{6}$$

The dependent variable, $\widehat{R}_{i,g,t}$, is an indicator for whether student $i$ receives a school-based CPS report in year $t$, residualized for observable characteristics as in Equation 1. The key predictor is the difference in reports-added between the destination and origin school, $\Delta\widehat{RA}_{j(1),j(0)}$. The event time indicators, $\text{event}_{i,t}^{k}$, denote years relative to the move. The coefficients of interest, $\beta_k$, trace out how reporting changes in event time relative to the school switch, scaled by the difference in reports-added between the destination and origin school, $\Delta\widehat{RA}_{j(1),j(0)}$. We estimate this specification on a sample of 106,103 students who switch schools exactly once during elementary school, in second through fourth grade.[13]

Interacting this change in reports-added with event-time indicators allows us to test whether reporting outcomes respond to changes in school reporting tendencies at the time of the move, and whether there are differential pre-trends for students who move to higher-reporting schools. We include grade and year fixed effects, along with the main effect of $\Delta\widehat{RA}_{j(1),j(0)}$, to absorb time-invariant differences in reporting outcomes associated with moves to higher-reporting schools.[14]

Panel (a) of Figure 3 presents the event-study estimates. In the year of the move, the estimated coefficient is statistically indistinguishable from one (1.06), indicating that students who move to schools with higher reports-added experience a nearly one-for-one increase in their own likelihood of being reported. In contrast, the coefficients on the pre-move event indicators are small and statistically insignificant, providing no evidence of differential pre-trends prior to the school switch.[15]

We further assess this relationship using binscatter plots that relate changes in students' reporting outcomes to the changes in reports-added induced by school moves. Panel (b) shows that the relationship is approximately linear, with an estimated slope of 1.00, supporting the assumption that reporting outcomes can be modeled using additively separable school and time effects (Finkelstein, Gentzkow and Williams, 2016).

[13]To avoid mechanical correlation between the outcome and the right-hand-side variable, we re-estimate reports-added using only students who never switch schools and use this as our independent variable. The resulting measure is similar to our baseline estimate (correlation = 0.913). More generally, because we do not include student fixed effects, our reports-added estimates do not rely on student movers for identification. Nevertheless, students are quite mobile, with 14.6% of student-year observations consisting of a move. Thus, the network of schools connected via student moves contains nearly all schools (99.5 percent) and students (99.99 percent).

[14]We do not restrict to the hazard sample in these models, since that sample mechanically contains no reports prior to event time $t-1$.

[15]Note that the estimated effect of switching to a higher-reporting school declines in subsequent years. Because our sample includes only students who switch schools once during elementary school, this pattern cannot reflect additional school moves. One interpretation is that schools become less likely to file subsequent reports after first reporting the same child to CPS. A second possibility is that the pattern reflects the unbalanced structure of our panel, since students who switch in fourth grade are observed for only two years after the move, whereas second-grade switchers are observed for four years. However, we find little evidence for this second explanation. Figure A4 shows that the post-switch time path is similar for students who move in second, third, or fourth grade.

We also test whether reports-added captures broader community reporting behavior – such as reporting by law enforcement or medical professionals – rather than school-specific reporting practices. To do so, we examine changes in reports filed by non-school reporters during the summer preceding the school year. As shown in Panels (c) and (d) of Figure 3, the relationship between changes in reports-added and summer reporting is essentially flat: for example, the estimated binscatter slope is 0.006 (standard error = 0.071). This placebo test suggests that reports-added does not capture changes in underlying maltreatment risk or broader community reporting behavior around the time of the move, but instead reflects school-specific reporting practices.

Finally, we estimate the extent to which cross-school variation in observed CPS reporting reflects persistent differences in schools' reporting tendencies rather than differences in underlying student risk. Because observed reporting rates combine both components, we estimate the coefficient from a regression of school reports-added, $\mu_j$, on schools' unadjusted reporting rates, $\overline{r}_j$: $\widehat{\beta}_\mu = \frac{\widehat{\mathrm{Cov}}(\mu_j, \overline{r}_j)}{\widehat{\mathrm{Var}}(\overline{r}_j)}$. Intuitively, $\widehat{\beta}_\mu$ measures the share of cross-school variation in observed reporting rates that is attributable to persistent differences in schools' reporting tendencies. We estimate both the covariance and variance using the same empirical Bayes framework used to estimate the variance of school reports-added. Applying this estimator yields $\widehat{\beta}_\mu = 0.41$, implying that approximately 41 percent of the variation in school reporting reflects persistent differences in schools' reporting tendencies, while the remaining 59 percent reflects differences in underlying student risk.

As a robustness check, we also implement the mover-based decomposition proposed by Finkelstein, Gentzkow and Williams (2016), who were interested in understanding geographic variation in healthcare spending. Specifically, we re-estimate the mover event study and binscatter using an indicator for receiving any school-based CPS report in year $t$ as the outcome and the school's unadjusted reporting rate as the measure of school reporting, which we also shrink using empirical Bayes methods for consistency (Figure A5). As in our main mover analysis, the event study exhibits a sharp discontinuous increase in reporting immediately following a school move, with no evidence of differential pre-trends. We likewise find no discontinuity in reports filed by non-school reporters during the preceding summer, supporting the identifying assumption that underlying student risk does not change discontinuously at the time of the move. Because unadjusted reporting rates reflect both school reporting practices and underlying student risk, however, the estimated jump is substantially smaller than the corresponding estimate using reports-added. As discussed in Finkelstein, Gentzkow and Williams (2016), under this identifying assumption, the size of the discontinuity identifies the share of cross-school variation attributable to supply-side factors rather than differences in student risk. The estimated discontinuity implies that approximately 41 percent of the variation in school-based reporting is attributable to differences in school reporting practices. Similarly, the estimated binscatter slope is 0.37 (SE = 0.06), with both estimates closely matching our preferred covariance-based estimate.

To summarize, students who move to schools with higher reporting tendencies experience an immediate increase in school-based CPS reports, with no evidence of differential pre-trends and a linear relationship between changes in reporting outcomes and changes in reports-added. The near one-for-one relationship between changes in reports-added and changes in reporting outcomes implies that our estimates are "forecast-unbiased" (Chetty, Friedman and Rockoff, 2014). We also show that these effects are specific to school-based reporting and do not reflect broader community reporting norms. Taken together, these results indicate that reporting behavior accounts for approximately 40 percent of the variation in student reports across schools.

**Characteristics of High-Reporting Schools.** Table 2 describes bivariate associations between school characteristics and school-level reporting behavior. Specifically, we regress average school-year residual reports on individual school "supply-side" characteristics – such as staff demographics, resources, and measures of school performance – one at a time. We use residual (non-shrunken) average school-year reports rather than reports-added estimates as the outcome because value-added estimates are unbiased predictors but not unbiased estimates of true school effects (Jacob and Rothstein, 2016). Column 1 reports sample means, while Column 2 presents coefficients from bivariate regressions that include county-by-year fixed effects.

Several broad patterns emerge. First, reporting is strongly related to schools' institutional capacity to observe and respond to student needs. Schools that employ a school nurse – though relatively uncommon – exhibit higher reporting rates. Second, reporting is systematically related to traditional measures of school performance: schools with lower test score value added tend to report more frequently. These patterns suggest that schools' approaches to identifying and responding to student needs do not necessarily align with traditional measures of academic effectiveness. Third, reporting varies with staff composition. Schools with higher shares of female educators and non-White teachers and administrators tend to report more frequently. Finally, reporting is negatively associated with teacher experience and formal credentials. Despite these descriptive patterns, observable school characteristics explain relatively little of the variation in reporting behavior. The characteristics in Table 2 jointly account for only about 9 percent of the variation in reporting tendencies, indicating that most differences in reporting across schools are driven by factors that are not captured by standard administrative measures.

# 6 Welfare Implications of Reporting Differences

The results above establish that institutional factors play an important role in shaping children's exposure to child welfare reporting, but they do not reveal why some schools report more than others. Higher-reporting schools may be more willing to report borderline cases, or they may be better at distinguishing children whose circumstances warrant CPS intervention from those whose circumstances do not. Distinguishing between these explanations is important because they imply different policy responses. If variation primarily reflects differences in reporting preferences, then standardizing reporting practices may improve welfare. If variation instead reflects differences in schools' ability to identify children whose cases warrant intervention, then constraining high-reporting schools may be counterproductive.

A useful way to frame this question is through the lens of statistical classification. Schools do not directly observe whether a child's circumstances warrant CPS intervention. Instead, they observe noisy and incomplete signals – such as behavioral indicators, student disclosures, and contextual information – and must decide whether these signals are sufficiently strong to warrant a report. From this perspective, reporting can be viewed as a two-step decision process. First, schools evaluate a case and form an assessment of the likelihood that the child's circumstances warrant CPS intervention. Second, they decide how to trade off the costs of missed cases against unnecessary reports and choose a reporting threshold based on that assessment. Differences across schools may therefore arise either because schools differ in their ability to assess risk or because they differ in how they trade off the costs of missed cases against unnecessary reports.

We adapt the framework developed by Chan, Gentzkow and Yu (2022) (CGY) to separate these mechanisms. The CGY framework is well suited to settings where researchers observe downstream outcomes that reveal decision quality and decision-makers face comparable cases (Currie, MacLeod and Musen, 2026).[16] Our setting shares these features. We observe the outcomes of reports, including whether they are screened in for investigation and whether they are ultimately substantiated by a CPS investigator. These outcomes provide information about the quality of reported cases. In addition, our value-added analyses

[16] In Appendix D, we show that the reduced-form approach motivated by Currie and MacLeod (2017) and Currie, MacLeod and Musen (2026) yields qualitatively similar conclusions.

isolate variation in school reporting behavior from differences in underlying student risk, allowing us to compare schools facing similar underlying cases.

## 6.A Reporting as a Statistical Classification Problem

**Setup.** We conceptualize reporting decisions as a statistical classification problem in which both the decision and the underlying state of the world are binary, and the agent does not observe the true state when making the decision. In our setting, the agent is a school, the decision is whether to report a child to CPS, and the underlying state is whether the child's circumstances warrant CPS action. In our baseline specification, we define this state using whether a report is screened in for formal investigation. We also estimate the model using a more stringent outcome – whether allegations are substantiated following an investigation – and obtain similar conclusions. We emphasize that all school-level rates discussed in this section – including both reporting rates and the outcomes of those reports (i.e., screen-in and substantiation rates conditional on reporting) – are risk-adjusted. As a result, differences across schools reflect variation in reporting behavior holding student risk constant, rather than differences in the underlying risk of maltreatment.[17]

Let $d_{ij} \in \{0, 1\}$ denote whether school $j$ reports child $i$ to CPS, and let $s_i \in \{0, 1\}$ denote whether child $i$'s case would warrant investigation if reported. Each child has an underlying latent risk, $\nu_i$, and the underlying state $s_i$ is modeled as $s_i = \mathbf{1}\{\nu_i > \bar{\nu}\}$. Schools receive a noisy signal of student risk ($x_{ij}$) and report cases when this signal exceeds the school's reporting threshold $\tau_j$: $d_{ij} = \mathbf{1}(x_{ij} > \tau_j)$.

This framework gives rise to four possible outcomes, which are illustrated in Figure A6:

$$TP_j \equiv \Pr(d_{ij} = 1, s_i = 1), \quad FP_j \equiv \Pr(d_{ij} = 1, s_i = 0),$$

$$TN_j \equiv \Pr(d_{ij} = 0, s_i = 0), \quad FN_j \equiv \Pr(d_{ij} = 0, s_i = 1).$$

These quantities represent population probabilities (or rates) for school $j$. Let $P_j = TP_j + FP_j$ denote the share of cases school $j$ reports, and let $S_j = TP_j + FN_j$ denote the prevalence of cases that would warrant investigation in the population served by school $j$.

[17]Note that our unit of analysis throughout is the school rather than individual staff members. Accordingly, we interpret the latent signal, reporting threshold, and resulting measures of skill as characterizing the school's reporting process rather than the information or reporting preferences of any particular staff member. Although reporting decisions may involve multiple individuals – including teachers, counselors, and administrators – our estimates summarize the school's overall reporting behavior. This interpretation is supported by the evidence in Section 4, which shows that reporting behavior varies substantially across schools but relatively little across teachers within the same school.

**ROC curves.** A standard way to summarize performance in the statistical classification literature is through the receiver operating characteristic (ROC) curve, which plots the true positive rate,

$$TPR_j = \Pr(d_{ij} = 1 \mid s_i = 1) = \frac{TP_j}{TP_j + FN_j},$$

against the false positive rate,

$$FPR_j = \Pr(d_{ij} = 1 \mid s_i = 0) = \frac{FP_j}{FP_j + TN_j}.$$

In our setting, the true positive rate measures the share of children whose cases would warrant investigation that $j$ actually reports, while the false positive rate measures the share of children whose cases would *not* warrant investigation that $j$ nonetheless reports.

Figure 4 illustrates how differences in skill and preferences appear in ROC space. For a given school with a fixed skill at identifying likely maltreatment, varying its reporting threshold traces out a set of $(FPR_j, TPR_j)$ pairs along a frontier of achievable classification outcomes. At one extreme, a school that reports no children lies at $(0, 0)$, while one that reports all children lies at $(1, 1)$. A school that perfectly classifies its students would lie at $(0, 1)$: it reports all children whose cases warrant investigation and does not report any children whose cases do not. Under a threshold-based reporting rule, increasing the propensity to report moves the school along this curve, raising both the true positive rate and the false positive rate. The ROC curve therefore summarizes the trade-off between the true positive rate and the false positive rate, and traces out the frontier of achievable classification outcomes associated with a given level of skill. Schools with more informative signals operate on higher ROC curves, achieving higher true positive rates for a given false positive rate.

Panel A of Figure 4 holds skill fixed and varies preferences, so all schools lie on a common ROC curve but choose different operating points along it. Schools $j$, $j'$, and $j''$ operate on the same curve, indicating equal skill, but differ in their reporting behavior: moving from $j$ to $j''$ corresponds to a greater willingness to report, resulting in both higher true positive rates and higher false positive rates. In contrast, Panel B holds preferences fixed and varies skill, so schools operate on different ROC curves. Schools $j$, $j'$, and $j''$ choose points corresponding to tangencies between their ROC curves and a common set of indifference curves. As skill increases, these tangency points move up and to the left, so higher-skill schools achieve higher true positive rates while generating fewer false positives.

Visualizing schools in ROC space provides a way to think about how differences in skill and preferences manifest in the data. Implementing this distinction empirically, however, requires recovering each school's true and false positive rates – which requires knowledge of the underlying prevalence of cases that would warrant investigation, $S_j \equiv \Pr(s_i = 1)$. An important identification challenge is that we do not directly observe $S_j$ because the underlying state for any individual child, $s_i$, is only revealed if the child is reported to CPS. For children who are not reported, we do not observe whether their cases would have been screened in had they been reported. As a result, the data do not directly reveal $S_j$, or thus the true and false positive rates, which means that we cannot place schools in ROC space without additional assumptions.

**What we can learn without $S_j$.** It is useful to begin by considering what can be learned from reporting rates and the conditional outcomes of those reports, without imposing additional assumptions. For each school $j$, we observe the reporting rate, $P_j$, and the share of reported cases that are screened in for investigation, $\mathcal{I}_j$, but not the underlying prevalence of cases that would warrant investigation, $S_j$.

Consider two schools $j$ and $j'$ where school $j$ reports more children, but a smaller share of its reports are screened in for investigation: $P_j > P_{j'}$ but $\mathcal{I}_j < \mathcal{I}_{j'}$. This pattern is consistent with multiple underlying explanations. In one scenario, the two schools have identical skill (i.e., the same ability to distinguish between children whose cases do and do not warrant investigation), but school $j$ places a higher weight on avoiding missed cases (false negatives) relative to unnecessary reports (false positives), leading it to adopt a lower reporting threshold, $\tau_j$. Alternatively, the same observed pattern could arise if school $j$ has a lower reporting threshold, but is *less* skilled at distinguishing between cases that do and do not warrant investigation. Consequently, reporting rates and the conditional outcomes of reported cases alone do not separately identify reporting preferences and skill.

Figure 5 shows the relationship between reporting rates $P_j$ and conditional investigation rates $\mathcal{I}_j$ in our data. We observe substantial variation in both reporting rates and conditional investigation rates, and even among schools with similar reporting rates there remains considerable variation in the share of reports screened in for investigation. However, the relationship is essentially flat and, if anything, slightly upward sloping. If schools differed only in their reporting preferences, one would instead expect a negative relationship, as more aggressive reporting would bring in lower-likelihood marginal cases and reduce the share of reports that are screened in for investigation. The absence of such a relationship suggests that differences in skill may also play an important role. In particular, the pattern is consistent with a setting in which higher-reporting schools are also more skilled at identifying cases that warrant CPS intervention. In that case, greater diagnostic skill would offset the decline in report quality that would otherwise arise from more aggressive reporting, helping to generate the observed flat relationship.

While this pattern is also consistent with alternative explanations – including a setting in which schools have little ability to distinguish between cases that do and do not warrant investigation and therefore report nearly at random – such explanations appear less consistent with the substantial variation in conditional investigation rates across schools. The reduced-form evidence thus suggests that differences in skill are important, although it does not by itself allow us to separately identify skill and preferences.

**What we can learn with $S_j$.** Now consider what we can learn if we are able to recover the population prevalence of cases that would warrant investigation, $S$. Note that, after risk adjustment, we treat $S$ as common across schools (i.e., it does not vary with $j$), reflecting the assumption that differences in underlying risk have been accounted for in the data. Still, in our setting, we do not directly observe $S$ because the underlying state is only revealed for reported cases. In principle, $S$ could be recovered using an "identification at infinity" approach that leverages schools with extreme reporting behavior (e.g., see Arnold, Dobbie and Hull (2022) and Baron et al. (2024)). In this approach, schools that report nearly all children would reveal the underlying prevalence of cases warranting investigation. However, in practice, most schools report only a small share of students in any given year, limiting the scope for such extrapolation in our data. Instead, we estimate $S$ as part of the structural model below.[18]

In our baseline specification, $\widehat{S} = 0.062$, indicating that approximately 6.2 percent of children in a given year would be screened in for investigation if reported. Given this estimate, we can compute each school's TPR and FPR, which we plot in Figure 6. The figure provides suggestive evidence that schools may differ in skill. There appear to be many instances in which one school has both a higher true positive rate *and* a lower false positive rate than another school, implying that the former school lies on a higher ROC curve and is more effective at distinguishing between cases that do and do not warrant investigation. Such dominance relationships cannot arise if all schools operate on a common ROC curve and therefore suggest there may be heterogeneity in skill. To further visualize this heterogeneity, in Figure A7 we fit a LOESS curve summarizing the relationship between TPR and FPR

[18]Intuitively, the structural model replaces identification at infinity with parametric extrapolation. As we describe below, by imposing additional structure on the relationship between latent risk, noisy signals, and reporting decisions, it allows the base rate $S$ to be estimated even though no schools report nearly all children.

across schools and bootstrap a 95% confidence interval. Many schools lie well above or below this curve, illustrating substantial dispersion in classification performance beyond what would be expected under a common ROC curve.

While this evidence suggests the presence of skill heterogeneity, it does not allow us to separately identify skill and preferences. Each school is observed at a single point in ROC space, which reflects both the informativeness of its signal about underlying risk and how it trades off false positives and false negatives. Without additional structure, multiple combinations of signal quality (skill) and preferences can generate the same observed point. Thus, the data alone do not allow us to uniquely decompose behavior into skill and preferences or to fully rank schools along either dimension. To make further progress, we follow Chan, Gentzkow and Yu (2022) and impose additional structure on how schools form assessments and make reporting decisions.

## 6.B A Structural Model of School Reporting Decisions

The model is as follows. Child $i$ has an underlying risk level $\nu_i \sim N(0,1)$. This risk determines a binary state $s_i$, indicating whether the child's case would be screened in for investigation if reported. Specifically, $s_i = \mathbf{1}\{\nu_i > \bar{\nu}\}$, where $\bar{\nu}$ governs the population prevalence of screen-in-eligible cases. School $j$ does not observe $\nu_i$ directly. Instead, it observes a noisy signal of the child's risk, $x_{ij}$, which it uses to determine whether to file a report. We assume that $(\nu_i, x_{ij})$ are jointly normally distributed and allow the correlation between a child's underlying risk and the signal received by a school to vary across schools:

$$\begin{pmatrix} \nu_i \\ x_{ij} \end{pmatrix} \sim N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \alpha_j \\ \alpha_j & 1 \end{pmatrix} \right).$$

The school-specific correlation between the signal and underlying risk, $\alpha_j \in (0,1)$, is our measure of school "skill." Preferences are summarized by a parameter $\beta_j$, which captures how schools trade off the costs of missed cases against unnecessary reports. We normalize the cost of a false positive to one, so that $\beta_j$ measures the cost of a false negative relative to the cost of a false positive. The school's utility is given by

$$u_{ij} = \begin{cases} -1 & \text{if } d_{ij} = 1,\ s_i = 0, \\ -\beta_j & \text{if } d_{ij} = 0,\ s_i = 1, \\ 0 & \text{otherwise.} \end{cases}$$

As shown by Chan, Gentzkow and Yu (2022), under this setup the school's optimal reporting decision takes the form of a threshold rule: the school files a report if and only if $x_{ij} > \tau_j$, the school-specific reporting threshold. The optimal threshold for a school equates the posterior probability that a reported case does not warrant investigation with the relative cost of a false negative versus a false positive. Specifically, the optimal cutoff $\tau_j^*$ satisfies $\Pr(s_i = 0 \mid x_{ij} = \tau_j^*) = \frac{\beta_j}{1+\beta_j}$, which implies

$$\tau_j^*(\alpha_j, \beta_j) = \frac{\bar{\nu} - \sqrt{1-\alpha_j^2}\,\Phi^{-1}\left(\frac{\beta_j}{1+\beta_j}\right)}{\alpha_j}.$$

The comparative statics of the threshold are intuitive. A higher $\bar{\nu}$, corresponding to a lower prevalence of screen-in-eligible cases ($S$), raises the reporting threshold. A higher $\beta_j$, reflecting a greater weight on avoiding missed cases, lowers the threshold and increases reporting. The effect of skill $\alpha_j$ is ambiguous: increases in skill make the signal more informative, which can either increase or decrease reporting depending on the prevalence of screen-in-eligible cases.

This structure is sufficient to map observed reporting behavior into skill and preferences. Under the joint normality assumption, schools with different values of $\alpha_j$ operate on different ROC curves, with higher values of $\alpha_j$ corresponding to higher curves. Given a school's reporting rate and conditional screen-in rate, each school can be located at a unique point in ROC space. Skill $\alpha_j$ is identified by the ROC curve on which that point lies, while preferences $\beta_j$ are identified by the school's location along the curve. Equivalently, preferences are recovered from the optimal threshold condition, which equates the slope of the ROC curve to the relative weight placed on false negatives versus false positives.[19] We estimate the model using risk-adjusted, school-level data on reporting behavior and downstream CPS responses to reports. Appendix C provides details of the estimation procedure.

**Model Estimates**

Table 3 reports estimates from the structural model. The top panel presents our baseline specification, which defines the underlying state as whether a report would be screened in for investigation. The bottom panel reports analogous estimates using substantiation following investigation as the underlying state. The qualitative conclusions are similar across both specifications.

In the baseline specification, we estimate the risk-adjusted prevalence of cases that would

[19] We can also allow for a share of children who are never at risk of CPS intervention. We abstract from this extension in our baseline specification and show in robustness exercises that incorporating it does not materially affect our estimates.

be screened in for investigation if reported, $\widehat{S}$, to be 6.2 percent. This exceeds the share of students who experience an actual screened-in report in a given year (4.3 percent). The difference between these quantities implies that a nontrivial share of investigation-worthy cases are missed by the reporting process, suggesting scope for improvements in schools' ability to identify and report high-risk children.

The average school's reporting skill, $\alpha_j$ – the correlation between the school's signal and underlying risk – is 0.65. There is substantial heterogeneity across schools: $\alpha_j$ ranges from 0.55 at the 10th percentile to 0.74 at the 90th percentile. These estimates indicate that schools differ substantially in their ability to distinguish between children whose cases do and do not warrant investigation. Schools also vary in their estimated reporting preferences. The average $\beta_j$ is 1.61, with a 90-10 difference of 0.56. Fewer than one percent of schools have $\beta_j < 1$, indicating that nearly all schools place greater weight on avoiding false negatives than on avoiding false positives.

The bottom panel reports analogous estimates using substantiation rather than screen-in as the underlying state. As expected, the estimated prevalence of substantiation-eligible cases is substantially lower, at 1.2 percent. The estimated level of skill is also lower, reflecting the fact that substantiation is a rarer and more difficult outcome to predict than whether a report will be screened in for investigation. At the same time, estimated preferences are substantially stronger: the average $\hat{\beta}_j$ rises from 1.6 to 18.4, implying that schools behave as if the cost of missing a substantiated case is much larger than the cost of making an unnecessary report. Despite these differences in levels, the rankings of schools are highly stable across specifications. As shown in the first row of Table A4, the correlation between skill estimates obtained using screen-ins and substantiations is 0.93, while the corresponding correlations for preferences and implied thresholds are 0.84 and 0.98, respectively.

For context, it is useful to benchmark these estimates against Chan, Gentzkow and Yu (2022). Radiologists in their study have an average skill of 0.86 and a much higher relative cost of false negatives than in our baseline estimates (mean $\hat{\beta}_j \approx 6.7$). Both differences are consistent with the underlying institutional environment. Diagnosing pneumonia from medical imaging involves specialized training and relatively high-quality signals, whereas school-based reporting relies on more diffuse and ambiguous information about children's circumstances, which likely limits diagnostic accuracy. At the same time, the consequences of false positives differ across settings. In the Chan, Gentzkow and Yu (2022) context, unnecessary treatment may be relatively low-cost, whereas in the child welfare context, a report can trigger a formal investigation – often lasting several weeks – and may lead to substantial intervention, including the possibility of child removal. Consistent with this, schools appear less strongly tilted toward avoiding missed cases that would warrant an investigation than radiologists.

We also find that skill and preferences are positively correlated ($\rho = 0.43$). This means that schools that are better able to distinguish between cases that do and do not warrant investigation place greater relative weight on avoiding missed cases. As shown in Figure A8, higher-skill schools tend to report at higher rates. At the same time, schools with stronger preferences for avoiding missed cases also report more frequently. Taken together, these results imply that higher reporting rates do not simply reflect a greater willingness to report marginal cases. Instead, schools that report more frequently also tend to be better at identifying children whose cases warrant investigation.

Next, Table A3 examines how estimated skill and preferences relate to observable school characteristics. Several patterns emerge. First, skill and preferences are both positively related to measures of school capacity. For example, schools employing a nurse exhibit substantially higher values of both $\alpha_j$ and $\beta_j$, consistent with greater ability to identify and respond to potential cases of maltreatment. Second, schools serving more disadvantaged populations – as proxied by Title I status – tend to exhibit both higher skill and stronger preferences for avoiding missed cases. Third, reporting behavior appears to capture a dimension of school practice that is distinct from traditional measures of school effectiveness. Schools with higher skill and stronger reporting preferences tend to have lower academic value-added, including test score and attendance measures. Lastly, staff composition is systematically related to both dimensions. Schools with higher shares of non-White teachers and principals tend to exhibit both higher skill and stronger preferences for reporting, while schools with more experienced teachers tend to exhibit lower skill and weaker preferences.

Finally, we assess the robustness of our estimates to alternative specifications in Table A4. In our baseline model, we construct risk-adjusted counts of reporting and screen-in outcomes by converting value-added measures for both the probability of being reported and the probability of being screened in conditional on reporting, and scale these by school enrollment to obtain adjusted counts of true positives and false positives. We then re-center these counts relative to the county distribution so that each school's outcomes are expressed relative to its local population baseline. Because we estimate a single statewide prevalence $\widehat{S}$, we show that results are similar when re-centering at the state level instead. We also consider two additional specifications. First, we redefine reporting outcomes using information from all reports a student receives within a school year, regardless of reporter type. Under this definition, a school report is classified as successful if the student has any report that is ultimately screened in (or substantiated), even if the report meeting that criterion was filed by a non-school reporter. This specification accounts for the possibility that multiple reporters identify the same child and tests whether our results are sensitive to attributing screening and substantiation outcomes exclusively to school reporters. Second, we allow for a fixed share of students to be not at risk of CPS involvement by introducing a parameter $\kappa$, capturing the possibility that some children lie outside the relevant risk set. Across these specifications, estimated skill, preferences, and implied thresholds remain highly correlated with the baseline estimates.

**Policy Counterfactuals**

We begin by assessing the relative importance of skill and preferences in generating variation in reporting behavior across schools. Specifically, we simulate counterfactual distributions of reporting decisions in which we eliminate variation in one dimension at a time. We fix school skill at the mean ($\alpha_j = \overline{\alpha}$) while allowing preferences to vary, and conversely fix preferences at the mean ($\beta_j = \overline{\beta}$) while allowing skill to vary. For each counterfactual distribution, we simulate reporting decisions and resulting reporting rates using the estimated model parameters. In our baseline specification, eliminating variation in skill reduces the standard deviation of risk-adjusted reporting rates by 38 percent, while eliminating variation in preferences reduces the standard deviation by 18 percent. When we measure reporting outcomes based on substantiated cases rather than screened-in calls, this difference is even larger: eliminating variation in skill reduces variation in reporting rates by 52 percent, while eliminating variation in preferences only reduces the variation by 11 percent. These results indicate that differences in skill account for a larger share of variation in risk-adjusted reporting behavior than differences in preferences.

Next, we use the estimated model to examine how alternative reporting policies affect reporting behavior and welfare. We focus on three policy-relevant interventions that target different sources of variation in reporting. First, we consider policies that improve schools' ability to identify high-risk cases by increasing reporting skill. Second, we consider policies that directly restrict reporting behavior by imposing caps on reporting rates. Third, we consider policies that reduce variation in reporting across schools by compressing reporting rates toward a common range. As discussed in Section 1, these counterfactuals capture central dimensions of current policy debates, which focus on whether to improve decision-making, limit reporting, or standardize reporting practices across institutions.

We evaluate the welfare implications of these policies using a welfare measure based on classification errors. Specifically, welfare depends on the rates of false positives and false negatives. Let $FP^0$ and $FN^0$ denote the corresponding rates under the status quo. We measure welfare relative to the status quo as $W = 1 - \frac{FP+\beta^S FN}{FP^0+\beta^S FN^0}$, where $\beta^S$ represents the social planner's relative weight on false negatives compared to false positives.

This normalization implies that $W = 0$ under the status quo and $W = 1$ when both false positives and false negatives are eliminated. Policies that reduce the weighted sum of classification errors increase welfare, while policies that increase classification errors reduce welfare. The parameter $\beta^S$ governs the relative importance of missed cases versus unnecessary reports: higher values of $\beta^S$ place greater weight on avoiding false negatives. In our baseline specification, we set $\beta^S = \overline{\beta}$, the average estimated preference parameter across schools, and later examine how policy effects vary across alternative values of $\beta^S$.

This welfare measure focuses on classification errors relative to the school's statutory role as a mandated reporter. School personnel are not responsible for determining whether maltreatment occurred or whether a child should be removed from the home; rather, their legal responsibility is to identify cases that warrant further assessment by CPS. Accordingly, our welfare framework evaluates schools based on their ability to identify cases that merit CPS intervention while avoiding unnecessary reports.[20] We view this as an important margin given the high-stakes nature of both types of mistakes: failing to report a child whose circumstances warrant intervention may leave a child in a harmful environment, while unnecessary reports can subject families to intrusive investigations and other costs associated with child welfare system involvement. In our baseline specification, the welfare weights are informed by schools' observed reporting decisions through the estimated preference parameters. These parameters provide information about how schools appear to perceive the relative costs and benefits of reporting decisions. They do not directly identify the underlying welfare consequences of child welfare system involvement, as they may also reflect legal incentives or other factors shaping reporting behavior.

We emphasize that our welfare criterion evaluates schools through their reporting decisions. A complete welfare analysis would also account for both the downstream effects of child welfare interventions and the broader effects of attending schools with different reporting practices. In our context in Michigan, existing evidence suggests that children on the margin of foster care placement experience improved outcomes following placement (Baron and Gross, 2024), indicating that failures to identify children whose cases warrant CPS intervention may carry substantial social costs.

[20] Our approach is conceptually similar to evaluations of other gatekeeping institutions, such as studies of bail judges that assess decision quality using observed pretrial misconduct (Arnold, Dobbie and Yang, 2018).

With this caveat in mind, Table 4 reports the effects of these counterfactual policies on reporting behavior and welfare. The top panel uses screen-ins to define the underlying state, while the bottom panel uses substantiation following investigation. Several clear patterns emerge. First, improving reporting skill generates significant welfare gains. This counterfactual can be interpreted as a stylized representation of policies that improve schools' ability to identify and assess high-risk cases, such as training, additional support staff, or better screening tools. Raising the skill of all schools below the 90th percentile to the 90th percentile level increases the true positive rate from 0.195 to 0.347 while only modestly increasing the false positive rate, yielding a welfare gain of 0.10 under the screen-in specification. This corresponds to closing 10 percent of the gap between the status quo and a benchmark with no classification errors. We obtain similar results when substantiation is used as the underlying state, with welfare increasing by 0.065. These gains arise because higher skill improves schools' ability to distinguish between higher- and lower-risk cases, increasing true positives much more than false positives.

In contrast, policies that directly restrict reporting behavior reduce welfare. Capping reporting rates at 1 percent – a stylized example of policies aimed at reducing reporting or discouraging reports – sharply lowers both true and false positives, but the reduction in true positives is substantially larger, resulting in a welfare loss of 0.026 under the screen-in specification and 0.013 under the substantiation specification. Similarly, compressing reporting rates across schools – a stylized representation of policies that seek to standardize reporting behavior across institutions – has smaller effects on reporting outcomes but nonetheless slightly reduces welfare in both specifications. These results stand in contrast to the gains generated by improving skill. Because higher-reporting schools also tend to be more skilled, policies that mechanically reduce or equalize reporting rates eliminate true positives as well as false positives, increasing the incidence of missed high-risk cases.

Figure 7 examines how the welfare effects of each policy vary with the social planner's preference parameter $\beta^S$. Rather than fixing a single welfare weight, the figure shows how the relative desirability of each policy depends on the weight placed on avoiding false negatives versus false positives. The central result is that the ranking of policies is remarkably stable across a wide range of welfare weights. Improving reporting skill generates positive welfare gains for nearly all values of $\beta^S$, while policies that restrict reporting generally reduce welfare. In contrast, policies that compress reporting rates toward a common range have relatively small welfare effects throughout. These conclusions are always true when false negatives receive at least 23 percent more weight than false positives (i.e., when $\beta^S > 1.23$), a threshold exceeded by 97 percent of schools in our sample.

The intuition is straightforward. Raising skill to the 90th percentile substantially increases true positives while generating only modest increases in false positives. As a result, welfare improves for nearly all values of $\beta^S$, except when the planner places very little weight on missed cases (approximately $\beta^S < 0.5$).[21] By contrast, capping reporting rates at 1 percent reduces both true and false positives, but the reduction in true positives is substantially larger. Consequently, welfare falls for most values of $\beta^S$, particularly when the planner places moderate or high weight on avoiding false negatives. Finally, compressing reporting rates within a narrow band around the mean generates only modest changes in average reporting behavior and therefore produces welfare effects that remain close to zero across the full range of welfare weights.

# 7 Conclusion

This paper studies the sources and consequences of variation in child maltreatment reporting, focusing on the role of institutional practices in public schools. Using linked administrative data from Michigan and a value-added framework validated with student moves, we show that reporting behavior varies substantially across schools even after accounting for student risk. Approximately 40 percent of the variation in reporting is attributable to differences in school reporting tendencies rather than differences in underlying risk.

To interpret this variation, we estimate a structural model that separates differences in diagnostic skill from differences in reporting preferences. We find that both dimensions vary across schools, but differences in skill explain a much larger share of the variation in reporting behavior. Moreover, schools that report more frequently tend to be both more skilled at identifying cases that would warrant CPS intervention and more willing to report.

These findings have direct implications for policy. Reporting decisions involve a tradeoff between missed cases and unnecessary intervention. Consistent with the model's decomposition, we show that ongoing policies that seek to restrict or equalize reporting behavior – such as tighter reporting standards or reporting caps – generally reduce welfare by increasing missed cases. In contrast, policies that improve the quality of information and decision-making – such as training, additional support staff, or access to better signals – generally increase welfare by improving detection with only modest increases in false positives.

[21] Increasing skill leads to schools reporting more students, meaning both true positive and false positive rates increase. This can reduce welfare when society is very averse to falsely reporting children to CPS.

More broadly, our findings illustrate the importance of moving beyond observed reporting rates when interpreting variation in children's exposure to the child welfare system. A first step is to distinguish differences in underlying risk from differences in institutional behavior. However, our results show that this distinction alone is not sufficient. Even when variation reflects institutional behavior, it remains important to distinguish differences in reporting preferences from differences in schools' ability to identify high-risk cases. These sources of variation have fundamentally different implications for how disparities in child welfare system involvement should be interpreted and for how policy should respond to them.

# References


**Abaluck, Jason, Leila Agha, Chris Kabrhel, Ali Raja, and Arjun Venkatesh.** 2016. "The determinants of productivity in medical testing: Intensity and allocation of care." *American Economic Review*, 106(12): 3730–3764.

**Abdulkadiroğlu, Atila, Parag A Pathak, Jonathan Schellenberg, and Christopher R Walters.** 2020. "Do parents value school effectiveness?" *American Economic Review*, 110(5): 1502–39.

**Angrist, Joshua, Peter Hull, and Christopher Walters.** 2023. "Methods for measuring school effectiveness." *Handbook of the Economics of Education*, 7: 1–60.

**Angrist, Joshua, Peter Hull, Parag Pathak, and Christopher Walters.** 2016. "Interpreting tests of school VAM validity." *American Economic Review: Papers and Proceedings*, 106(5): 388–392.

**Angrist, Joshua, Peter Hull, Russell Legate-Yang, Parag A Pathak, and Christopher R Walters.** 2026. "Putting school surveys to the test." *American Economic Review*.

**Arnold, David, Will Dobbie, and Crystal S Yang.** 2018. "Racial bias in bail decisions." *The Quarterly Journal of Economics*, 133(4): 1885–1932.

**Arnold, David, Will Dobbie, and Peter Hull.** 2022. "Measuring racial discrimination in bail decisions." *American Economic Review*, 112(9): 2992–3038.

**Badinski, Ivan, Amy Finkelstein, Matthew Gentzkow, and Peter Hull.** 2023. "Geographic variation in healthcare utilization: The role of physicians." National Bureau of Economic Research.

**Bald, Anthony, Joseph J Doyle Jr, Max Gross, and Brian A Jacob.** 2022. "Economics of foster care." *Journal of Economic Perspectives*, 36(2): 223–246.

**Baron, E Jason, and Max Gross.** 2024. "Is There a Foster Care-To-Prison Pipeline? Evidence from Quasi-Randomly Assigned Investigators." *Review of Economics and Statistics*.

**Baron, E Jason, Ezra G Goldstein, and Cullen T Wallace.** 2020. "Suffering in silence: How COVID-19 school closures inhibit the reporting of child maltreatment." *Journal of public economics*, 190: 104258.

**Baron, E. Jason, Ezra G. Goldstein, and Joseph Ryan.** 2023. "The push for racial equity in child welfare: Can blind removals reduce disproportionality?" *Journal of Policy Analysis and Management*, 42(2): 456–487.

**Baron, E Jason, Joseph J Doyle Jr, Natalia Emanuel, Peter Hull, and Joseph Ryan.** 2024. "Discrimination in multiphase systems: Evidence from child protection." *The Quarterly Journal of Economics*, 139(3): 1611–1664.

**Bayer, Patrick, Fernando Ferreira, and Robert McMillan.** 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy*, 115(4): 588–638.

**Benson, Cassandra, Maria D Fitzpatrick, and Samuel Bondurant.** 2025. "Beyond reading, writing, and arithmetic: The role of teachers and schools in reporting child maltreatment." *Journal of Human Resources*, 60(1): 153–186.

**Beuermann, Diether W, C Kirabo Jackson, Laia Navarro-Sola, and Francisco Pardo.** 2023. "What is a good school, and can parents tell? Evidence on the multidimensionality of school output." *The Review of Economic Studies*, 90(1): 65–101.

**Biasi, Barbara.** 2021. "The labor market for teachers under different pay schemes." *American Economic Journal: Economic Policy*, 13(3): 63–102.

**Campos, Christopher, and Caitlin Kearns.** 2024. "The impact of public school choice: Evidence from Los Angeles's zones of choice." *The Quarterly Journal of Economics*, 139(2): 1051–1093.

**Casey Family Programs.** 2025. "How Are Some Jurisdictions Transforming Mandated Reporting?" Casey Family Programs.

**Chan, David C, Matthew Gentzkow, and Chuan Yu.** 2022. "Selection with variation in diagnostic skill: Evidence from radiologists." *The Quarterly Journal of Economics*, 137(2): 729–783.

**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.

**Child Welfare Information Gateway.** 2023. "Mandatory Reporters of Child Abuse and Neglect." *https://www.childwelfare.gov/topics/safety-and-risk/mandated-reporting/*, U.S. Department of Health and Human Services.

**Cullen, Julie Berry, Brian A. Jacob, and Steven D. Levitt.** 2006. "The Effect of School Choice on Participants: Evidence from Randomized Lotteries." *Econometrica*, 74(5): 1191–1230.

**Currie, Janet, and W Bentley MacLeod.** 2017. "Diagnosing expertise: Human capital, decision making, and performance among physicians." *Journal of labor economics*, 35(1): 1–43.

**Currie, Janet, W. Bentley MacLeod, and Kate Musen.** 2026. "Doctor Decision-Making and Patient Outcomes." *Journal of Economic Literature*, 64(1): 141–94.

**Deming, David J.** 2014. "Using school choice lotteries to test measures of school effectiveness." *American Economic Review*, 104(5): 406–411.

**Dettlaff, Alan J, and Reiko Boyd.** 2020. "Racial disproportionality and disparities in the child welfare system: Why do they exist, and what can be done to address them?" *The ANNALS of the American Academy of Political and Social Science*, 692(1): 253–274.

**Doyle, Joseph J., and Becky Staiger.** 2025. "Physician Group Influences on Treatment Intensity and Health: Evidence from Physician Switchers." *American Economic Journal: Economic Policy*, 17(2): 471–505.

**Doyle, Joseph J., Natalia Emanuel, and Raimundo Eyzaguirre.** 2025. "Cumulative Risks of Foster Care Placement for U.S. Children: A Research Note Comparing Birth Cohort and Synthetic Cohort Approaches." *Working paper.*

**Drake, Brett, Dylan Jones, Jun-Hong Chen, Sarah Font, Emily Putnam-Hornstein, Richard P Barth, and Melissa Jonson-Reid.** 2023. "Poverty or Racism? A Re-Analysis of Briggs et al. 2022." *Research on Social Work Practice*, 10497315231179648.

**Edwards, Frank, Sara Wakefield, Kieran Healy, and Christopher Wildeman.** 2021. "Contact with Child Protective Services is pervasive but unequally distributed by race and ethnicity in large US counties." *Proceedings of the National Academy of Sciences*, 118(30): e2106272118.

**Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams.** 2016. "Sources of geographic variation in health care: Evidence from patient migration." *The quarterly journal of economics*, 131(4): 1681–1726.

**Fong, Kelley.** 2019. "Neighborhood inequality in the prevalence of reported and substantiated child maltreatment." *Child Abuse & Neglect*, 90: 13–21.

**Fong, Kelley.** 2020. "Getting eyes in the home: Child protective services investigations and state surveillance of family life." *American Sociological Review*, 85(4): 610–638.

**Grimon, Marie-Pascale, and Christopher Mills.** 2025. "Better together? a field experiment on human-algorithm interaction in child protection." *arXiv preprint arXiv:2502.08501.*

**Jackson, C Kirabo.** 2018. "What do test scores miss? The importance of teacher effects on non–test score outcomes." *Journal of Political Economy*, 126(5): 2072–2107.

**Jackson, C Kirabo, Shanette C Porter, John Q Easton, Alyssa Blanchard, and Sebastián Kiguel.** 2020. "School effects on socioemotional development, school-based arrests, and educational attainment." *American Economic Review: Insights*, 2(4): 491–508.

**Jacob, Brian A, Lars Lefgren, and David P Sims.** 2010. "The persistence of teacher-induced learning." *Journal of Human resources*, 45(4): 915–943.

**Jacob, Brian, and Jesse Rothstein.** 2016. "The measurement of student ability in modern assessment systems." *Journal of Economic Perspectives*, 30(3): 85–108.

**Jones, Kristin.** 2024. "States Find a Downside to Mandatory Reporting Laws Meant to Protect Children." NPR.

**Kane, Thomas J, and Douglas O Staiger.** 2008. "Estimating teacher impacts on student achievement: An experimental evaluation." National Bureau of Economic Research.

**Kim, Hyunil, Christopher Wildeman, Melissa Jonson-Reid, and Brett Drake.** 2017. "Lifetime Prevalence of Investigating Child Maltreatment Among US Children." *American Journal of Public Health*, 107(2): 274–280. PMID: 27997240.

**New York State Legislature.** 2024. "An Act to Amend the Social Services Law, in Relation to Reports of Child Abuse and Maltreatment." *https://www.nysenate.gov/legislation/bills*, Ending anonymous reporting to the Statewide Central Register.

**Petek, Nathan, and Nolan G Pope.** 2023. "The multidimensional impact of teachers on students." *Journal of Political Economy*, 131(4): 1057–1107.

**Raz, Mical.** 2020. "Calling child protectives services is a form of community policing that should be used appropriately: Time to engage mandatory reporters as to the harmful effects of unnecessary reports." *Children and Youth Services Review*, 110: 104817.

**Rittenhouse, Katherine, Emily Putnam-Hornstein, and Rhema Vaithianathan.** 2022. "Algorithms, Humans, and Racial Disparities in Child Protective Services: Evidence from the Allegheny Family Screening Tool." Working paper.

**Rose, Evan K, Jonathan T Schellenberg, and Yotam Shem-Tov.** 2025. "The effects of teacher quality on adult criminal justice contact." *NBER Working Paper #w30274*.

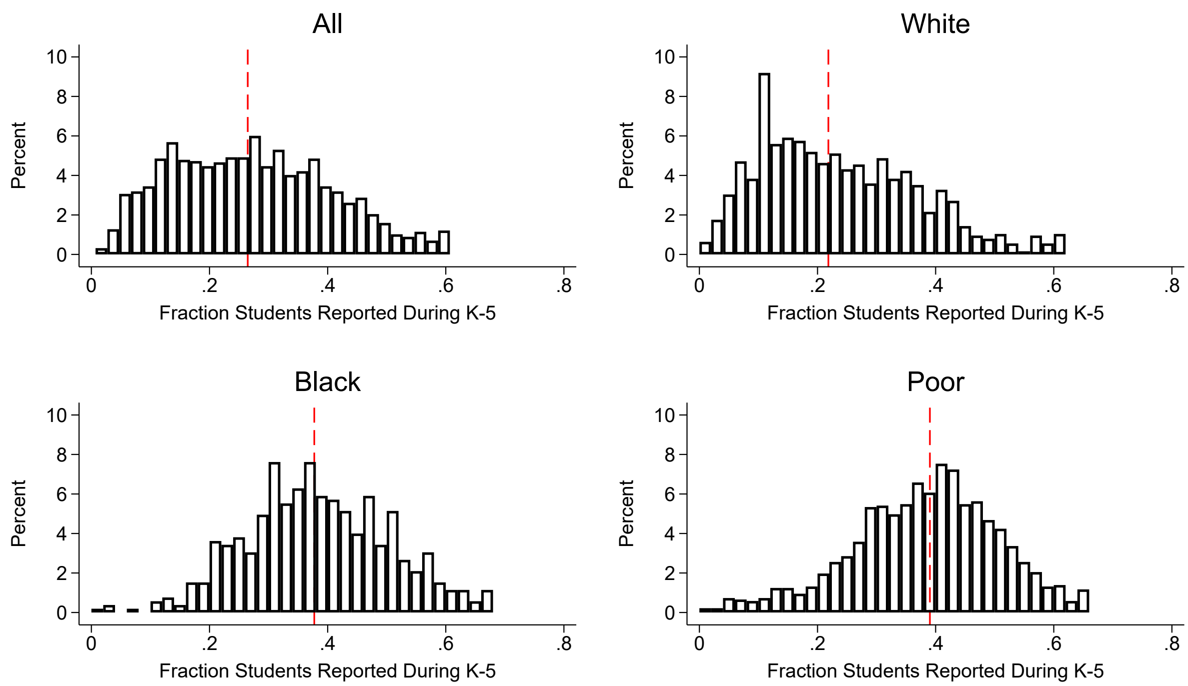
**Safe & Sound.** 2025. "A Landmark Moment: California Codifies Move from Mandated Reporting to Supporting Families." Accessed 2026.

**Sarkar, Susanti.** 2024. "Child Welfare Officials in New York Get an Earful on Hazards of CPS Hotline Practices." The Imprint.

**Yi, Youngmin, Frank Edwards, Natalia Emanuel, Hedwig Lee, John M Leventhal, Jane Waldfogel, and Christopher Wildeman.** 2023. "State-level variation in the cumulative prevalence of child welfare system contact, 2015–2019." *Children and youth services review*, 147: 106832.

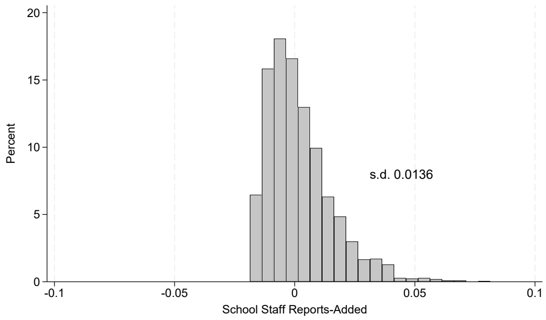
**Yi, Youngmin, Frank R Edwards, and Christopher Wildeman.** 2020. "Cumulative prevalence of confirmed maltreatment and foster care placement for US children by race/ethnicity, 2011–2016." *American journal of public health*, 110(5): 704–709.

Figure 1: Proportion of Students Ever Reported in K-5

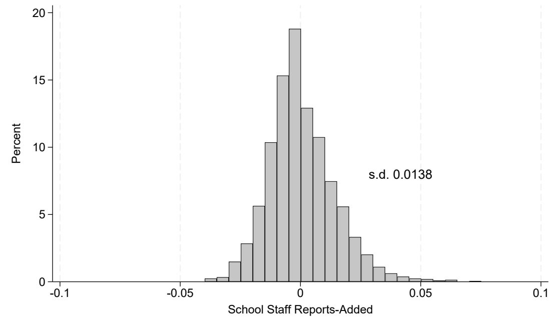



*Note:* The figure displays the distribution across schools of the percent of students reported to CPS at least once between kindergarten and fifth grade. The sample includes children who entered a Michigan public or charter school in Fall 2017 or Fall 2018. The four panels report the distribution for all students, White students, Black students, and economically disadvantaged students. Schools with fewer than 25 students meeting the sample criteria are excluded, as are reports made between March 2020 and June 2021. The red dashed lines indicate the median school.

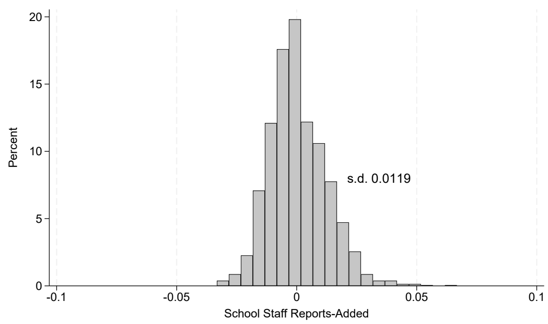
Figure 2: Distribution of School Reports-Added Estimates




(a) No Controls



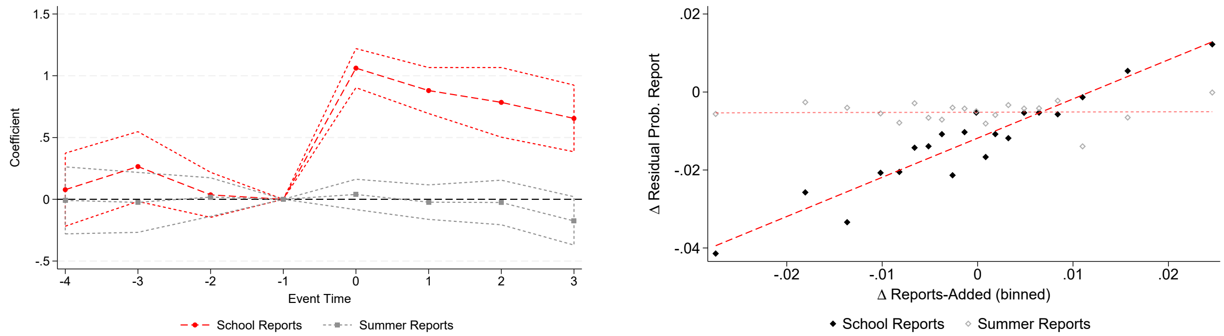

(b) Add Student and School Controls




(c) Add Controls for Lagged Reports

*Note:* The figure shows the distribution of school-level "reports-added" estimates across specifications. Reports-added is estimated using a value-added framework and measures a school's contribution to the probability that a student is reported to CPS by school staff. The dependent variable is an indicator for whether a student receives a school-based CPS report in a given year. Panel (a) includes no controls. Panel (b) adds student, school, and neighborhood characteristics. Panel (c) further includes controls for students' prior CPS reports. All specifications exclude student-school observations following a student's first report at that school. Estimates are empirical Bayes–shrunken to account for sampling variation.

Figure 3: Evidence on Changes in Reporting Around School Switches

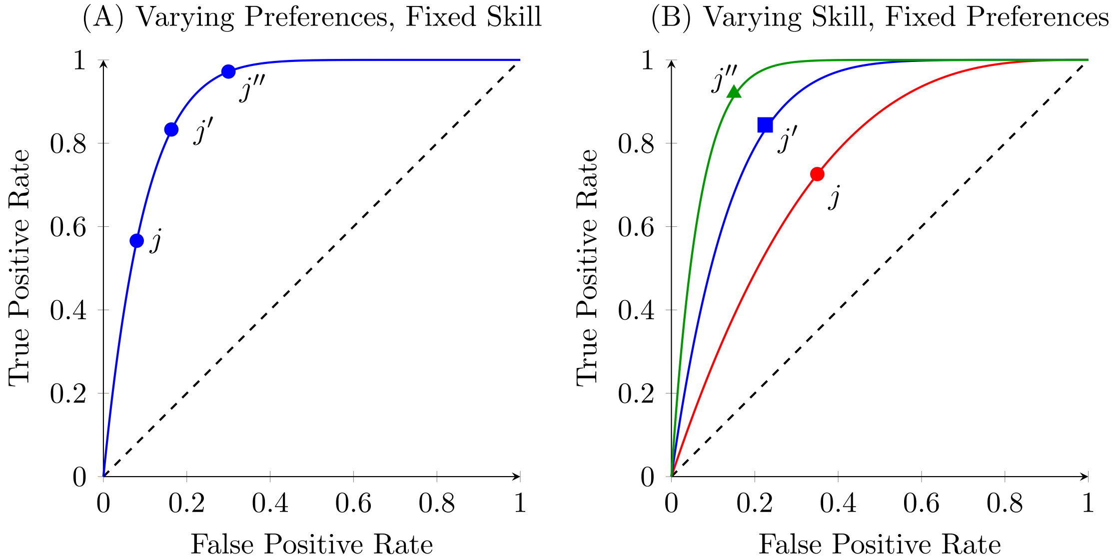



(a) Event Study

(b) Binscatter

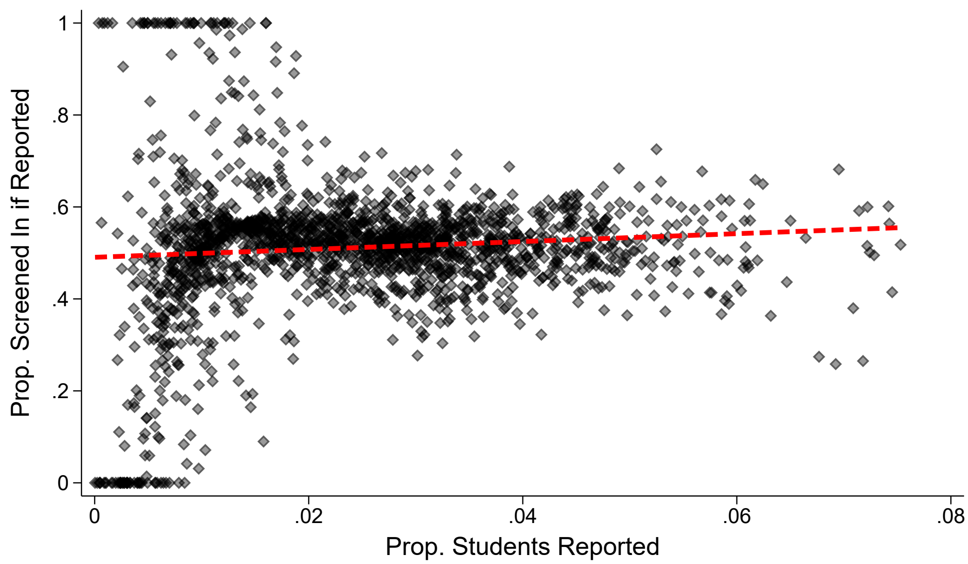
*Note:* The figure reports event-study and binscatter estimates of changes in student-level CPS reporting following school switches. Event-study coefficients are estimated from regressions of an indicator for receiving a school-based CPS report on event-time indicators interacted with the difference in reports-added between the destination and origin schools. Panel (a) compares changes in risk-adjusted school-based reports with placebo estimates based on reports from non-school reporters occurring during the preceding summer. Event-study specifications include grade and year fixed effects, the main effect of the change in reports-added, and standard errors clustered at the student and destination-school levels. Panel (b) plots the year-over-year change in risk-adjusted school-based reporting against the change in reports-added induced by the school switch. The estimated binscatter slope is 1.00 (standard error = 0.06). The sample includes students who switch schools once during elementary school (in second, third, or fourth grade). In all panels, reports-added is estimated using only students who never switch schools.

Figure 4: Illustrative Variation in Preferences and Skill in ROC Space



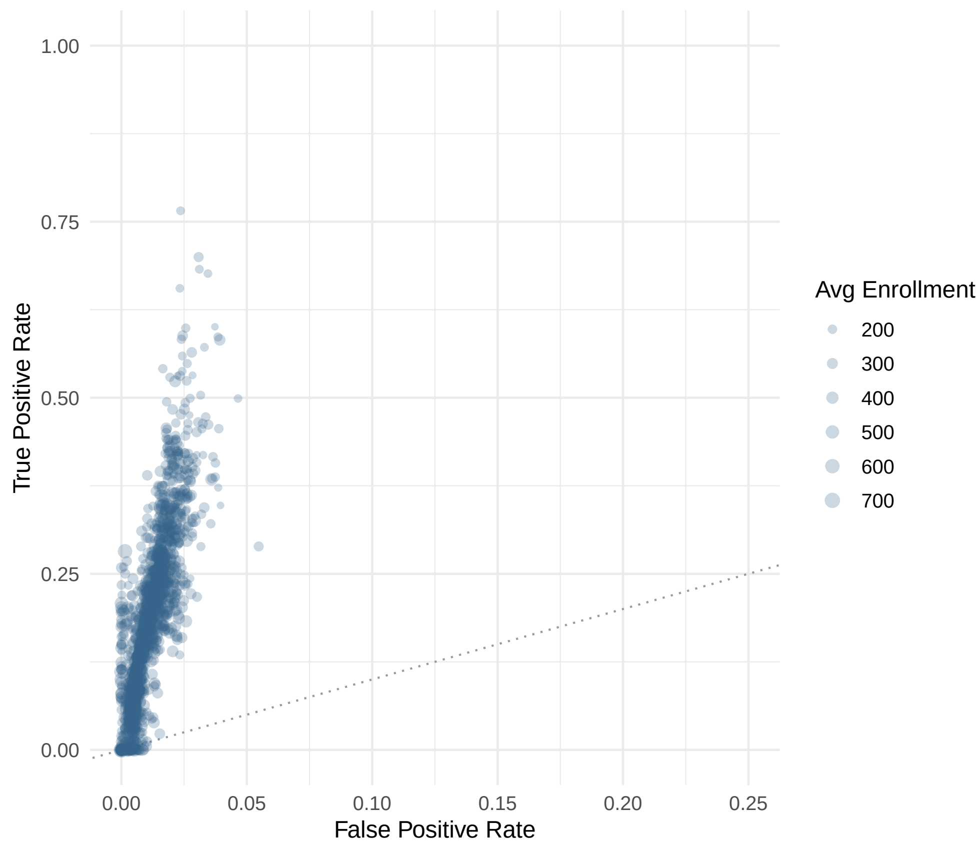

*Note:* The figure is adapted from Chan, Gentzkow and Yu (2022). It illustrates how differences in preferences and skill appear in ROC space. Each ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) as reporting behavior varies. The ROC curve represents the frontier of achievable classification outcomes for a school with a given level of skill. A school with no informative signal lies on the 45-degree line, while more skilled schools have ROC curves that lie weakly above those of less skilled schools. In Panel A, schools $j$, $j'$, and $j''$ operate on the same ROC curve, indicating equal skill, but differ in their reporting behavior: moving from $j$ to $j''$ corresponds to a greater willingness to report, increasing both true and false positive rates. In Panel B, schools differ in skill but share the same preferences. Each school chooses a point corresponding to a tangency between its ROC curve and a parallel set of indifference curves. In this framework, skill determines which ROC frontier a school operates on, while preferences determine where along that frontier it chooses to operate.

Figure 5: School Reporting Rates and Conditional Investigation Rates



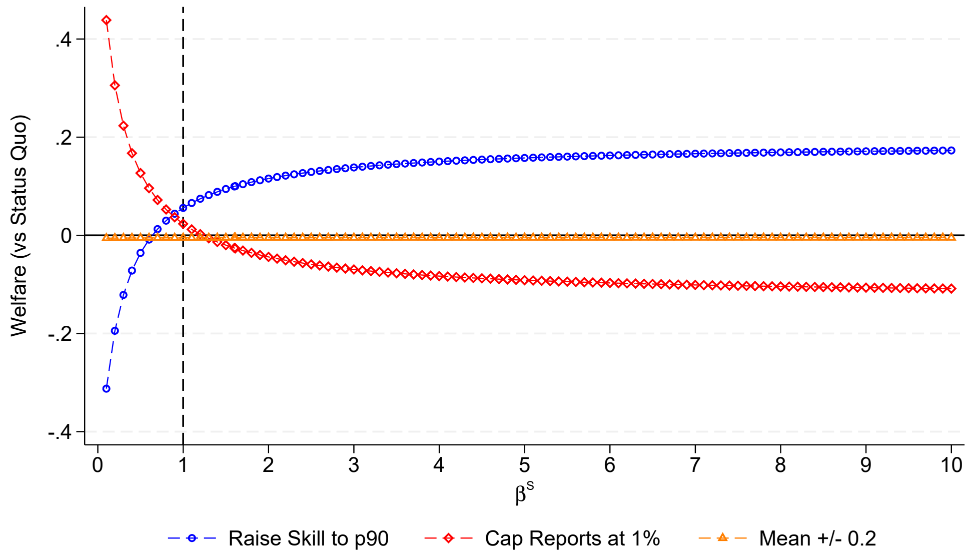

*Note:* The figure plots school-level reporting rates against the share of reported cases that are screened in for investigation. Both reporting and screen-in rates are risk-adjusted. Each point represents a school. The relationship is approximately flat, indicating that higher-reporting schools do not systematically generate reports that are more or less likely to be screened in.

Figure 6: School True and False Positive Rates in ROC Space




*Note:* The figure plots each school's true positive rate (TPR) against its false positive rate (FPR), assuming a population prevalence of cases warranting investigation of $\widehat{S} = 0.062$, as estimated in the structural model. Each point represents a school. Schools that achieve both higher true positive rates and lower false positive rates than other schools lie on higher ROC curves and are therefore more skilled at identifying cases that warrant investigation. Estimates are risk-adjusted. To reduce the influence of measurement error, the sample is restricted to schools with at least 100 students per year.

Figure 7: Welfare Effects of Counterfactual Policies Across Social Preferences




*Note:* This figure shows the change in welfare relative to the status quo for three counterfactual policies across different values of the social planner's preference parameter $\beta^S$. The vertical axis reports welfare gains relative to the status quo, and the horizontal axis shows $\beta^S$, which captures the relative weight placed on false negatives versus false positives. The vertical dashed line denotes where $\beta^S = 1$. Outcomes are simulated using the estimated structural model. Policy (1) increases reporting skill for all schools below the 90th percentile of the skill distribution. Policy (2) caps school reporting rates at 1 percent. Policy (3) reduces variation in reporting by restricting school reporting rates to lie within $\pm 0.2$ percentage points of the mean.

Table 1: Characteristics of Michigan K–5 Students by CPS Reporting Status and Race

| | All | By CPS Reporting Status | | By Race/Ethnicity | | |
|---|---|---|---|---|---|---|
| | | Never Reported | Ever Reported | White | Black | Hispanic |
| N | 1,232,220 | 970,304 | 261,916 | 778,211 | 279,302 | 106,302 |
| Prop. Students | | 0.787 | 0.213 | 0.632 | 0.227 | 0.086 |
| Ever Reported Before KG | 0.110 | | 0.516 | 0.096 | 0.165 | 0.100 |
| | (0.313) | | (0.500) | (0.295) | (0.371) | (0.300) |
| Ever Reported, KG-5th Grade | 0.197 | | 0.927 | 0.174 | 0.285 | 0.195 |
| | (0.398) | | (0.260) | (0.379) | (0.451) | (0.396) |
| Ever Reported Through 5th Grade | 0.213 | | 1.000 | 0.188 | 0.306 | 0.209 |
| | (0.409) | | (0.000) | (0.391) | (0.461) | (0.407) |
| Any Substantiated Report Before KG | 0.047 | | 0.223 | 0.041 | 0.073 | 0.042 |
| | (0.213) | | (0.416) | (0.199) | (0.261) | (0.201) |
| Any Substantiated Report, KG-5th Grade | 0.054 | | 0.255 | 0.047 | 0.082 | 0.052 |
| | (0.226) | | (0.436) | (0.211) | (0.274) | (0.223) |
| Any Substantiated Report Through 5th Grade | 0.086 | | 0.406 | 0.075 | 0.132 | 0.081 |
| | (0.281) | | (0.491) | (0.263) | (0.339) | (0.273) |
| *Demographics* | | | | | | |
| White | 0.632 | 0.651 | 0.559 | | | |
| | (0.482) | (0.477) | (0.496) | | | |
| Black | 0.227 | 0.200 | 0.326 | | | |
| | (0.419) | (0.400) | (0.469) | | | |
| Hispanic | 0.086 | 0.087 | 0.085 | | | |
| | (0.281) | (0.281) | (0.279) | | | |
| Other Race | 0.056 | 0.062 | 0.030 | | | |
| | (0.229) | (0.242) | (0.170) | | | |
| Female | 0.487 | 0.489 | 0.478 | 0.484 | 0.491 | 0.490 |
| | (0.500) | (0.500) | (0.500) | (0.500) | (0.500) | (0.500) |
| Age in KG | 5.179 | 5.178 | 5.186 | 5.188 | 5.158 | 5.180 |
| | (0.395) | (0.392) | (0.406) | (0.401) | (0.378) | (0.396) |
| Economically Disadvantaged | 0.578 | 0.502 | 0.859 | 0.469 | 0.851 | 0.766 |
| | (0.494) | (0.500) | (0.348) | (0.499) | (0.356) | (0.424) |
| IEP | 0.142 | 0.122 | 0.216 | 0.145 | 0.139 | 0.146 |
| | (0.349) | (0.327) | (0.411) | (0.352) | (0.346) | (0.353) |
| ELL | 0.086 | 0.099 | 0.036 | 0.048 | 0.015 | 0.376 |
| | (0.280) | (0.299) | (0.186) | (0.213) | (0.120) | (0.484) |
| Chronically Absent (Attendance <90%) | 0.244 | 0.215 | 0.353 | 0.173 | 0.431 | 0.286 |
| | (0.429) | (0.411) | (0.478) | (0.378) | (0.495) | (0.452) |
| Attendance Rate | 0.920 | 0.926 | 0.897 | 0.935 | 0.879 | 0.915 |
| | (0.095) | (0.089) | (0.111) | (0.077) | (0.124) | (0.093) |
| Great Start Readiness Program | 0.308 | 0.288 | 0.383 | 0.262 | 0.433 | 0.391 |
| | (0.462) | (0.453) | (0.486) | (0.440) | (0.495) | (0.488) |
| Head Start | 0.086 | 0.069 | 0.151 | 0.062 | 0.157 | 0.101 |
| | (0.281) | (0.253) | (0.358) | (0.241) | (0.364) | (0.302) |

*Note:* The sample includes students observed in kindergarten through fifth grade in Michigan public or charter schools between the 2017–18 and 2023–24 school years. We report characteristics for each student in their first observed year. "Ever reported" indicates whether the student was ever the subject of a CPS report through fifth grade. Measures of prior CPS involvement include reports occurring before kindergarten entry. Economically disadvantaged indicates eligibility for free or reduced-price lunch, receipt of SNAP or TANF, homelessness, foster care placement, or migrant status. IEP denotes receipt of special education services, and ELL denotes English Language Learner status. Standard deviations are reported in parentheses.

Table 2: School Characteristics and Reporting Behavior (Bivariate Associations)

| | Mean (sd) | Association with Reporting |
|---|---|---|
| *Staff Demographics* | | |
| Prop. Teachers Female | 0.837 | 0.027** |
| | (0.128) | (0.012) |
| Prop. Teachers Black | 0.071 | 0.070*** |
| | (0.175) | (0.010) |
| Prop. Teachers Hispanic | 0.016 | 0.046*** |
| | (0.040) | (0.008) |
| Prop. Teachers Other Race | 0.024 | 0.045*** |
| | (0.050) | (0.010) |
| Prop. Principals Black | 0.123 | 0.140*** |
| | (0.322) | (0.032) |
| Prop. Principals Hispanic | 0.013 | 0.371*** |
| | (0.108) | (0.067) |
| Prop. Principals Other Race | 0.016 | 0.149* |
| | (0.122) | (0.077) |
| *Staff Experience and Credentials* | | |
| Average Teacher Age | 43.578 | 0.000 |
| | (4.919) | (0.011) |
| Teacher Years at School | 10.762 | -0.036*** |
| | (4.469) | (0.012) |
| Prop. Teacher Master's Degree | 0.513 | -0.022* |
| | (0.216) | (0.012) |
| *School Characteristics* | | |
| Student-Teacher Ratio | 17.633 | -0.014 |
| | (35.308) | (0.023) |
| Charter | 0.131 | 0.044 |
| | (0.338) | (0.036) |
| Title I | 0.385 | 0.257*** |
| | (0.487) | (0.023) |
| *School Performance* | | |
| Math Test Score VA | 0.001 | -0.093*** |
| | (0.934) | (0.014) |
| ELA Test Score VA | 0.001 | -0.133*** |
| | (0.934) | (0.015) |
| *Support Staff* | | |
| Has Guidance Counselor | 0.141 | -0.063** |
| | (0.348) | (0.027) |
| Has Social Worker | 0.540 | -0.065*** |
| | (0.498) | (0.021) |
| Has School Psychologist | 0.317 | 0.003 |
| | (0.465) | (0.022) |
| Has Nurse | 0.062 | 0.233*** |
| | (0.242) | (0.042) |
| Observations | 11,530 | 11,530 |

*Note:* Each coefficient reports the bivariate association between a school characteristic and the average residual probability of a school-based CPS report at school $j$ in year $t$. Regressions are estimated separately for each characteristic and include county-by-year fixed effects. Continuous variables are standardized. The unit of observation is a school-year. Standard errors, clustered at the school level, are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Structural Model Estimates

| | Mean | p10 | p25 | p50 | p75 | p90 |
|---|---|---|---|---|---|---|
| | | | **Screen-Ins** | | | |
| $\widehat{S}$ | 0.062 | | | | | |
| | (0.007) | | | | | |
| $\widehat{\alpha}_j$ | 0.653 | 0.554 | 0.607 | 0.664 | 0.703 | 0.737 |
| | (0.023) | (0.019) | (0.022) | (0.024) | (0.026) | (0.026) |
| $\widehat{\beta}_j$ | 1.611 | 1.365 | 1.450 | 1.579 | 1.732 | 1.925 |
| | (0.062) | (0.028) | (0.033) | (0.054) | (0.088) | (0.130) |
| | | | **Substantiations** | | | |
| $\widehat{S}$ | 0.012 | | | | | |
| | (0.001) | | | | | |
| $\widehat{\alpha}_j$ | 0.371 | 0.295 | 0.325 | 0.376 | 0.412 | 0.450 |
| | (0.020) | (0.013) | (0.016) | (0.020) | (0.024) | (0.028) |
| $\widehat{\beta}_j$ | 18.39 | 17.48 | 17.75 | 18.18 | 18.82 | 19.65 |
| | (0.64) | (0.38) | (0.47) | (0.62) | (0.81) | (1.02) |

*Note:* This table reports estimates from the structural model of school reporting decisions. $\widehat{S}$ is the estimated risk-adjusted share of children whose cases would warrant investigation if reported. $\widehat{\alpha}_j$ denotes school reporting skill, measured as the correlation between the school's signal and the child's latent risk. $\widehat{\beta}_j$ denotes the school's relative cost of false negatives versus false positives; values above one imply greater weight on missed cases than on unnecessary reports. The table also reports the distribution of empirical Bayes posterior means across schools. Bootstrapped standard errors are shown in parentheses.

Table 4: Welfare Effects of Counterfactual Reporting Policies

| | Reporting Rate | TPR | FPR | Prop. Schools Affected | $\Delta$ Welfare |
|---|---|---|---|---|---|
| **Screen-Ins** | | | | | |
| 0. Status quo | 0.023 | 0.195 | 0.012 | – | – |
| | (0.014) | (0.116) | (0.007) | | |
| 1. Raise skill to p90 ($\widehat{\alpha} = 0.76$) | 0.039 | 0.347 | 0.018 | 0.900 | 0.100 |
| | (0.010) | (0.062) | (0.006) | | |
| 2. Cap reporting at 1% | 0.009 | 0.093 | 0.004 | 0.821 | -0.026 |
| | (0.002) | (0.028) | (0.001) | | |
| 3. Compress reporting to mean ± 0.2 pp | 0.023 | 0.191 | 0.012 | 0.887 | -0.005 |
| | (0.001) | (0.047) | (0.002) | | |
| **Substantiations** | | | | | |
| 0. Status quo | 0.023 | 0.144 | 0.022 | | |
| | (0.014) | (0.096) | (0.013) | | |
| 1. Raise skill to p90 ($\widehat{\alpha}$=0.47) | 0.041 | 0.279 | 0.038 | 0.900 | 0.065 |
| | (0.008) | (0.042) | (0.008) | | |
| 2. Cap reporting at 1% | 0.009 | 0.072 | 0.008 | 0.804 | -0.013 |
| | (0.002) | (0.029) | (0.002) | | |
| 3. Compress reporting to mean ± 0.2 pp | 0.023 | 0.140 | 0.022 | 0.898 | -0.004 |
| | (0.002) | (0.045) | (0.002) | | |

*Note:* This table reports simulated reporting outcomes and welfare under the status quo and alternative policy counterfactuals based on the estimated structural model. The top panel reports results from our baseline specification, in which the underlying state is whether a report would be screened in for investigation. The bottom panel reports analogous counterfactuals when the underlying state is defined using substantiation following investigation. Outcomes are generated by simulating reporting decisions for a large population of schools (10,000) drawn from the estimated distribution of school-level skill and preference parameters. The reporting rate denotes the share of students reported, TPR denotes the true positive rate, and FPR denotes the false positive rate. Welfare is measured using a weighted loss function that places weight $\beta^S$ on false negatives relative to false positives and is reported relative to the status quo: $W = 1 - \frac{FP+\beta^S FN}{FP^0+\beta^S FN^0}$. For this table, welfare is evaluated at $\beta^S = \overline{\beta}$, the average estimated preference parameter across schools. The first row of each panel reports outcomes under the status quo. Policy (1) raises the skill of all schools below the 90th percentile of the skill distribution to the 90th percentile level. Policy (2) caps school reporting rates at 1 percent. Policy (3) compresses reporting behavior by restricting school reporting rates to lie within ±0.2 percentage points of the mean. "Prop. Schools Affected" reports the share of schools whose reporting decisions change under each policy. Standard deviations across schools are reported in parentheses.

# Online Appendix

## Table of Contents

# A Additional Tables

Table A1: School and Neighborhood Characteristics by CPS Reporting Status and Race

| | All | By CPS Reporting Status | | By Race/Ethnicity | | |
|---|---|---|---|---|---|---|
| | | Never Reported | Ever Reported | White | Black | Hispanic |
| N | 1,232,220 | 970,304 | 261,916 | 778,211 | 279,302 | 106,302 |
| Prop. Students | | 0.787 | 0.213 | 0.632 | 0.227 | 0.086 |
| *School Characteristics* | | | | | | |
| Charter School | 0.121 | 0.119 | 0.129 | 0.054 | 0.304 | 0.130 |
| | (0.326) | (0.324) | (0.336) | (0.226) | (0.460) | (0.337) |
| Total School Enrollment | 454.5 | 458.9 | 438.2 | 444.1 | 480.3 | 452.9 |
| | (229.0) | (226.8) | (236.4) | (208.7) | (281.1) | (226.6) |
| Number Teachers | 26.258 | 26.518 | 25.293 | 25.860 | 26.835 | 26.650 |
| | (11.547) | (11.497) | (11.683) | (10.714) | (13.680) | (11.515) |
| Title I School | 0.486 | 0.438 | 0.667 | 0.413 | 0.690 | 0.594 |
| | (0.500) | (0.496) | (0.471) | (0.492) | (0.462) | (0.491) |
| Mean Std. Test Score (Reading) | -0.001 | 0.040 | -0.153 | 0.127 | -0.368 | -0.109 |
| | (0.442) | (0.433) | (0.443) | (0.347) | (0.471) | (0.421) |
| Mean Std. Test Score (Math) | -0.005 | 0.039 | -0.171 | 0.137 | -0.421 | -0.123 |
| | (0.484) | (0.475) | (0.480) | (0.375) | (0.507) | (0.451) |
| Urban District | 0.280 | 0.269 | 0.320 | 0.156 | 0.568 | 0.396 |
| | (0.449) | (0.444) | (0.466) | (0.363) | (0.495) | (0.489) |
| Suburban District | 0.501 | 0.525 | 0.415 | 0.550 | 0.386 | 0.445 |
| | (0.500) | (0.499) | (0.493) | (0.497) | (0.487) | (0.497) |
| Rural District | 0.219 | 0.206 | 0.266 | 0.294 | 0.046 | 0.158 |
| | (0.413) | (0.404) | (0.442) | (0.456) | (0.209) | (0.365) |
| *Census Tract Characteristics* | | | | | | |
| Prop. BA+ | 0.282 | 0.299 | 0.218 | 0.304 | 0.204 | 0.234 |
| | (0.196) | (0.204) | (0.149) | (0.192) | (0.164) | (0.178) |
| Prop. Married | 0.469 | 0.480 | 0.427 | 0.522 | 0.321 | 0.438 |
| | (0.158) | (0.157) | (0.155) | (0.126) | (0.151) | (0.140) |
| Unemployment Rate | 0.073 | 0.068 | 0.091 | 0.056 | 0.124 | 0.076 |
| | (0.063) | (0.059) | (0.070) | (0.039) | (0.088) | (0.058) |
| Median HH Income | 58,888 | 61,755 | 48,265 | 64,752 | 41,660 | 51,851 |
| | (29,290) | (30,391) | (21,710) | (28,186) | (22,968) | (24,996) |
| Prop. Public Assistance | 0.029 | 0.027 | 0.036 | 0.024 | 0.042 | 0.036 |
| | (0.030) | (0.029) | (0.032) | (0.023) | (0.038) | (0.035) |
| Family Poverty Rate | 0.125 | 0.115 | 0.160 | 0.089 | 0.222 | 0.156 |
| | (0.125) | (0.121) | (0.133) | (0.089) | (0.154) | (0.131) |

*Note:* The sample includes students observed in kindergarten through fifth grade in Michigan public or charter schools between the 2017–18 and 2023–24 school years. We report characteristics for each student in their first observed year. "Ever reported" indicates whether the student was ever the subject of a CPS report through fifth grade. School characteristics are measured at the school attended in the observation year. Census tract characteristics are based on the student's residential location. Standardized test scores are normalized to have mean zero and standard deviation one within grade and year. Median household income is measured in 2025 dollars. Standard deviations are reported in parentheses.

## Table A2: Prediction Model for CPS Reports

| | School Staff | Summer, Non-School Staff |
|---|---|---|
| N (Student x Year) | 3,448,920 | 3,448,920 |
| School Fixed Effects | X | X |
| County x Year Fixed Effects | X | X |
| Adjusted R-Squared | 0.061 | 0.052 |
| *Total Number Reports Pre-KG* | | |
| All Reports | 0.0157*** (0.0005) | 0.0128*** (0.0004) |
| Substantiated Reports | -0.0074*** (0.0014) | -0.0132*** (0.0013) |
| *Lagged # School-Based Reports* | | |
| 1 Year Ago | 0.0587*** (0.0016) | 0.0470*** (0.0013) |
| 2 Years Ago | 0.0318*** (0.0013) | 0.0191*** (0.0010) |
| 3+ Years Ago | 0.0049*** (0.0006) | 0.0004 (0.0005) |
| *Lagged # Substantiated School-Based Reports* | | |
| 1 Year Ago | -0.0175*** (0.0031) | 0.0137*** (0.0028) |
| 2 Years Ago | -0.0060** (0.0025) | 0.0088*** (0.0022) |
| 3+ Years Ago | 0.0044*** (0.0014) | 0.0078*** (0.0013) |
| *Lagged # Non-School Reports* | | |
| 1 Year Ago | 0.0243*** (0.0024) | 0.0465*** (0.0023) |
| 2 Years Ago | 0.0220*** (0.0020) | 0.0189*** (0.0019) |
| 3+ Years Ago | -0.0017* (0.0010) | 0.0001 (0.0008) |
| *Lagged # Substantiated Non-School Reports* | | |
| 1 Year Ago | 0.0058 (0.0045) | 0.0035 (0.0043) |
| 2 Years Ago | 0.0086** (0.0037) | 0.0092*** (0.0034) |
| 3+ Years Ago | 0.0077*** (0.0020) | 0.0095*** (0.0017) |
| *Grade (omitted: KG)* | | |
| 1st Grade | 0.0051*** (0.0004) | 0.0038*** (0.0003) |
| 2nd Grade | 0.0064*** (0.0004) | 0.0042*** (0.0003) |
| 3rd Grade | 0.0068*** (0.0005) | 0.0036*** (0.0003) |
| 4th Grade | 0.0060*** (0.0005) | 0.0036*** (0.0003) |
| 5th Grade | 0.0057*** (0.0005) | 0.0035*** (0.0004) |
| *Race (omitted: white)* | | |
| Black | 0.0026*** (0.0004) | -0.0007** (0.0003) |
| Hispanic | -0.0007 (0.0004) | -0.0009*** (0.0003) |
| Other Race | 0.0000 (0.0004) | -0.0004 (0.0003) |
| Female | -0.0015*** (0.0002) | 0.0005*** (0.0001) |
| Age in KG | 0.0001 (0.0002) | -0.0004** (0.0002) |
| Poor | 0.0157*** (0.0003) | 0.0075*** (0.0002) |
| IEP | 0.0155*** (0.0004) | 0.0015*** (0.0002) |
| ELL | -0.0052*** (0.0004) | -0.0050*** (0.0002) |
| Ever in MI Pre-K Program | -0.0030*** (0.0003) | -0.0019*** (0.0002) |
| Ever in Head Start | 0.0044*** (0.0005) | 0.0019*** (0.0003) |
| Ever in Great Start Readiness Pgm | -0.0014*** (0.0002) | -0.0006*** (0.0002) |

| | | |
|---|---|---|
| Same School Last Year | -0.0067*** | -0.0053*** |
| | (0.0004) | (0.0003) |
| Same School Two Years Ago | -0.0023*** | -0.0011*** |
| | (0.0003) | (0.0002) |
| Same School 3+ Years Ago | -0.0020*** | -0.0006** |
| | (0.0003) | (0.0003) |
| *Student Neighborhood Characteristics* | | |
| Prop. Adults BA+ | 0.0134 | -0.0047 |
| | (0.0123) | (0.0101) |
| Prop. Married | -0.0198 | 0.0037 |
| | (0.0123) | (0.0101) |
| Unemployment Rate | 0.0016 | -0.0002 |
| | (0.0026) | (0.0023) |
| Log Median Household Income | 0.0001 | 0.0000 |
| | (0.0001) | (0.0001) |
| Prop. Receiving Public Assistance | 0.0045 | 0.0060* |
| | (0.0044) | (0.0035) |
| Family Poverty Rate | 0.0066*** | -0.0006 |
| | (0.0015) | (0.0013) |
| *School Leave-Out Average Characteristics* | | |
| Non-School Reports Per Student Last Year | -0.0004 | -0.0186 |
| | (0.0154) | (0.0113) |
| Prop. Students Black | -0.0024 | -0.0034 |
| | (0.0074) | (0.0044) |
| Prop. Students Hispanic | -0.0021 | 0.0058 |
| | (0.0084) | (0.0055) |
| Prop. Students Other Race | 0.0057 | 0.0010 |
| | (0.0071) | (0.0047) |
| Prop. Students Female | -0.0005 | 0.0003 |
| | (0.0049) | (0.0035) |
| Average Age in KG | -0.0025 | -0.0025 |
| | (0.0048) | (0.0034) |
| Prop. Students Poor | -0.0069* | 0.0030 |
| | (0.0035) | (0.0023) |
| Prop. Students IEP | -0.0015 | 0.0035 |
| | (0.0059) | (0.0041) |
| Prop. Students ELL | 0.0055 | -0.0092** |
| | (0.0056) | (0.0039) |
| Prop. Students Ever in MI Pre-K Program | -0.0006 | -0.0028 |
| | (0.0036) | (0.0024) |
| Prop. Students Ever in Head Start | -0.0089 | -0.0040 |
| | (0.0055) | (0.0039) |
| Prop. Students Ever in Great Start Readiness Pgm | -0.0061* | -0.0038* |
| | (0.0036) | (0.0023) |
| Prop. Students Same School Last Year | 0.0028 | 0.0038** |
| | (0.0023) | (0.0016) |
| Prop. Students Same School Two Years Ago | 0.0060* | -0.0006 |
| | (0.0032) | (0.0019) |
| Prop. Students Same School 3+ Years Ago | -0.0046 | -0.0012 |
| | (0.0037) | (0.0023) |
| Average Prop. Adults BA+ | -0.0659 | 0.0597 |
| | (0.0587) | (0.0467) |
| Average Prop. Married | 0.0593 | -0.0566 |
| | (0.0580) | (0.0465) |
| Average Unemployment Rate | -0.0221* | -0.0010 |
| | (0.0124) | (0.0082) |
| Average Median Household Income | 0.0006 | 0.0001 |
| | (0.0007) | (0.0005) |
| Average Prop. Receiving Public Assistance | 0.0091 | -0.0080 |
| | (0.0225) | (0.0125) |
| Average Family Poverty Rate | -0.0041 | 0.0056 |
| | (0.0092) | (0.0059) |
| *Urbanicity (vs. Suburban)* | | |
| Urban | -0.0219*** | 0.0043*** |
| | (0.0015) | (0.0010) |
| Rural | 0.0127*** | -0.0046*** |
| | (0.0025) | (0.0015) |

*Note:* The sample includes K–5 students enrolled in Michigan public schools from 2018 to 2024, excluding the 2020–21 school year. The dependent variable is an indicator for whether a student receives a CPS report in a given year, separately for reports filed by school staff and non-school reporters. Observations are at the student-year level, and we drop observations after a student's first report while attending a given school. All regressions include school and county-by-year fixed effects. Standard errors are clustered at the school level. Additional controls include student demographics, prior CPS involvement, school characteristics, and neighborhood characteristics. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: School Characteristics and Estimated Reporting Skill and Preferences

| | Skills ($\alpha_j$) | Preferences ($\beta_j$) |
|---|---|---|
| Prop. Teachers Female | 0.013 | 0.005 |
| | (0.022) | (0.023) |
| Prop. Teachers Black | 0.187*** | 0.077*** |
| | (0.025) | (0.022) |
| Prop. Teachers Hispanic | 0.101*** | 0.052** |
| | (0.016) | (0.020) |
| Prop. Teachers Other Race | 0.062*** | 0.078*** |
| | (0.018) | (0.020) |
| Prop. Principals Black | 0.293*** | 0.222*** |
| | (0.080) | (0.069) |
| Prop. Principals Hispanic | 0.690*** | 0.513*** |
| | (0.125) | (0.137) |
| Prop. Principals Other Race | 0.304** | 0.188 |
| | (0.119) | (0.197) |
| Average Teacher Age | 0.035 | -0.005 |
| | (0.023) | (0.024) |
| Teacher Years at School | -0.057** | -0.061** |
| | (0.024) | (0.025) |
| Prop. Teacher Master's Degree | -0.048* | -0.028 |
| | (0.027) | (0.024) |
| Student-Teacher Ratio | -0.065 | -0.037 |
| | (0.063) | (0.045) |
| Charter | 0.048 | 0.167** |
| | (0.093) | (0.075) |
| Title I | 0.416*** | 0.414*** |
| | (0.043) | (0.056) |
| Math Test Score VA | -0.203*** | -0.119*** |
| | (0.026) | (0.032) |
| ELA Test Score VA | -0.288*** | -0.167*** |
| | (0.030) | (0.036) |
| Attendance VA | -0.247*** | -0.406*** |
| | (0.025) | (0.038) |
| School Has Counselor | -0.131*** | -0.093 |
| | (0.047) | (0.060) |
| School Has Nurse | 0.585*** | 0.597*** |
| | (0.103) | (0.157) |
| N | 10,922 | 10,922 |

*Note:* Each coefficient reports the association between a school characteristic and the estimated skill ($\alpha_j$) or preference ($\beta_j$) parameter from the structural model. Each characteristic is entered separately in a regression that includes county-by-year fixed effects. Continuous variables are standardized to have mean zero and unit variance. The unit of observation is a school-year. Standard errors, clustered at the school level, are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Correlation of Estimated Parameters Across Specifications

| | $\widehat{\alpha}_j$ | $\widehat{\beta}_j$ | $\widehat{\tau}_j$ |
|---|---|---|---|
| Substantiations | 0.925 | 0.844 | 0.976 |
| Re-Mean Within State | 0.808 | 0.833 | 0.759 |
| All Reports (Any Reporter) | 0.997 | 0.979 | 0.999 |
| All Substantiations (Any Reporter) | 0.921 | 0.829 | 0.976 |
| $\kappa = 0.4$ | 0.999 | 0.999 | 0.999 |

*Note:* Each entry reports the correlation between school-level parameter estimates from an alternative specification and the baseline model. The baseline model defines true positives using school-based reports that are screened in for investigation and uses risk-adjusted reporting outcomes. Columns report correlations for estimated skill ($\widehat{\alpha}_j$), preferences ($\widehat{\beta}_j$), and implied reporting thresholds ($\widehat{\tau}_j$). "Substantiations" defines true positives using substantiated cases rather than screened-in reports. "Re-Mean Within State" re-centers risk-adjusted reporting measures at the state rather than county level. "All Reports" and "All Substantiations" redefine reporting outcomes using information from all reports received by a student within the school year, regardless of reporter type. Under these specifications, a school report is treated as successful whenever any report received during the school year is screened in (or substantiated), even if the report meeting that criterion was filed by a non-school reporter. The $\kappa = 0.4$ specification assumes that 40 percent of students are not at risk of CPS involvement, where 40 percent is defined as one minus the maximum share of students ever reported to CPS across schools between kindergarten and fifth grade. This provides a conservative lower bound on the fraction of students plausibly outside the risk set.

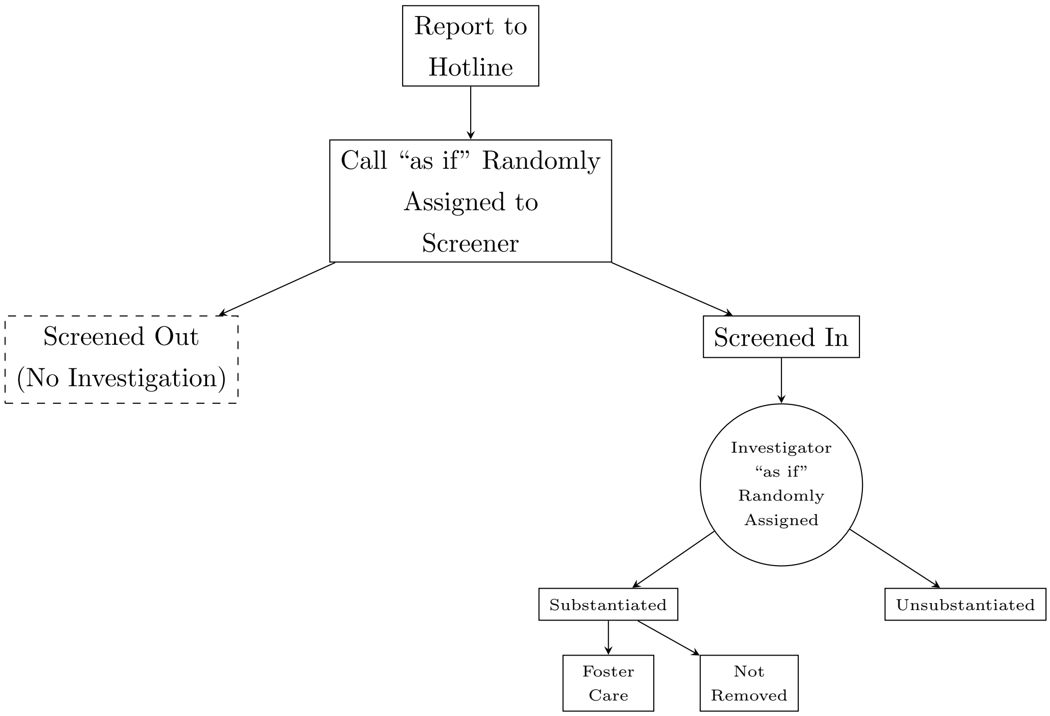
Table A5: Relationship Between Chan, Gentzkow and Yu (2022) and Currie, MacLeod and Musen (2026) Measures of Skill and Preferences

| | Reduced-Form Measure | |
|---|---|---|
| | Skill ($a_j$) | Preferences ($B_j$) |
| Structural Skill ($\alpha_j$) | 0.280 | 0.285 |
| Structural Preferences ($\beta_j$) | 0.218 | 0.421 |
| Reports-Added | 0.170 | 0.675 |
| Observations | 1,815 | 1,815 |

*Note:* The table reports pairwise correlations between the reduced-form measures of school skill and preferences motivated by Currie, MacLeod and Musen (2026) and the corresponding structural estimates from the CGY framework. The reduced-form skill measure captures how strongly a school's reporting behavior responds to predicted student risk, while the reduced-form preference measure captures reporting rates for students with very low predicted risk.

# B Additional Figures

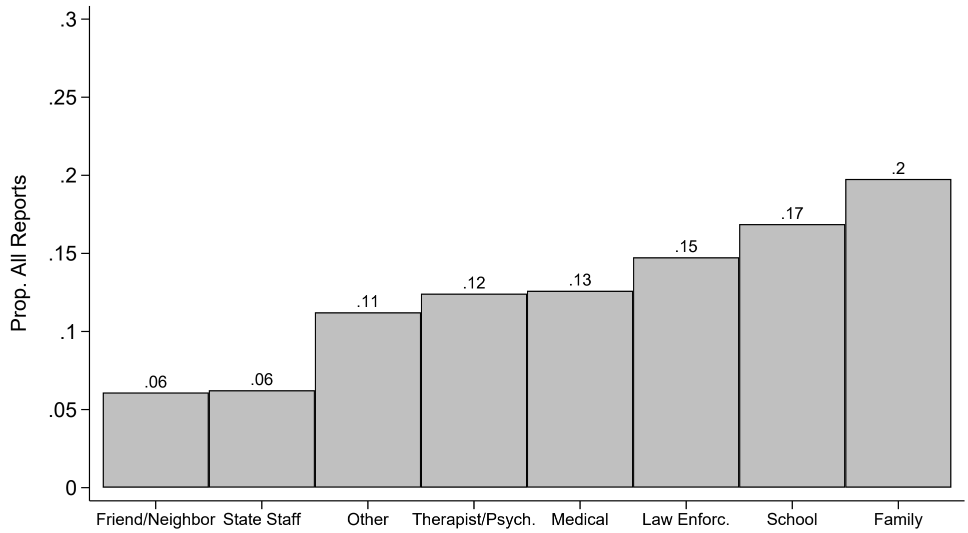
Figure A1: Child Protective Services Reporting and Investigation Process in Michigan



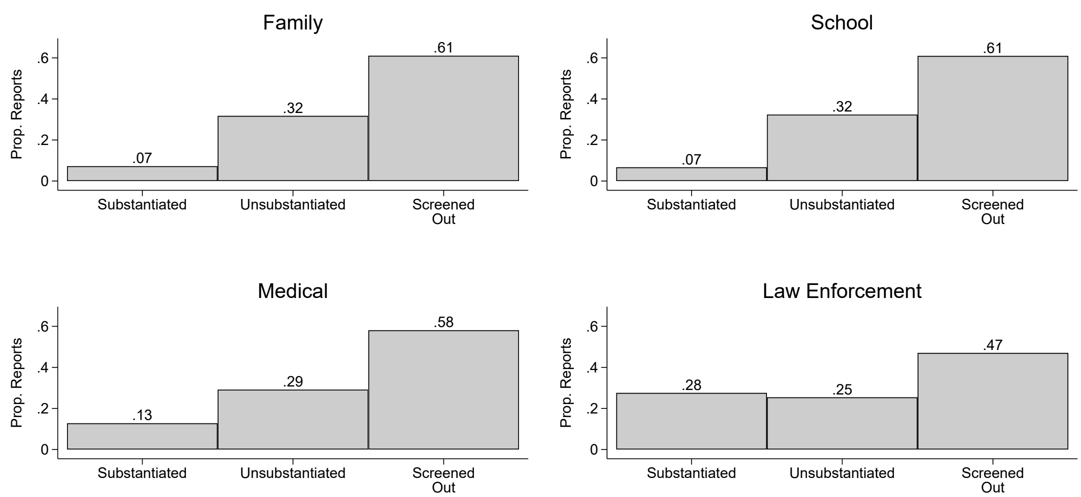

*Note:* The figure illustrates the sequence of decisions in Michigan's Child Protective Services (CPS) system. Reports are made to a centralized hotline and assigned to screeners, who determine whether the case is screened in for investigation. Screened-in cases are assigned to investigators, who determine whether allegations are substantiated and whether further action, such as foster care placement, is warranted. Approximately 40 percent of reports are screened in for investigation, and about 25 percent of investigations are substantiated.

Figure A2: CPS Reporting and Outcomes by Reporter Type

(a) Distribution of CPS Reports by Reporter Type

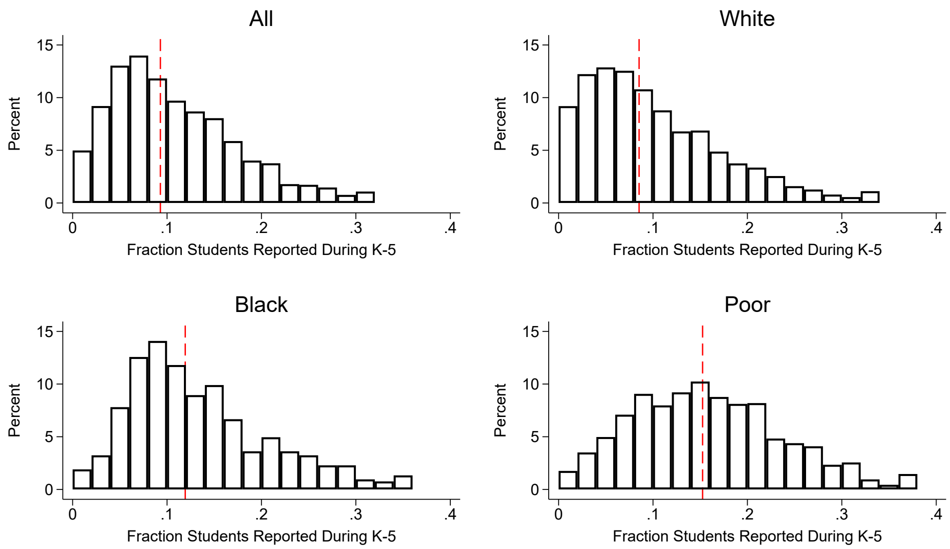



(b) Investigation and Substantiation Rates by Reporter Type

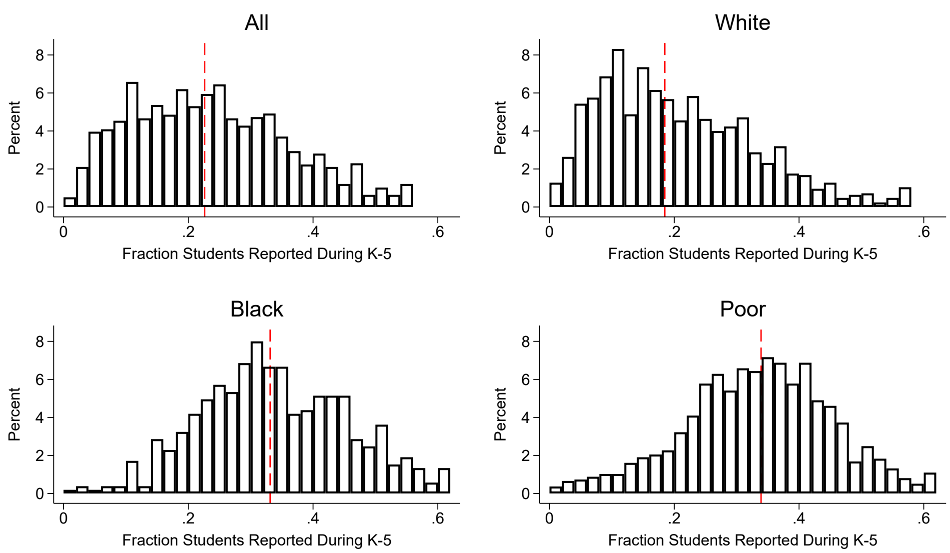



*Note:* Panel A shows the distribution of CPS reports by reporter type for Michigan public and charter school students between 2017 and 2024. Panel B shows the share of reports that are screened in for investigation and the share that result in substantiated findings, by reporter type. Reporter type is recorded at the time of the call and includes categories such as school staff, medical professionals, law enforcement, family members, and other community reporters. The sample excludes reports filed between March 2020 and June 2021 due to disruptions associated with the COVID-19 pandemic.

Figure A3: Variation in School and Non-School CPS Reporting Across Schools



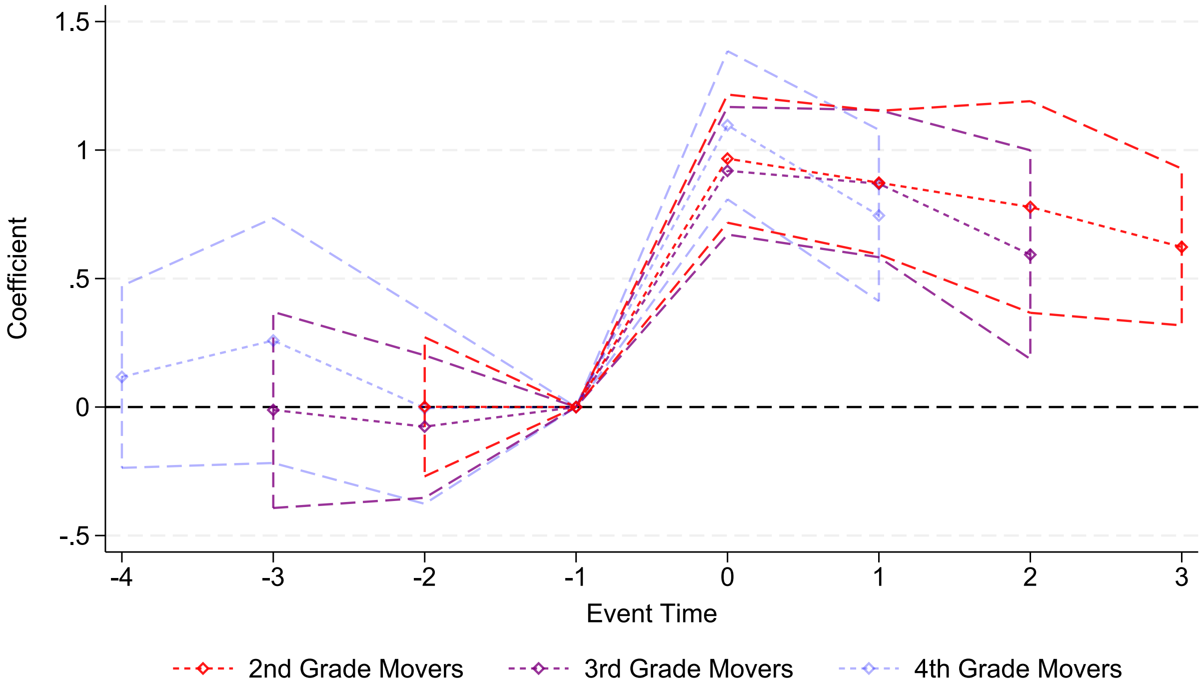

(a) School Staff Reports




(b) Non-School Reports

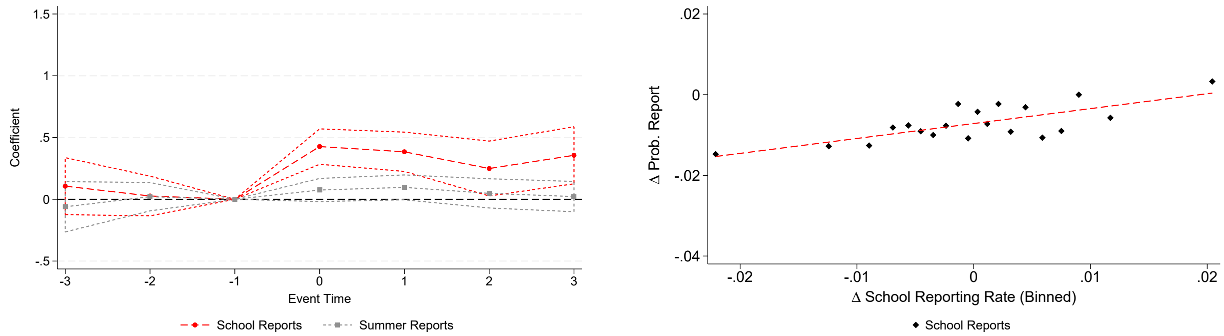
*Note:* The sample includes students who entered kindergarten in a Michigan public or charter school in Fall 2017 or Fall 2018 and are observed through fifth grade. Panels (a) and (b) display the distribution across schools of the percent of students reported to CPS at least once between kindergarten and fifth grade by school staff and non-school reporters, respectively. Schools are ordered by their overall reporting rates, and the red dashed lines indicate the median school. Schools with fewer than 25 students meeting the sample criteria are excluded, as are reports made between March 2020 and June 2021.

Figure A4: Evidence on Changes in Reporting Around School Switches by Grade




*Note:* The figure reports event-study estimates of changes in student-level CPS reports when students switch schools, separately by the grade in which students moved. Event-study coefficients are estimated from regressions of an indicator for receiving a school-based CPS report on event-time indicators interacted with the difference between the destination and origin school's reports-added. The sample includes students who switch schools once during elementary school (in second, third, or fourth grade). The independent variable is constructed using only students who never switch schools.

Figure A5: Evidence on Changes in Unadjusted Reporting Around School Switches

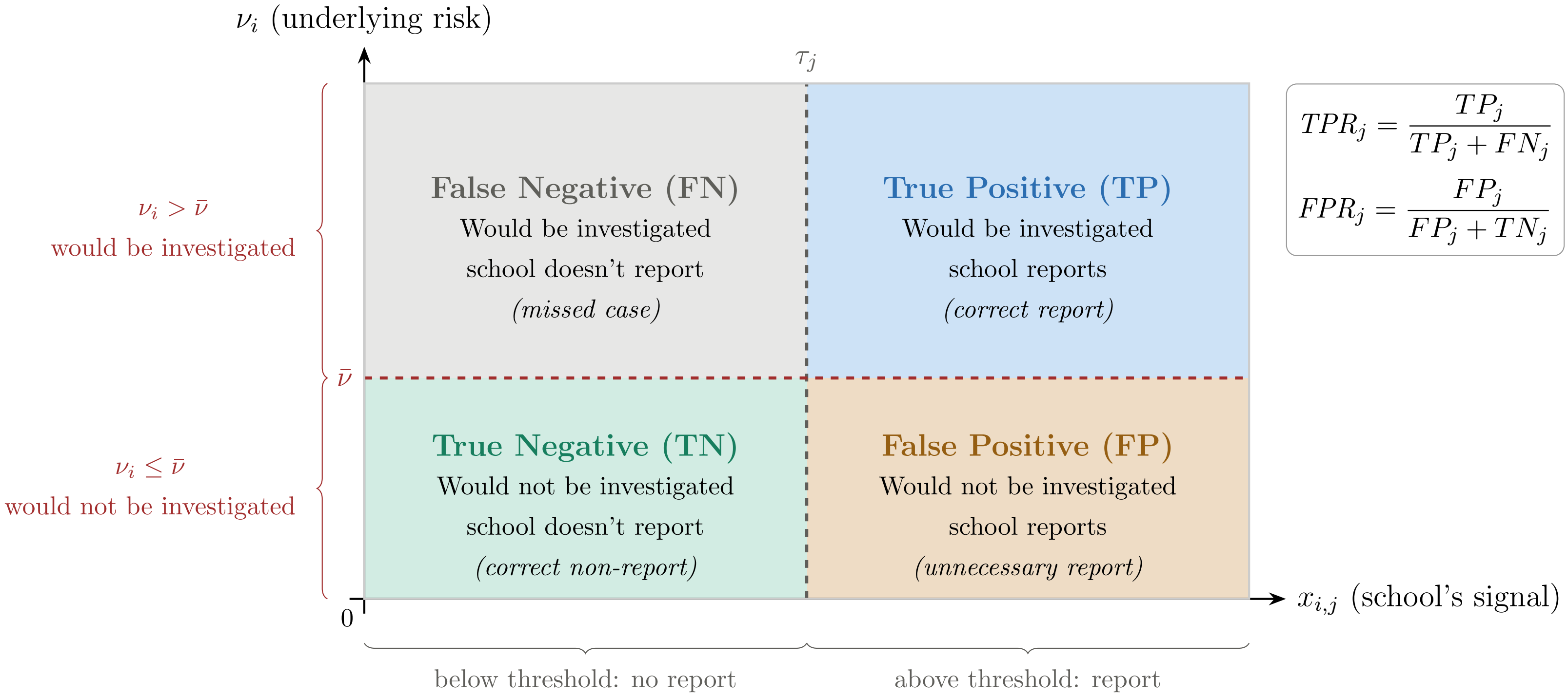



(a) Event Study

(b) Binscatter

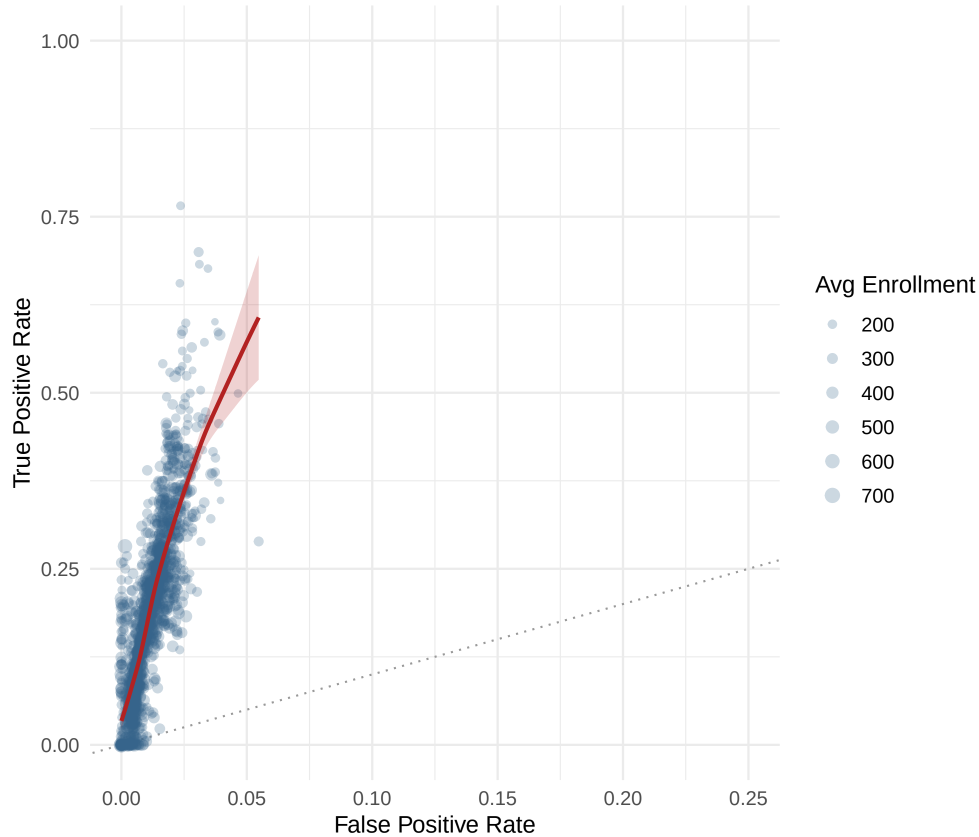
*Note:* The figure reports event-study and binscatter estimates of changes in student-level CPS reporting following school switches. Unlike Figure 3, both the dependent variable and the school-level reporting measure are based on unadjusted reporting rates. Event-study coefficients are estimated from regressions of an indicator for receiving a school-based CPS report on event-time indicators interacted with the difference in the destination and origin schools' average reporting rates (shrunken using empirical Bayes). Panel (a) also reports placebo estimates based on reports filed by non-school reporters during the preceding summer. The event-study specifications include grade and year fixed effects, the main effect of the change in school reporting rates, and standard errors clustered at the student and destination-school levels. Panel (b) plots the year-over-year change in the probability of receiving a school-based CPS report against the change in school reporting rates induced by the school switch. The sample includes students who switch schools exactly once during elementary school (in second, third, or fourth grade). School reporting rates are estimated using only students who never switch schools.

Figure A6: Reporting as a Statistical Classification Problem



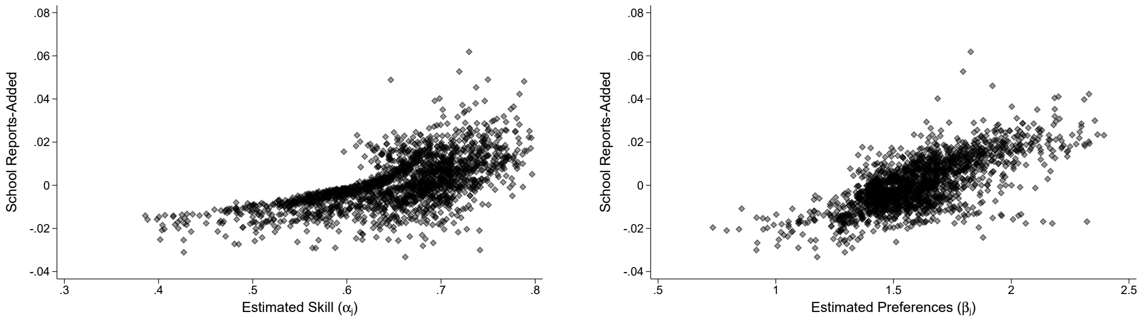

*Note:* The figure illustrates the reporting decision as a statistical classification problem. Each child has an underlying level of risk $\nu_i$, and schools observe a noisy signal $x_{i,j}$ of that risk. Cases with $\nu_i > \bar{\nu}$ are those that would be screened in for investigation by CPS if reported. Schools report when the signal exceeds a threshold $\tau_j$, generating true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). The true positive rate ($TPR$) and false positive rate ($FPR$) summarize reporting performance conditional on the underlying state.

Figure A7: School True and False Positive Rates in ROC Space With Fitted LOESS Curve




*Note:* The figure plots each school's true positive rate (TPR) against its false positive rate (FPR), assuming a population prevalence of cases warranting investigation of $\widehat{S} = 0.062$, as estimated in the structural model. Each point represents a school. Schools that achieve both higher true positive rates and lower false positive rates than other schools lie on higher ROC curves and are therefore more skilled at identifying cases that warrant investigation. Estimates are risk-adjusted. The LOESS curve summarizes the relationship between TPR and FPR across schools, with 95% confidence interval based on 1,000 bootstrap replications. To reduce the influence of measurement error, the sample is restricted to schools with at least 100 students per year.

Figure A8: Correlation of Reporting Skill and Preferences with Reports-Added




(a) Skill (b) Preferences

*Note:* The figure plots the relationship between school-level reports-added and estimated structural parameters. Panel A shows reporting skill ($\alpha_j$), and Panel B shows reporting preferences ($\beta_j$). Each point represents a school. We omit schools above the 99th percentile of reports-added for visual clarity.

# C Structural Model Estimation Details

We estimate the model using school-level data on reporting behavior and downstream CPS responses to those reports. For each school $j$, let $n_j^{TP}$, $n_j^{FP}$, and $n_j^{NR}$ denote the risk-adjusted counts of true positives, false positives, and non-reports, respectively. Given $(\alpha_j, \beta_j, \bar{\nu})$, the likelihood of these outcomes is given by the multinomial likelihood implied by $(p_{1j}, p_{2j}, p_{3j})$:

$$L_j(\alpha_j, \beta_j; \bar{\nu}) = p_{1j}^{n_j^{TP}} \; p_{2j}^{n_j^{FP}} \; p_{3j}^{n_j^{NR}}. \tag{7}$$

Because each school contributes a finite number of observations, we assume that $(\alpha_j, \beta_j)$ are drawn from a population distribution. Specifically,

$$\tilde{\alpha}_j \sim N(\mu_\alpha, \sigma_\alpha^2), \quad \alpha_j = \frac{1 + \tanh(\tilde{\alpha}_j)}{2}, \tag{8}$$

$$\tilde{\beta}_j \sim N(\mu_\beta, \sigma_\beta^2), \quad \beta_j = \exp(\tilde{\beta}_j), \tag{9}$$

which ensures $\alpha_j \in (0, 1)$ and $\beta_j > 0$.

Let $\theta = (\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta, \bar{\nu})$ denote the vector of hyperparameters. We estimate $\theta$ by maximizing the integrated likelihood:

$$\hat{\theta} = \arg\max_\theta \sum_j \log \int L_j(\alpha, \beta; \bar{\nu}) \, dF(\alpha, \beta; \theta), \tag{10}$$

where $F$ is the joint distribution of $(\alpha_j, \beta_j)$ implied by the parameterization above. As in Chan, Gentzkow and Yu (2022), we approximate the integral numerically using a discrete grid over $(\alpha, \beta)$ and apply empirical Bayes shrinkage to account for finite sampling variation. We obtain standard errors using a school-level bootstrap.

To account for non-random assignment of students across schools, we construct risk-adjusted counts of reporting and screen-in outcomes using school-level value-added estimates. Specifically, we use value-added measures for both the probability of being reported and the probability of being screened in conditional on reporting, and scale these by school enrollment to obtain adjusted counts of true positives and false positives. We then re-center these counts relative to the county distribution so that each school's outcomes are expressed relative to its local population baseline. Because we estimate a single statewide prevalence $\widehat{S}$, we show that results are similar when re-centering at the state level instead. We bottom-code a small number of schools with negative adjusted counts and restrict the sample to schools with at least one report, for which screen-in rates are well defined.[22]

# D Currie–MacLeod–Musen Measures of Skill and Preferences

To provide additional intuition for what the structural model's skill and preference parameters capture, we also implement a reduced-form approach motivated by Currie and MacLeod (2017) and Currie, MacLeod and Musen (2026). In this framework, schools with greater "skill" are those whose reporting decisions respond more strongly to underlying student risk, while schools with stronger reporting "preferences" report more low-risk students. As discussed by Currie, MacLeod and Musen (2026), these objects can also be interpreted through the lens of true and false positive rates, though they rely on different identifying assumptions than our structural model. Importantly, the reduced-form approach yields the same main qualitative conclusion as the structural model: schools with higher estimated skill tend to report students at higher rates.

Specifically, we estimate

$$d_{ijt} = B_j + a_j \widehat{X}_{it} + \varepsilon_{ijt}, \tag{11}$$

where $d_{ijt}$ is an indicator for whether school $j$ reports student $i$ to CPS in year $t$, $\widehat{X}_{it}$ is a predicted measure of student maltreatment risk, $B_j$ captures school-specific reporting propensity, and $a_j$ captures school-specific responsiveness to risk.

[22] We omit approximately ten percent of schools without well-defined screen-in rates. These schools are typically small, enrolling only three percent of all students.

We estimate $\widehat{X}_{it}$ using the same linear probability model used to estimate school reports-added. We focus on the hazard sample of students who had not previously been reported at their current school. In practice, we estimate school-specific slopes and intercepts by interacting predicted risk $\widehat{X}_{it}$ with school fixed effects, controlling for student covariates predictive of risk in the regression. In this framework, the intercept $B_j$ captures the probability that school $j$ reports a student with very low predicted risk and can therefore be interpreted as reporting preferences or reporting propensity. The slope coefficient $a_j$ captures how strongly reporting behavior responds to underlying student risk and is interpreted as diagnostic skill.[23] Currie, MacLeod and Musen (2026) show that these objects can be related to false positive and true positive rates:

$$B_j = FPR_j, \tag{12}$$

and

$$a_j = TPR_j - FPR_j. \tag{13}$$

Finally, we apply empirical Bayes shrinkage to the slopes and intercepts using their estimated standard errors, analogous to our construction of reports-added measures. The variance in true school parameter $\omega$ is estimated as the observed variance in the parameter minus the average variance (the square of the standard error) of each school's estimated parameter, $\sigma^2_\omega = \mathrm{Var}(\omega) - \overline{\mathrm{Var}(\widehat{\omega}_j)}$. The noise component for each school is the school-specific variance of $\omega$, $\mathrm{Var}(\widehat{\omega}_j)$. The shrinkage factor for school $j$ is then: $\lambda = \frac{\sigma^2_\omega}{\sigma^2_\omega + \mathrm{Var}(\widehat{\omega}_j)}$.

Table A5 shows the correlations between the reduced-form and the structural model's estimates of skill and preferences. The reduced-form skill measure is positively correlated with the structural model's estimate of diagnostic skill, while the reduced-form preference measure is positively correlated with the structural model's estimate of reporting preferences. Importantly, both frameworks yield the same qualitative conclusion: schools with higher estimated skill also tend to report students at higher rates.

[23] Note that this measure of skill is just the slope coefficient itself, not the sum of $a_j$ and $B_j$.