Households in Conflict Network
hicn.org

Polling Place Matters:
How Voting Location Influences Election Fraud

Tannishtha Ghosh[1] Robert M. Gonzalez[2] and Kelly Lifchez[3]

HiCN Working Paper 458

July 2026

Abstract
Research on the determinants of election fraud typically focuses on conventional factors such as election monitoring, legal punishments, and voter education. This paper examines an often-overlooked factor: whether polling location (e.g., school, place of worship) influences election fraud. Combining validated fraud measures from the 2009 Afghan presidential election with a novel instrumental variable approach, we find that polling centers within schools report an 8 percentage-point lower likelihood of fraud compared to those within mosques. Two mechanisms may explain this difference. First, mosques designated as polling centers were more likely to be attacked by the Taliban, likely suppressing turnout and creating incentives for fraudulent votes through a vote-substitution channel. Second, election-related complaint data indicate systematic differences in the behavior of voters and polling officials between schools and mosques. Compared to mosques, voters at schools are more willing to file complaints, complainants are more often women, and polling officials are less likely to be the subject of complaints.

JEL Classifications
D72, D73, K42

Keywords
election fraud, polling location, insurgent violence

∗School of Economics, Georgia Institute of Technology. E-mail: tghosh32@gatech.edu
†School of Economics, Georgia Institute of Technology. E-mail: robert.gonzalez@gatech.edu
‡School of Economics, Georgia Institute of Technology. E-mail: klifchez3@gatech.edu

1 Introduction

Election fraud undermines democratic legitimacy and erodes public trust in institutions [Bank, 2017]. Understanding the determinants of election fraud is thus crucial for both researchers and policymakers. Existing studies primarily focus on conventional determinants of fraud, such as institutional capacity [Birch and Van Ham, 2017], electoral design [Gieczewski and Shadmehr, 2024], monitoring and enforcement mechanisms [Callen and Long, 2015, Gonzalez, 2021], and conflict [Weidmann and Callen, 2013]. While these factors are important, the role of contextual elements—such as the physical location of polling centers (e.g., a school, a place of worship)—has received little attention.

Contextual features of polling environments may influence both the opportunities for fraud and the selection and behavior of electoral agents. For example, polling centers located in places of worship may be perceived as more trustworthy or sacred, potentially encouraging polling officials to follow rules while leading authorities to reduce oversight under the assumption that misconduct is unlikely in such settings. Similarly, schools may prime notions of civic duty or benefit from stronger administrative oversight and infrastructure. This can shape both the selection and behavior of polling officials and voters. These environments can also affect the presence (or absence) of community monitors. As such, the physical and social setting of a polling center is not a neutral backdrop but an active component of the election process that may indirectly influence fraud.

This paper examines whether and how polling location type affects the likelihood of election fraud. We focus on the 2009 Afghan presidential election, a setting where fraud was both widespread and systematically documented. Using validated fraud measures, we compare polling centers in schools and mosques, the two most common types of polling locations.

We explore this question by relying on several unique sources of data. First, we combine primary data on the location and characteristics of polling centers collected during security assessments conducted by International Security Assistance Force (ISAF) inspection teams with polling-center-level fraud measures from a UN-sponsored audit conducted shortly after the 2009 Afghan presidential election. We classify polling center types (school, mosque, private house, or other) using a combination of natural language processing (NLP) models and rule-based classification applied to the recorded names of polling centers. We focus on Kabul province in our study because polling center classification was more accurate there due to higher-quality reporting of center names and fewer ambiguous center names than in other large provinces.

Research on this question faces a key issue: the endogeneity of polling center type.[1] Two factors related to our context help mitigate endogeneity concerns. First, information on the location of polling centers was withheld from the public until one week before the election to reduce the risk of Taliban attacks [National Democratic Institute, 2010]. This limited the ability of political actors and voters to adjust their behavior based on polling center type. Second, we further rely on the ISAF security assessment data, which document the two key determinants of polling center choice: the security environment in the area (including security and the defensibility of the location) and accessibility. We control for these factors in our regressions by including measures of pre-election violence near the polling center, topographic characteristics that affect defensibility, and whether the center was accessible. Additionally, all specifications include district fixed effects to account for confounding factors that vary across districts.

To strengthen causal identification, we also implement a novel instrumental variable (IV) strategy that exploits a sharp increase in the likelihood of school assignment when a location surpasses 3,000 expected voters. This threshold is operational: the Independent Election Commission (IEC) allocated one voting station for every 600 expected voters, and we observe a sharp increase in the likelihood of a polling center being assigned to a school once a center requires more than five stations (i.e., more than 3,000 expected voters). This is driven by logistical considerations rather than political manipulation—schools typically have more rooms that can be used as polling stations than mosques. We provide evidence that other polling center characteristics do not change significantly at this threshold.

Our findings indicate that the choice of polling center matters. We find significant evidence that fraud is less prevalent in schools than in mosques. Specifically, polling centers located within schools exhibit an 8 percentage-point lower likelihood of fraud than those located within mosques. Our findings are robust to different model specifications, including various functional forms and controls. Our results are also robust to different measures of fraud, suggesting that polling centers located in schools exhibit less fraud than those in mosques along both extensive and intensive margins. On the intensive margin, the results from linear models suggest that being in a school reduces the share of fraudulent votes by approximately 7 percentage points compared to the share within mosques.

Results from our IV strategy suggest that assigning a polling center to a school significantly reduces the incidence of electoral fraud relative to a mosque. Across specifications, school

[1] For example, schools may be assigned as polling centers in urban and better-governed areas, where fraud detection could be stronger.

locations are associated with a 10-17 percentage-point decrease in the share of fraudulent votes. These results are robust to controls, district fixed effects, and bandwidth restrictions around the 3,000-voter threshold. They also closely align with our OLS estimates, reinforcing our causal interpretation.

We explore several channels through which polling center location may influence fraud. First, we examine whether the effect operates through election-related violence.[2] Violence can facilitate fraud via a vote-substitution mechanism: if certain types of polling centers are more likely to be targeted, turnout may fall, prompting corrupt officials to replace lost votes with fraudulent ones. To test this channel, we use declassified Significant Actions (SIGACTs) data on insurgent attacks, which include precise timestamps, geolocations, and incident types. We also exploit the exogenous timing of the public release of polling center locations—one week before election day—to identify attacks that occurred after this disclosure. We find a sharp increase in attacks following the announcement, consistent with strategically targeted violence. Importantly, the pattern varies by location type: mosques are attacked at higher rates than schools and other sites. Using an IV causal mediation approach following Dippel et al. [2020], we estimate that 58–64% of the fraud-reducing effect of assigning a school as a polling center operates through lower exposure to post-announcement violence. Placebo tests using pre-announcement violence yield small and statistically insignificant indirect effects, reinforcing the interpretation that the violence channel plays a substantial role.

Second, we consider that differences in electoral fraud across polling centers may stem from variations in the composition and behavior of voters and polling center officials. While we lack data on individual voters and polling officials, we leverage polling center-level complaint records from the 2010 parliamentary election to infer behavioral differences. Using a Poisson specification, we find that voters in schools are more likely to report complaints than in mosques, with more complaints filed by women and individual citizens (rather than candidates or organizations). Complaints in schools are also more likely to implicate candidates rather than polling officials as suspects of fraud. These patterns suggest that schools may attract voters and officials who are intrinsically different than their mosque counterparts. Alternatively, this could also be interpreted as schools affecting the behavior of voters and officials, as reflected in the higher levels of engagement in the monitoring process.

This paper fits into the economics literature studying the determinants of election fraud and

[2]Election-related violence here excludes incumbent-driven voter intimidation. Although such violence occurs in other contexts, existing evidence indicates that in our setting, pre-election violence is largely perpetrated by non-state insurgent groups and affects incumbents and challengers symmetrically (Weidmann and Callen, 2013), thereby reducing concerns that our measure captures omitted strategic political violence.

possible deterrents [Callen and Long, 2015, Stockemer et al., 2013, James and Clark, 2019]. In Afghanistan, specifically, Weidmann and Callen [2013] study the relationship between violence and election fraud, finding evidence that the relationship follows an inverted U-shape and is sensitive to the security situations faced by incumbent and challenger networks. Gonzalez [2021] examines the impact of cell phone access on election fraud and suggests that increased access improves social monitoring capacity, thus decreasing fraud.

This paper also contributes to the literature on polling center locations and election outcomes. Studies within this literature focus on the impacts of contextual priming on voter behavior and election outcomes [Ajzenman and Durante, 2023, Berger et al., 2008]. While this literature examines how contextual factors affect electoral outcomes and voter behavior, our paper addresses how contextual factors affect fraud. To the best of our knowledge, this is the first paper to study whether contextual factors, such as the features of the voting location (e.g., whether it is a school or a place of worship), affect fraud.

2 Background on the 2009 Afghan Presidential Election

The 2009 election marked the second presidential election following the collapse of the Taliban regime in 2001. A distinguishing feature of this election was the creation of the Electoral Complaints Commission (ECC). The ECC was an independent body tasked with investigating and adjudicating fraud-related complaints. In addition, the ECC had authority to order audits, recounts, and runoff elections as needed [Electoral Complaints Commission, 2010]. To safeguard its independence during the 2009 election, three of the five ECC commissioners—including the chair—were international experts appointed by the UN Secretary-General's special representative. The remaining two were Afghan nationals selected from the Afghanistan Independent Human Rights Commission and the Supreme Court, respectively [National Democratic Institute, 2010].

The 2009 Afghan presidential election was marked by widespread allegations of electoral fraud, prompting an official audit process led by the Electoral Complaints Commission (ECC). A central concern was the pattern of reported vote totals at the polling station level. Each polling station had been issued a fixed supply of 600 blank ballots, yet an unexpectedly large number of stations reported exactly 600 or more votes. This pattern stood in contrast to the overall low voter turnout, which had been suppressed by fears of election-day violence. Figure 1 presents a histogram of total votes per station for Kabul province. Note the significant spike at exactly 600 total votes and the incidence of stations with more than 600 reported votes. Further concerns included a high incidence of polling stations—including those with 600 or more recorded votes—reporting over 95% support for a single candidate [Khadhouri, 2010].

In response, the ECC established three categories of suspicious stations based on (i) excessively high turnout, i.e., 600 or more votes cast (Category A), (ii) lopsided vote shares, i.e., 95% vote share for a single candidate (Category B), and (iii) the intersection of both, i.e., stations with more than 600 votes cast and more than 95% going to a single candidate (Category C). A total of 3,376 stations—nearly 15% of all polling stations—were flagged as potentially fraudulent. Given the urgency of determining whether a runoff election was necessary, the ECC audited a 10% random sample of the suspect stations. The audits involved physical inspection of ballot boxes and documentation, identifying signs of fraud such as broken seals, uniform ballot markings, and discrepancies between tally sheets and vote totals. Refer to Gonzalez [2021] for more details about the 2009 election and audit.

3 Data

We rely on two primary sources of data: the geolocation and names of polling centers collected by International Security Assistance Force (ISAF) inspection teams shortly after the election, and polling center level data on election fraud metrics created by a UN-sponsored audit following the election.

3.1 Measures of Fraud

As in Weidmann and Callen [2013] and Gonzalez [2021], we rely on the ECC suspected-fraud categories described above to construct our measure of fraud. Specifically, the two fraud outcomes that we define are: (i) the percentage of votes cast by stations falling under *Category A* at the polling station level. In particular, if a polling center c has s stations overall, of which $n \leq s$ fall under *Category A*, the number of votes cast in the n stations divided by the total number of votes cast in center c is the measure of fraud at center c. This measure is referred to as the *Share of Votes under Category A*, and (ii) an indicator for whether there is at least one *Category A* fraud station at polling center c.[3] Table A1 presents summary statistics for our primary outcomes and covariates. On average, 5% of votes at a polling center are cast in stations falling under our Category A fraud measure. Additionally, 9% of centers have at

[3] We note that our Category A fraud definition refers to stations that satisfy the original ECC categories A and C. In the original ECC definitions, Category A refers to stations that report 600 or more votes and are not included in Category C. For our study, we redefine Category A to include both Category A and C stations so that we account for all stations for which 600 or more votes are reported.

least one such station affected by fraud. These figures indicate that while outright fraud is not widespread across all centers, a nontrivial number of polling centers exhibit signs of significant manipulation.

We focus on Category A for our analysis because, by definition, it has a higher incidence than Category C. In our sample of 507 polling centers in Kabul Province, approximately 9% report Category A fraud, while about 2% report Category C fraud. Using Category A ensures reasonable variation in our outcome of interest.[4]

A station's eligibility for one of the categories does not necessarily indicate that fraud occurred there. For example, there might be stations with exceptionally high voter turnout or significant partiality toward a particular candidate. In light of this, our measure of fraud should be interpreted as a proxy for fraud. However, we are confident that our proxy measure presents an accurate indication of fraud for two key reasons. First, of all the audited stations qualifying as Category A (our measure of fraud), 83% showed clear evidence of tampering and were deemed by electoral experts to be fraudulent [Electoral Complaints Commission, 2010].[5] Second, within the same Afghan setting, these measures have been cross-validated with digit-based fraud measures commonly used in the election fraud measurement literature [Weidmann and Callen, 2013]. We recognize that although these reasons mitigate concerns, there may still be measurement error in our fraud outcome. Given this, we also present results using an instrumental variable approach that accounts for any correlation between polling center choice and measurement error in our fraud outcome.

3.2 Classifying Polling Location Type

We classify polling center location type using rule-based and NLP methods that leverage keywords within the polling center names to assign each center to a predefined category. For instance, if the word "mosque" or "madrassa" is found in a polling center name, the center is categorized as a "mosque". Certain categories are consolidated to streamline analysis. For instance, all school types (primary, high, secondary, university, other) are grouped under a single category, "schools," simplifying the classification to four main types: Mosque, Schools, House, and Other. To address variation in findings across different types of schools, we further employ

[4] Appendix Table A2 presents results using Category C as the outcome, and results are consistent with our main results, albeit less significant, as expected.

[5] This number comes from calculations made by the authors based on the reported results on page 37 of the ECC post-election report [Electoral Complaints Commission, 2010]. Specifically, there were 55, 30, 82, and 14 stations qualifying for the audit under categories A1, A2, C1, and C2, respectively. Of those, there were 40, 22, 79, and 10 stations that were deemed fraudulent in each specified categories. This yields the 83% (151 fraudulent stations out of 181 audited stations)

NLP techniques to categorize institutions into distinct groups: primary, secondary, high school, and university. For schools that do not fit these classifications, we assign them to the "other schools" category.

3.3 Polling Center Characteristics

Before the 2009 presidential election, the Independent Elections Commission (IEC) conducted a thorough examination of every potential voting location. The inspection's goal was to evaluate the accessibility and security conditions of the selected polling places in preparation for the September 2010 legislative election. Teams from the Afghan National Security Force and ISAF worked together to complete the evaluations. Four items of information were included in each assessment: the name and identification code of the polling station, an MGRS grid that showed the precise position of the polling station, and the state of the road's accessibility.

We generate a sample that includes the fraud measures for each center as well as the locations of the centers. Next, we combine the polling-center-level data on fraud outcomes mentioned at the beginning of this section with the 2010 center evaluation data mentioned above. The name and code of the polling center are used to integrate the data. In the 100 instances where the codes and names did not line up, the match was made solely on the basis of the names. Of the remaining 6,160 polling center observations, 5,904 (95.8%) used coordinates directly from the 2010 evaluation. The following coordinates were imputed for the remaining 256 centers: Six (0.1%) centers simply used the coordinates of the district capital where the center was located, 169 (2.7%) used the centroid coordinates of the village or settlement where the center was located, and 81 (1.3%) used the coordinates of the center with the closest identifier code. We focus on Kabul province in our study because polling center classification was more accurate there due to higher-quality reporting of center names, and it had fewer ambiguous center names than other large provinces. This leaves us with 507 observations in the final sample. Figure 2 presents a map of these polling centers distributed across Kabul province, classified according to polling center type.

Additional election-related outcomes, such as the number of anticipated voters before election day, the total number of votes cast at the center, the total number of stations per center, the voter turnout rate, and the percentages won by the two front-runners, are obtained using the electoral results that have been made public. The IEC's pre-election data on the kind of voting place (school, mosque, or other) and the proportion of stations reserved for women and Kuchis, a minority ethnic group, are added to these figures. Resources from geographic information systems are used to extract the physical and economic development features of the region in which each center is situated. Using vector files gathered by the Afghanistan Information Management Service (AIMS) and acquired from the Empirical Studies of Conflict Project (AIMS 1997–2005), we specifically compute the distances between polling places and primary and secondary roads, district hospitals, basic health centers, and primary and secondary rivers. NASA's Shuttle Radar Topography Mission (SRTM30) provides data on exogenous geographic features such as polling center elevation and slope (National Aeronautics and Space Administration and the National Geospatial Intelligence Agency 2000). Finally, the US Agency for International Development (USAID)-sponsored Measuring Impacts of Stabilization Initiatives (MISTI) project provides demographic information on the population and ethnic composition in the vicinity of the polling center. Between 2012 and 2013, the MISTI project (MISTIs 2013) collected demographic information and geographic coordinates for over 37,000 villages in Afghanistan from a variety of data sources. In the village nearest to the polling center, we use these data to generate variables representing the population size and the languages spoken there ("Pashto", "Dari", and "Other").

4 Empirical Strategy

We are interested in examining the relationship between polling center location and election fraud. We employ multiple empirical strategies, using both linear and nonlinear specifications to ensure robustness of our findings and to explore different potential aspects of fraudulent voting behavior. We use two measures of fraud in our study based on ECC categorization and the number of votes and stations in a center. The first is (i) the share of suspected fraudulent votes in a polling center, and the second is (ii) an indicator of whether there was at least one suspected fraudulent station in a polling center. See section 3 for details on variable construction. We begin with a series of ordinary least squares (OLS) specifications that model the relationship between polling place location and the share of votes under Category A fraud. The following equation summarizes our baseline econometric strategy:

$$\text{Fraud}_{id} = \alpha + \mathbf{C}'_{id}\beta + \mathbf{X}'_{id}\gamma + \Omega_d + \epsilon_i \tag{1}$$

where Fraud_{id} is one of the measures of fraud mentioned above for polling center i in district d. $\mathbf{C}_{id}$ is a vector of categorical variables denoting the location of the polling center: mosque, school, house, or other. We leave mosques as the reference category. $\mathbf{X}_{id}$ is a vector of polling center characteristics that include the number of polling stations, share of female and Kuchi polling stations, distances from the polling center to: the closest village, primary, secondary, and tertiary roads, basic health facility, district health facility, primary and secondary rivers, elevation, terrain slope, primary language in the village (Pashto, Dari, and Other), indicator for whether the center is in a district capital, and population in the closest village. Most importantly, $\mathbf{X}_{id}$ contains the key drivers of polling center selection based on our context: security and accessibility of the center. Specifically, we include controls for the number of attacks within a five-kilometer radius of the center that occurred in the year of the election and up until election day, as well as indicators for the accessibility of the center (accessible, limited, or not accessible), obtained from the security assessment performed following the election.[6] Ω_d are district fixed effects, which capture unobserved heterogeneity across districts. ϵ_i is the error term. All specifications use district-clustered standard errors.

In addition to our OLS model, we estimate Equation (1) using a Poisson model to account for the large number of polling centers without any documented fraud. In cases where we use a binary outcome—Fraud_{id} equal to 1 if any polling station in polling center i exhibits fraud—we estimate versions of Equation (1) using Linear Probability and Probit models.

Causal interpretation of our estimates requires that polling center type $\mathbf{C}_{id}$ is uncorrelated with unobserved determinants of fraud. We believe this assumption is plausible in our context for two main reasons. First, the locations of polling centers were withheld from the public until one week before the election, thereby limiting officials' or voters' ability to adjust their behavior in response to polling place type [National Democratic Institute, 2010]. Second, we have access to detailed data from ISAF security assessments identifying the main operational criteria driving polling center choice: security conditions (including defensibility) and physical accessibility. As mentioned above, we control for these factors directly by including measures of pre-election violence near soon-to-be polling centers and accessibility indicators. These contextual features and conditioning on the key drivers of polling center choice mitigate concerns about endogeneity. Nonetheless, we also implement an instrumental variables strategy, which we describe next.

4.1 Instrumental Variable Approach

Polling center assignment may be endogenous to political, logistical, or security considerations correlated with fraud. For example, corrupt officials may strategically place centers in locations where fraud is harder to detect. Given that we have clear information on assignment procedures, we believe the OLS estimates present a plausible case for causal interpretation. However, to

[6] As a robustness check, we replicate our main results using only the indicators for accessibility and security reported in the ISAF security assessments. Appendix Table A4 presents the results of the specifications that use this parsimonious set of controls, which are highly similar to our main findings in direction and magnitude, and which are all statistically significant at the 1% level.

strengthen our identification strategy and address residual concerns about selection into polling place type, we implement a novel instrumental variables (IV) approach.

Independent Election Commission (IEC) guidelines required that each center should have one station for every 600 expected voters. Once a center requires more than five stations (i.e., crossed the 3,000-expected voters threshold), we observe a sharp increase in the likelihood that it is assigned to a school. Although the underlying reasoning behind this is not formally documented, we hypothesize that it reflects practical constraints: schools are more naturally compartmentalized than mosques, making them better suited to host multiple simultaneous stations. As a result, school assignments appear to become the default for centers requiring more than five stations/rooms.

This is clearly visible in Figure 3, which plots the probability of assignment to a school versus a mosque by expected number of voters. For centers below the threshold, assignment is roughly balanced. Immediately after the cutoff, the probability of being assigned to a school rises rapidly, while the probability of being assigned to a mosque falls to near zero. Although this resembles the structure of a fuzzy RD, the change at the threshold is not a discrete jump but a steep, continuous rise. Because there is no discontinuous shift at the cutoff, we do not treat this as an RD and instead rely on a standard IV strategy.

Based on this context, we define School_i, a binary indicator equal to one if the polling center is located in a school, as our endogenous regressor of interest and restrict the analysis to polling centers assigned to either schools or mosques (87% of all centers). We instrument School_i with a binary indicator equal to one if the polling center was expected to serve more than 3,000 voters.

Specifically, our 2SLS specification is given by:

$$\text{Fraud}_{id} = \alpha + \beta \text{School}_{id} + \mathbf{X}'_{id}\gamma + \Omega_d + \varepsilon_{id} \tag{2}$$

$$\text{School}_{id} = \pi_0 + \pi_1 \mathbf{1}\{\text{Voters}_{id} > 3000\} + \mathbf{X}'_{id}\rho + \Omega_d + \nu_{id} \tag{3}$$

where $\mathbf{1}\{\text{Voters}_{id} > 3000\}$ is an indicator for whether center i had more than 3,000 expected voters. The remaining variables are defined as in equation (1). Equations (3) and (2) denote the first and second stage, respectively. We estimate the system using two-stage least squares (2SLS), with standard errors clustered at the district level. In addition to full-sample estimates, we also report specifications restricted to polling centers within 1,200 expected voters of the 3,000 threshold to improve local comparability. This translates into centers with 4 to 7 stations (number of stations per center in the sample varied between 2 to 19 stations). We assess instrument strength and inference using Kleibergen–Paap F-statistics and Anderson–Rubin p-

values.

A potential concern with our IV strategy is that polling centers expected to serve more than 3,000 voters may differ systematically from smaller centers in ways that are directly correlated with fraud. For instance, larger centers might attract more oversight, or expected voter counts could be strategically manipulated in anticipation of fraud. We view these concerns as unlikely in our context. Expected voter counts came from pre-existing population estimates, reducing the likelihood of strategic manipulation. In Section 5.3, we find no significant differences in polling center characteristics or other observables–aside from school assignment–across the 3,000-voter threshold. This supports the exclusion restriction by indicating that the threshold only affects the type of polling place used. Additionally, we estimate versions of our model that restrict the sample to polling centers within 1,200 voters of the cutoff. This limits our analysis to centers with comparable populations and operational needs (i.e., 4–7 voting stations or populations between roughly 1,800 and 4,200).

5 Results

5.1 Polling Center Location

Table 1 presents our main results based on Equation (1). Columns 1 through 4 show estimates of the effect of location type on the share of votes under Category A fraud in a polling center. Column 1 shows the results from our baseline OLS model. Column 2 introduces controls for polling center and demographic characteristics; Column 3 adds district fixed effects to the specification. Our OLS estimates suggest that the use of a school as a polling center is associated with a 7.1 to 7.4 percentage-point decrease in fraud (Columns 1 and 2).

Column 4 presents the results of the Poisson specification using the share of fraudulent votes as the outcome. The reported coefficient for the school category is -1.21 and is statistically significant at the 1% level. Consistent with the results from the OLS regression analyses, this suggests that polling centers located in schools are associated with a lower share of fraudulent votes than those located in mosques. To provide a more intuitive interpretation of the coefficient, we apply the linear transformation $(1 - exp(\beta))$. For the school category, this transformation gives us $(1 - exp(-1.21))$, equal to approximately 0.70. This indicates that a polling center residing in a school decreases the expected share of fraudulent votes by approximately 70% relative to that in mosques.

Columns 5 - 6 of Table 1 report estimated effects on the probability of Category A fraud in a polling center based on nonlinear specifications of Equation (1), described in Section 4. The estimates are comparable to those from the OLS specifications, suggesting that the probability of fraud is lower if a polling center is located in a school. Specifically, the coefficient from the linear probability specification reported in Column 5 suggests that the probability of fraud is reduced by 8.2 percentage points if a polling center is located in a school. This is consistent with the estimated effect from the Probit specification reported in Column 6. The results of the Probit model indicate a marginal effect of -0.088, meaning that a polling center being in a school decreases the probability of fraud by 8.88 percentage points.

5.2 School Type

In Table 2, we report our estimated effects on election fraud when the polling center location is disaggregated by school type. We include primary schools, secondary schools, high schools, and universities in the disaggregated set of polling center locations. We present results from the same six specifications as in Table 1, using the share of fraudulent votes in a polling center as the outcome in columns 1 through 3 and the probability of at least one station displaying Category A fraud within a center as the outcome in columns 4 through 6. Our estimates suggest that all types of schools contribute to the lower levels of fraud observed in the broad category of schools compared to mosques in Table 1. In other words, we do not see one type of school driving the estimated effects discussed in Section 5. The signs of the coefficients across equation specifications and school categories are negative, and the estimated effects for secondary schools and universities are statistically significant at the 1% level. The results of the Probit specification reported in Column 6 provide additional suggestive evidence that the probability of fraud decreases (relative to mosques) consistently across schools. Again, we take this to mean that the type of school matters less than the fact that it is a school. Table 2 reports z-scores for the Probit model. Marginal effects for the different school types indicate that if a polling center is in a school, the probability of fraud decreases by 6 to 11 percentage points. Columns 3, 4, and 6 report a loss of statistical significance, particularly for the primary school category, which has fewer observations. This is expected, as the corresponding specifications include district fixed effects that remove all cross-sectional variation. We note that these results should be interpreted with caution due to the limitations of our classification methods.[7]

[7] If the type of school is not included in the school name, the NLP procedure predicts the most likely school type.

5.3 IV Estimates

Table 3 presents the two-stage least squares (2SLS) estimates of the effect of polling center school assignment on the share of fraudulent votes. Polling centers assigned to mosques are the comparison category. Panel A reports the first-stage estimates, where the instrument is an indicator for whether the polling center was expected to serve more than 3,000 voters. Across all specifications, the instrument is strongly predictive of school assignment: crossing the 3,000-voter threshold increases the probability of a polling center being assigned to a school by 32 to 53 percentage points. All estimates are statistically significant at the 1% level. The first stage is especially strong in specifications with additional controls and district fixed effects (columns 2 and 3), with Kleibergen–Paap F-statistics well above conventional thresholds.

Panel B reports the second-stage results. The estimated effect of a polling center being a school is consistently negative and statistically significant across all specifications. In the full sample without controls (column 1), school assignment reduces the share of fraudulent votes by approximately 10 percentage points. This effect remains stable when controls are added (column 2) and when district fixed effects are included (column 3).

In order to avoid comparing centers with widely different numbers of expected voters–and thus populations, columns 4 and 5 restrict the sample to polling centers within 1,200 expected voters of the 3,000 threshold. This means we are comparing centers with four to seven polling stations only. The estimates in these bandwidth-restricted models are slightly larger in magnitude (15 to 17 percentage point reductions) and remain statistically significant at the 5% level, despite the smaller sample size. Importantly, all Anderson–Rubin p-values remain below 0.05, confirming that the results are robust to weak instrument concerns.

Finally, Table A5 examines whether observable characteristics change significantly at the 3,000-voter threshold used in our IV strategy. For each variable, we replicate our first stage using the specified variable as the outcome and including district fixed effects and clustering standard errors at the district level. Across a broad range of geographic, demographic, and security-related variables, we find no statistically significant changes at the cutoff. Population and the share of female stations are exceptions. The differences in population are expected given the link between population size and expected voters. However, note that the magnitude–difference of 80 individuals–is small relative to the control mean of 3,718. A similar argument holds for the share of female stations, which is roughly 5 percentage points higher in post-cutoff centers. Taken together, these results suggest that our instrument does not predict other observable features of polling centers, lending support to the validity of our exclusion restriction.

Together, the IV estimates reinforce the OLS results and strengthen the case for a causal interpretation: relative to mosques, polling centers assigned to schools show significantly lower levels of electoral fraud.

6 Theoretical Framework

The goal of this section is to present a theoretical model that illustrates the link between polling location choice and electoral fraud and to define channels that can affect fraud through polling center location choice. We follow Gonzalez [2021] by defining two players: the polling center official and the candidate.

The Election Official's Problem The candidate purchases fraudulent votes from an election official overseeing polling center i.[8] The price of these votes reflects two factors: the probability that fraud in the center is detected and the official is penalized, and the threat of election-related violence. We assume the candidate and the official share a common belief that fraud will be detected with probability π, where $\pi = \phi(H, T)$ is a function of the level of election-related violence H and the polling center type T = {mosque, school, other}. H refers to violence targeting voters, officials, and election-related personnel and infrastructure.[9] This is relevant in the Afghan context, where the Taliban issued warnings aimed at polling centers and voters on election day (Gall [2009], Filkins [2009]). We assume that detection probability π decreases with violence (i.e., $\frac{\partial \pi}{\partial H} < 0$) since violence can impede monitoring efforts such as sending election observers to violent polling locations. We will discuss the relationship between polling center type T and the probability of detection later.

The official faces a two-stage decision. First, he must decide whether to participate as a polling center manager. This decision depends on whether the expected gains from fraud outweigh the potential disutility from violence. Second, conditional on participating, he must decide whether to engage in fraud.

Assuming the official is an expected income maximizer, the participation constraint is given by:

$$\pi(p_f v_f - F v_f) + (1 - \pi) p_f v_f \geq H v_f + M v_f \tag{4}$$

where v_f, p_f, and F denote the number of fraudulent votes, their price, and the marginal fine,

[8] Although there are other methods for committing fraud, ballot stuffing and manipulation of total counts by officials were the most prevalent forms of fraud during the 2009 Afghan election (see Weidmann and Callen [2013]).

[9] One can think of H as the expected level of violence, i.e., a probability that a violent event will take place times the payoff of being exposed to that violence.

respectively. The official receives fraud revenue $p_f v_f$, but risks a penalty $F v_f$ if detected.[10] H captures the official's disutility from facing violence. The term M captures moral costs that the official incurs from committing fraud.

Conditional on participation, the official decides whether to engage in fraud if:

$$\pi(p_f v_f - F v_f) + (1-\pi)(p_f v_f) \geq M v_f \tag{5}$$

where we assume the payoff from not engaging in fraud is zero.[11] Since the participation constraint already satisfies the fraud constraint, then the minimum price per fraudulent vote that guarantees both participation and fraud is:

$$p_f = \pi F + H + M \tag{6}$$

At a minimum, the price of fraudulent votes has to offset the expected penalties from being detected (πF), the expected violence faced by officials operating in violent polling locations (per-vote "hazard pay" H), and the moral cost M from engaging in fraud. We note that prices in this context are broadly defined. They do not necessarily imply direct monetary per-vote transfers, but may reflect financial benefits that a polling official expects to gain from facilitating one additional fraudulent vote, or that a voter associates with casting their vote for a particular candidate.

The Candidate's Problem The candidate must decide how many legal and fraudulent votes to buy from each center. We assume that if fraud is detected, the candidate loses all fraudulent votes and only receives legal votes v_l.[12] If fraud is not detected, the candidate keeps all votes $v_l + v_f$.

We define the price of legal votes v_l as $p_l = f(H, a)$, where a is a parameter that captures affinity toward the candidate among potential voters in the area. This reflects average norms in the area that favor the candidate or tribal links between the candidate and potential voters. As in the election official's problem, H refers to election-related violence faced by potential voters. We assume that higher affinity toward a candidate requires a lower price per vote to secure

[10] This assumes that once a fraud is detected, the polling center manager is penalized. Although we have no concrete information on forms of penalty, one can imagine that F also captures alternative forms of punishment such as reputational loss or electoral consequences.

[11] This implicitly assumes that the official can only engage in fraud favoring one candidate, i.e., the outside option is not to receive fraud benefits from another candidate. This assumption simplifies the analysis, but it's also consistent with patterns observed in the data, in which fraud was typically skewed towards one candidate.

[12] Audits performed during the 2009 election dropped all votes coming from suspect stations within a polling center from the vote counts.

support ($\frac{\partial f(.)}{\partial a} < 0$) while higher violence increases the price required to get potential voters to turn out ($\frac{\partial f(.)}{\partial H} > 0$).

Assuming quasilinear preferences over legal and fraudulent votes, the candidate maximizes expected utility by solving:[13]

$$\max_{v_l, v_f} \quad \pi v_l + (1-\pi)\left[v_l + v_f^{\alpha}\right]$$
$$\text{subject to} \quad p_f v_f + p_l v_l \leq E$$

where $\alpha < 1$ and E is the candidate's endowment for that center. The resulting demand for fraudulent votes is:

$$v_f = \left[\alpha(1-\pi)\frac{p_l}{p_f}\right]^{\frac{1}{1-\alpha}} \tag{7}$$

Substituting the expressions for p_l and p_f, the equilibrium level of fraud at a given polling center is given by:

$$v_f^* = \left[\alpha(1-\pi)\frac{f(H,a)}{\pi F + H + M}\right]^{\frac{1}{1-\alpha}} \tag{8}$$

We highlight two key mechanisms underlying the relationship between polling center type and fraud at equilibrium.

Violence mechanism: Election-related violence H may vary across polling center types. This can lead to different levels of fraud in equilibrium across polling center types through three avenues: the price of legal votes p_l, the price of fraudulent votes p_f, and a direct impact on the demand for fraudulent votes.

First, if violence H is higher in polling center type $t \in T$, then the price of legal votes p_l required for potential voters to turn out in that center type will be higher. This, in turn, will lead to a higher demand for substitute fraudulent votes v_f in center type t relative to other centers, given their lower relative price. Other things equal, this will lead to an upward shift in the demand for fraudulent votes and thus higher fraud in equilibrium. This is represented in the movement from point A to B in Panel A of Figure 4.

Second, from Equation (6), higher levels of election-related violence H in polling center type t will lead to a higher price of fraudulent votes p_f since officials require a higher per-vote hazard pay H to operate there. However, since H also lowers the detection probability π–monitoring

[13] The quasilinear specification is different from the perfect substitutes utility in Callen and Long [2015]. The advantage of this setup is that it avoids corner solutions. In the data, among all fraudulent centers, 86% show a combination of legal and fraudulent votes, while only 14% are fully fraudulent.

efforts such as the deployment of election observers will be lower in more violent locations–then p_f will decrease in center type t. [International, 2010] Therefore, the overall effect of violence on the price of fraudulent votes p_f is ambiguous; while violence raises p_f through the hazard pay component, it decreases it by lowering the expected punishment πF.

Third, a higher H for polling center type t will lead to lower detection probability π. Note from the candidate's optimization problem that a lower π increases the marginal utility of fraudulent votes v_f relative to legal votes v_l. All else equal, a lower π will shift the demand for fraudulent votes up and thus lead to higher fraud in equilibrium. This is represented in the movement from point A to B in Panel A of Figure 4 that is explained by the change from π to π'.

These insights can be summarized using Equation (8). Higher violence in polling center type t will increase fraud in equilibrium by raising the relative price of legal votes $f(H, a)$, lowering the price of fraudulent votes $\pi F + H + M$ via a lower expected punishment πF, and increasing the expected utility of fraudulent votes $(1 - \pi)$. The price reduction in p_f, however, is offset by the higher hazard pay H required to operate in a more violent center type.

We also note that violence can affect equilibrium fraud through selection. If polling center type t was more likely to be selected in areas with higher H, then this selection can explain differences in fraud in equilibrium. We account for this latter explanation empirically by using district fixed effects, controlling for determinants of polling center choice documented in our security assessments data, and our instrumental variable approach.

Behavioral mechanism: Differences in fraud across polling center types may result from variation in the composition or behavior of voters and polling officials. In the context of the model, both cases operate through their effects on the perceived probability of detection, π, which enters directly into the price of fraudulent votes, $p_f = \pi F + H$, and the expected utility of vote types in the candidate's optimization problem.

On the voter side, if more civically engaged voters select into polling center type t, then the pool of voters in t will be more likely to monitor and report misconduct. This increases the perceived likelihood of detection (π). In addition, the polling center context may influence behavior. For example, voting in a school may prime a sense of civic duty or responsibility to future generations. These can make voters more proactive in reporting misconduct, thus raising π. In either case, a higher π increases the price of fraudulent votes p_f via a higher expected punishment πF and lowers the expected utility of fraudulent votes $1 - \pi$. This will reduce equilibrium fraud.

A similar logic applies on the official side. If polling officials differ systematically across center types, this can generate variation in the moral cost of fraud. For instance, more rule-abiding officials may be assigned to (or self-select into) polling center type t and have a higher per-vote disutility M. This will lead to a higher price of fraudulent votes in equilibrium in centers t. The polling center environment may also prime official behavior even in the absence of selection. If, for example, the institutional context of schools primes civic norms, this can also increase M. In both cases, a higher M results in lower incentives for officials to supply fraud and thus less fraud in equilibrium. The difference in equilibrium fraud due to selection and behavioral differences in π and M are represented by points A and B in Panel B of Figure 4. Point B represents a center where officials have lower dishonesty costs M (a drop in the price of fraudulent votes p_f) and perceive a lower probability of detection π (a downward shift in the demand curve along with a lower price of fraudulent votes p_f).

While the model does not explicitly model selection into polling center roles or locations, note from Equation (8) that any factor that increases either the perceived probability of detection π or the moral cost M in polling center type t—whether through selection or behavioral responses to context—will raise the price of fraudulent votes.

Additional Channels: In addition to the main channels discussed above, several other factors may plausibly influence fraud levels across polling center types. First, differences in the affinity parameter a may arise if polling center types are systematically assigned to areas with stronger baseline support for a given candidate. For instance, if the agency responsible for assigning polling centers colludes with the incumbent, it may allocate more strategically advantageous locations to incumbent-friendly areas. This can lead to a being systematically different between mosques and schools. In the model, a higher a reduces the price of legal votes p_l, which in turn lowers the demand for fraudulent votes. Thus, selection on affinity should reduce fraud in equilibrium and cannot account for the higher fraud observed in mosques. Second, the model includes the marginal penalty for fraud F. Because F is fixed across polling center types, it also cannot explain variations in fraud. Finally, officials may perceive that fraud committed in a mosque is less likely to attract suspicion due to the sanctity of the setting. If mosques are viewed as moral shields, this may lower the subjective probability of detection π, thereby increasing the incentive to engage in fraud. While these additional channels are not the central focus of the model, they offer complementary behavioral and institutional explanations for the observed differences.

7 Mechanisms

7.1 Election-Related Violence

Political violence can affect the incidence of electoral fraud in a number of ways. First, limited institutional capacity for election oversight in conflict settings creates opportunities for manipulation [Weidmann and Callen, 2013]. Election monitors, media agents, and other actors that might observe and report on fraudulent behavior are less likely to operate in areas with higher levels of insurgent activity. In other words, election oversight is most constrained where violence is the most salient. Second, insurgent violence that occurs on and around election day reduces voter turnout, which provides space for corrupt officials to manipulate vote tallies [Condra et al., 2018]. In some contexts, election-related violence strategically targets voters, polling officials, or polling locations in order to help or harm specific candidates. This is not the case in Afghanistan, as the Taliban aimed to disrupt the political process in general, rather than to aid one candidate or another [Weidmann and Callen, 2013, Greer, 2009]. Based on these channels and the fact that election-related violence did not favor either candidate, we expect more fraud to occur in the polling center locations that experience higher levels of violent attacks.

One week before the election, the IEC officially announced the location of polling centers across Afghanistan. This disclosure introduced a potential strategic element to insurgent attacks, as rebel groups could now precisely target locations where voters would gather. We examine the role of violence during the 2009 election by analyzing how violence near polling centers changes following the announcement and up to election day.

We use SIGACTs data to calculate the number of violent incidents occurring every day within a two-week window around the announcement day (one week before announcement to one week after announcement, i.e., election day) within a 5-kilometer radius around a polling center. We do this to assess whether polling centers were more likely to be attacked, given that the locations are known.

We estimate the following equation:

$$V_{idt} = \beta_0 + \sum_{\substack{j=-8 \\ j\neq -1}}^{7} \beta_j \mathbf{1}(t=j) + \mathbf{X}_i\theta + \gamma_d + \varepsilon_{idt} \tag{9}$$

where V_{idt} is the number of attacks within 5 km of polling center i on day t, $\mathbf{1}(t=j)$ is an indicator variable that equals 1 if it has been j days since the announcement of polling center locations and 0 otherwise. We leave $j=-1$ (one day before the announcement) as the reference category. $\mathbf{X}_i$ is a vector of controls at the polling center level.[14] γ_d is a district fixed effect. ε_{it} is the error term, clustered at the district level.[15]

Figure 5 shows that attacks go up right after the announcement, suggesting that armed groups responded strategically to polling center disclosures. This suggests that violence was likely a key determinant of suppressed turnout and, therefore, a mechanism for influencing fraud.

While the increase in violence around polling centers suggests a link between insurgent strategy and electoral manipulation, in order for violence to lead to differential levels of fraud across different polling centers, the effect of violence must differ across polling center types.

To assess whether exposure to violence varied by polling center category, we estimate the regression below that allows the effect of violence to vary by whether the polling center is a school (or other type) relative to a mosque.

$$\mathrm{V}_{idt} = \beta_0 + \sum_{\substack{j=-8 \\ j\neq -1}}^{7} \beta_j \mathbf{1}(t=j)\cdot \mathrm{School}_i + \sum_{\substack{j=-8 \\ j\neq -1}}^{7} \delta_j \mathbf{1}(t=j)\cdot \mathrm{Other}_i + \mathbf{X}_i\theta + \gamma_d + \lambda_t + \varepsilon_{idt} \qquad (10)$$

where all terms are defined as in Equation (9) and λ_t represents time fixed effects. $\mathbf{1}(t = j)\cdot \mathrm{School}_i$ and $\mathbf{1}(t = j)\cdot \mathrm{Other}_i$ capture the interactions of time indicators with indicators for whether polling center i is a school (School_i) or other type (Other_i). mosques are left as the comparison category. We are interested in coefficient β_j, which gives the difference in the daily number of attacks between schools and mosques in the window around the announcement of the polling locations and election day. [16]

Figure 6 plots the estimates of β_j for all attacks and for IED-specific attacks. We find that following the announcement of polling center locations, schools are less likely to be targeted by the Taliban relative to mosques and other polling center types. Because attacks disproportionately affected mosques, these locations experienced lower voter turnout, creating greater opportunity for officials to substitute lost votes with fraudulent ballots. The sharp increase in attacks following polling center announcements provides empirical support for this violence-fraud compensation mechanism, demonstrating how targeted insurgent violence strategically shaped electoral outcomes. Additionally, more violence in and around mosques implies a lower

[14] Included controls are distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, slope, indicators for language (Pashto and Other), population, and indicators for access (Limited and Other).

[15] Refer to Table A6 in the Appendix where we present the results for Equation (9).

[16] Refer to Table A7 in the Appendix where we present the results for Equation (10).

likelihood of election monitors being able to operate in these polling centers.

There are several reasons why mosques would be targeted by insurgent forces more than schools. First, insurgents were known to target individuals or groups appearing to be associated with or supportive of the government. This included attacks on pro-government religious leaders in the years leading up to the election [OHCHR, 2009, Bureau of Democracy and Labor, 2008][17] Second, there would likely have been more people present at mosques on election day, as schools would most likely close for election days, while mosques would have to remain open for prayer. This means that there could be many people in and out of a mosque throughout the day, even if they are not all there to vote. Because of this, attacking a mosque could have been perceived to have a larger impact.

7.2 Mediation Analysis

Section 7.1 highlights how schools may be perceived as less attractive targets for insurgent violence, thereby reducing opportunities for electoral manipulation. This motivates an investigation into whether violence acts as a mediator for the effect of polling center types on fraud outcomes. To formally assess this channel, we conduct a causal mediation analysis within an instrumental variable framework following Dippel et al. [2020]. Following the IV approach outlined in section 4.1, we instrument for the school indicator using a binary variable equal to one if the number of expected voters at the polling center exceeds 3000. The validity of this IV-based mediation analysis relies on the key identifying assumption outlined in Dippel et al. [2020], which is that any unobserved factors that influence the treatment cannot directly influence the outcome except through their effect on the mediator. Put more simply, the only way unobserved factors can link the treatment to the outcome is by first affecting the mediator. Confounders jointly affecting the mediator and outcome are permissible. The plausibility of this assumption in our context follows from the IEC's operational rule for assigning polling centers to schools when the expected number of voters exceeds 3000. This rule, driven by logistical rather than political or security concerns, makes it unlikely that it is correlated with unobserved factors that directly affect fraud. While such unobservables may plausibly impact both school assignment and post-announcement violence, there is no clear mechanism by which they could affect fraud independent of violence. This supports the assumption following Dippel et al. [2020] that endogeneity in treatment operates through the mediator, rather than directly through the outcome.

[17] The U.S. Bureau of Democracy, Human Rights, and Labor also reported targeted attacks against progovernment religious leaders in the year of the election "for supporting the government or for stating that activities conducted by terrorist organizations were against the tenets of Islam" [Bureau of Democracy and Labor, 2010].

We implement the analysis through the following system of equations:

$$\text{School}_{id} = \pi_0 + \pi_1 \mathbf{1}\{\text{Voters}_{id} > 3000\} + X'_{id}\rho + \Omega_d + \eta_{id} \tag{11}$$

$$\text{Violence}_{id} = \delta_0 + \delta_1 \widehat{\text{School}}_{id} + X'_{id}\theta + \Omega_d + \nu_{id} \tag{12}$$

$$\text{Violence}_{id} = \gamma_0 + \gamma_1 \mathbf{1}\{\text{Voters}_{id} > 3000\} + \gamma_2 \text{School}_{id} + X'_{id}\delta + \Omega_d + \xi_{id} \tag{13}$$

$$\text{Fraud}_{id} = \beta_0 + \beta_1 \widehat{\text{Violence}}_{id} + \beta_2 \text{School}_{id} + X'_{id}\gamma + \varepsilon_{id} \tag{14}$$

Equation (11) represents the standard first stage linking the instrument, $\text{Voters}_{id} > 3000$, to the treatment, School_{id}, where all terms are as defined in Equation (3). Equation (12) estimates the effect of the predicted school assignment, $\widehat{\text{School}}_{id}$, on the mediator, Violence_{id}, defined as the number of violent attacks within a 5 km radius of the polling center after the public announcement of polling center locations. This captures the first link in the mediation chain, isolating the causal effect of polling center type on violence. Equation (13) constitutes the mediator first stage. Here, the threshold is used as an instrument for Violence_{id}, conditional on School_{id}, exploiting the identifying assumption of Dippel et al. [2020] that the instrument is valid for the mediator when conditioned on the treatment. This step isolates exogenous variation in violence that is uncorrelated with unobserved determinants of fraud except through school assignment. Finally, Equation (14) estimates the effect on the outcome, Fraud_{id}, by regressing on the fitted values from the mediator first stage, $\widehat{\text{Violence}}_{id}$ and School_{id}.

In the framework provided in Dippel et al. [2020], the *total effect* of the treatment on the outcome can be decomposed into two components. A *direct effect* and an *indirect effect* operating through the mediator. Formally, the *total effect* captures the combined influence of the treatment on outcome, both through the mediator and through any remaining channels. The *direct effect* measures the impact of the treatment on the outcome *holding the mediator constant*, thereby capturing all paths that affect the outcome without operating through the mediator. The *indirect effect* measures the part of the treatment that affects the outcome *only through* the mediator. In our context, substituting (12) into (14) gives us the *total effect* $(\beta_2 + \beta_1\delta_1)$ which reflects how school assignment impacts fraud directly (β_2), through physical layout, feasibility of monitoring, etc., and indirectly $(\beta_1\delta_1)$, through its effect on the level of violence, which in turn affects the opportunity for fraud. This decomposition is valuable because it allows us the differentiate the violence channel from other mechanisms linking polling center type to fraud, thereby providing a more nuanced understanding of how operational decisions on polling center assignment may influence the integrity of the election process.

Table 4 reports the IV mediation estimates. The *total effect* of assigning a school as the polling center simply refers to the IV estimate reported in Column 3 of Table 3. Note that it is large, negative, and statistically significant, indicating that school polling centers experience lower rates of fraud relative to mosques.

In columns 1 and 2, we use the number of violent events within 5 km of the polling center occurring *after* the official announcement of polling center locations as the mediator. The *direct effect*—the portion of the total effect not operating through the violence channel—is smaller in magnitude than the corresponding *indirect effect*. The indirect effects are sizable and statistically significant: -0.065 in Column 1 (all attacks) and -0.059 in Column 2 (IEDs), implying that 58-64% of the total effect is mediated through reduced violence in schools relative to mosques. These findings are consistent with the mechanism proposed in Section 7.1. First-stage statistics indicate a strong instrument for the treatment (F-statistics above 35), and the mediator first stage is also strong for post-announcement violence (F-statistics of 9–10).

Columns 3 and 4 implement a placebo-style test in which the mediator is violence in the same geographic radius, but *before* polling center locations were announced. The logic of this exercise is that, prior to the announcement, neither voters nor insurgents knew which specific locations would be used as polling centers. Therefore, pre-announcement violence should not plausibly mediate the relationship between polling center type and fraud. Consistent with this prediction, the indirect effects in columns 3 and 4 are small in magnitude (-0.048 and -0.026) and statistically insignificant. The mediator first-stage F-statistics are also low (< 1).

Taken together, the results indicate that a substantial share of the fraud-reducing effect of school polling centers operates through lower exposure to strategically targeted violence after their locations become known.

7.3 Voter and Polling Official Behavior

Differences in fraud between different types of polling centers can be explained if there are intrinsic differences in either who selects to manage a polling center or who selects into voting in a given polling center. For example, if more civic-minded or honest individuals choose to vote or manage polling centers in schools, then these locations may have more oversight from voters and unofficial observers, or the polling center manager may feel more reticent about committing fraud. In our model, this means that the perceived probability of detection, π, and therefore the price of fraudulent votes, are higher in schools than in mosques.

We do not have data on voters or polling officials; therefore, we cannot directly test this channel. However, we can test whether there are differences in behavior related to election complaints. This can provide insight into the differences in the composition of individual voters and officials in mosques and schools. Specifically, we use polling center-level data on all election-related complaints submitted to the ECC during the 2010 parliamentary election. These data include the type of complainant (e.g., individual, candidate, organization), the gender of the complainant, and the type of suspect (e.g., candidate, polling official).[18] We note that the election complaints data come from the 2010 parliamentary election and not the election studied in this paper. However, the two elections used the same polling centers, the same overseeing bodies, and were to be held on the same date to reduce costs, but disagreements led to the different dates [Faiez, 2008]. Despite this caveat, we believe that if polling center type drives the selection of officials, voters, or community engagement in one election, it will do so in similar elections. Therefore, our insights remain relevant to our results even if they are from a different but closely related election.

Table 5 presents results by estimating the number of complaints by polling center type using a Poisson specification. The outcome variables are the total number of complaints (Column 1), the number of complaints filed by women (Column 2), the number of complaints filed by individuals (Column 3), where the polling official is the respondent (Column 4), and where the candidate is the respondent (Column 5). Panel B of Table 5 replicates these results using the share of complaints of the specified type (e.g., number of complaints filed by women divided by total complaints in that center).

Schools report a significantly higher number of complaints than mosques (Column 1). The number of complaints raised by women is higher in schools (Panel A, Column 2), although the share is not different across schools and mosques (Panel B, Column 2). The number and share of total complaints filed by individuals, rather than candidates or organizations, is significantly higher in schools than in mosques (Column 3). The share of complaints where the polling official is accused is lower in schools (Panel B, Column 4), while the number and share of complaints where the candidate is accused are higher in schools (Column 5). In all, this provides suggestive evidence that voters and officials may be intrinsically different (or behave differently) in schools relative to mosques. Individuals are more engaged in schools, as evidenced by their increased likelihood of raising complaints. Additionally, women and individual citizens are more likely to participate in the process. Similarly, there is potential evidence that officials might be different in schools: they are less likely to be accused in complaints than their mosque counterparts. Instead, candidates are more likely to be accused in schools.

[18] Refer to Weidmann and Callen [2013] and Gonzalez [2021] for more information on these data.

7.4 Selection

7.4.1 Pre-Election Violence

The IEC originally planned to open 26,877 polling stations across 6,970 polling centers on Election Day. Security agencies evaluated the accessibility and security of potential polling center locations between March and August of 2009 to determine which were best suited to service different areas.[19] Many intended polling centers were eliminated due to security incidents during this time period, and only 6,210 polling centers and 24,183 polling stations were included on the final list. [IEC Afghanistan, 2009][20]

We explore whether violence during the assessment period influenced the selection of polling center locations. To do so, we again use SIGACTs data to calculate aggregate violent incidents in the catchment area of a polling center during the period when we believe the assessments were taking place, from March 01, 2009, to August 13, 2009.[21] We do this to assess whether certain types of polling centers were more likely to be assigned based on the level of violence in the months leading up to the election.

We estimate the following equation:

$$School_{icd} = \beta_0 + \beta_1 V_{cd} + \mathbf{X}_i \theta + \gamma_d + \varepsilon_{icd} \tag{15}$$

where School_{icd} is a binary indicator equal to one if polling center i in catchment area c is located in a school. V_{cd} is the number of attacks within catchment area c between March 01 and August 13. $\mathbf{X}_i$ is a vector of controls at the polling center level.[22] γ_d is a district fixed effect. ε_{icd} is the error term, clustered at the district level. We use linear probability and probit models of this equation to estimate the probability that polling center i is located in a school, given the level of violence in catchment area c.

We find no evidence of a correlation between aggregate violence levels in the area around a polling center during the assessment period and the probability that a polling center is located in a school. Table 6 presents the results of this analysis. The observed marginal effects are neg-

[19] Although we do not have explicit records documenting the dates of the assessment period, we estimate that they took place between March 01, 2009, and August 13, 2009, based on information reported by Van Bijlert [2020].

[20] International [2010] provides similar figures, reporting that only 6,289 of the intended 6,969 polling centers were used on election day due to security issues.

[21] The catchment areas are constructed using Voronoi polygons, which divide an area into regions where all locations closest to a particular sample point are enclosed within a single polygon [Pearce, 2000]. Figure A1 shows a map of catchment areas for each polling center in the Kabul province.

[22] Included controls are distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, slope, indicators for language (Pashto and Other), population, and indicators for access (Limited and Other).

ligible, approaching zero, and none of the results are statistically significant at any conventional level. These null findings indicate that polling centers were not assigned to different types of locations based on the overall level of violence in the area.

7.4.2 Candidate-Voter Affinity

Loyalty networks are another channel through which polling center selection can be affected. As discussed in Section 6, the price of legal votes v_l is a function of election-related violence and voter affinity toward the candidate. High candidate-voter affinity in an area lowers the price of legal votes, thereby decreasing fraud as the relative price of fraudulent votes increases.[23] Candidate-voter affinity can also increase the potential for fraud; however, if candidates and election officials use it to better coordinate fraudulent activity. Polling center selection is a channel through which such coordination might occur, although the potential effects on location assignment are ambiguous. Imagine an election official who assigns polling center locations has strong ties to one candidate. This official knows that the Taliban is more likely to target mosques, which will decrease voter turnout and increase the price of legal votes. Given this knowledge, the election official assigns polling centers to mosques in areas where voter affinity for the official's candidate is low, to increase the price of legal votes for the opposing candidate. Additionally, if the polling center manager assigned to the location has the same candidate affinity, the probability of being detected committing fraud π decreases.

Alternatively, perhaps the official assigns polling centers to schools in areas where expected turnout is high, to maximize potential legal votes. Another possibility is that the election official knows that some intrinsic feature of schools makes people more likely to report fraud. Given this expectation, the official assigns more polling center managers with the same candidate affinity to mosques so that they may engage in fraud with less oversight. Similarly, the official may assign more mosques in an area with both high candidate-voter affinity and higher levels of violence. In doing so, when election-related violence decreases turnout and leads to increased substitution of fraudulent votes, the probability of detection decreases both because violence limits oversight and because voters with strong candidate affinity may be less likely to notice or report suspiciously favorable results.

Although we are unable to test tribal links and candidate-voter affinity directly, we use language as a proxy for ethnic ties. The incumbent, Hamid Karzai, has strong ties to the Pashto-speaking Pashtun ethnic group. The voting block of the opposition candidate, Abdullah

[23] See Gonzalez [2021] and Callen and Long [2015] for additional information on the link between legal votes, fraud, and candidate-voter affinity via tribal and ethnic links.

Abdullah, is primarily comprised of Tajik- and Dari-speaking ethnic groups [Qazi, 2020]. We explore whether the dominant language, and by extension, ethnic group, in an area influenced the selection of polling center locations. To do so, we use data from the MISTI project, described in more detail in Section 3, to identify the languages spoken in the village nearest to each polling center and in the catchment area of each village. We follow the methods described in the previous section to assess whether certain types of polling centers were more likely to be assigned based on the most common language spoken in the area.

We find no evidence of correlation between the language spoken in the area around a polling center and the probability of a polling center being a school, according to the results shown in Table 7. Panel A shows the marginal effects on the probability of a polling center being assigned to a school when Pashto is the dominant language spoken in the nearest village, while Panel B presents the same when Pashto is the most common language spoken in the village(s) of the catchment area of a polling center. In both cases, only the Probit specification in Column 4 is significant at the 10% level but does not remain significant when we add controls in Column 5. Overall, the results provide no consistent evidence that ethnic composition, proxied by dominant language spoken, systematically influenced polling center assignment.

7.5 Other Channels

In addition to the primary mechanisms identified in this study, there are several alternative, albeit untestable channels, that may help explain the observed differences in electoral fraud across polling locations. These mechanisms are theoretically possible and consistent with the findings of the paper, but cannot be tested directly due to limitations of the available data.

One potential channel relates to the perception of safety in religious institutions. It is possible that mosques, being sacred spaces, are perceived as sanctuaries where fraud is less likely to be suspected or detected. This perception of safety could result in polling officials feeling less scrutinized and thereby more likely to engage in corrupt behavior. Although this hypothesis is compelling, it remains speculative, as the available data is unable to quantify the extent to which voters or officials perceive these religious spaces as immune to scrutiny.

Another mechanism could stem from intrinsic differences between schools and mosques that influence the selection and behavior of both voters and polling officials. While this hypothesis suggests that the type of institution may attract different individuals, it is difficult to test directly due to the lack of individual-level data. The election complaints data shows that individuals in schools seem to exhibit different behaviors compared to those in mosques. This may reflect a different set of motivations among individuals at these locations. However, without data on the personal traits of those involved, this channel is not directly testable.

Another channel includes the strategic targeting of mosques by insurgent groups. Given the significant role that mosques play in religious importance as well as the voting process, they are often viewed as high-value targets by the insurgent groups seeking to disrupt the election process. If insurgents targeted mosques in order to suppress voter turnout, this could have created an environment where polling officials felt the need to substitute fraudulent votes to compensate for the lost turnout. The evidence from this study suggests that mosques experienced a higher frequency of violence compared to schools, which aligns with the idea that violence may have disproportionately affected voter participation in mosque-based polling centers. However, the precise mechanism through which violence led to increased fraud is difficult to isolate.

Lastly, the size of polling centers may play a role in the likelihood of electoral fraud. Larger polling centers, which are more common in schools due to their greater physical infrastructure, may be subject to more intense monitoring and scrutiny. This increased oversight could, in turn, reduce the opportunities for fraud. The complaints suggest that schools, which often have the capacity to handle larger expected voter turnouts, may experience higher levels of scrutiny, thereby limiting the chances for fraudulent activity. However, this hypothesis also remains speculative due to the lack of available data on the intensity of monitoring across different polling locations.

8 Policy Implications

The findings of this study have important implications for election administration and monitoring in unstable or conflict-affected settings. According to the 2025 *Freedom in the World* report, violence affected 27 of the 66 countries that held national elections in 2024 [Gorokhovskaia and Grothe, 2025], making our findings highly relevant in this context. Our results suggest that the location of polling centers matters and can be a powerful tool for reducing electoral fraud.

The differential fraud patterns between mosques and schools highlight that polling location choice represents a readily available and cost-effective policy intervention. Unlike other approaches to reducing electoral fraud, such as increasing security measures or reducing violence, changing polling center locations requires minimal resources and can be implemented relatively quickly. Election administrators should prioritize polling center locations that are less susceptible to insurgent targeting and manipulation when designing the polling process. This idea extends beyond the specific context of schools versus mosques in Afghanistan. The key insight is that in conflict settings, election officials should strategically evaluate potential polling locations based on their vulnerability to violence, voter suppression, and fraud. Places of worship, which are commonly used as polling centers across many countries, may be particularly susceptible to targeted violence and subsequent fraud compensation mechanisms. However, there are likely to be settings where the opposite is true. Election administrators should assess the security profile and symbolic significance of potential polling locations within their specific context.

Additionally, our findings provide guidance for election monitoring organizations operating under limited capacity. Rather than distributing observers uniformly across polling centers, monitoring agencies may consider concentrating resources on high-risk locations identified through violence and location-type risk assessments. In the context of Afghanistan, this would mean prioritizing mosque-based polling centers for intensive monitoring. In this context, though, election monitors may not be able to operate in high-violence areas because the risk is too high. Additional tools, such as monitoring technology, may be needed for high-risk polling locations. This targeted approach allows monitoring organizations to maximize their impact with limited resources. By focusing on locations where fraud is most likely to occur, monitors can more effectively detect and deter electoral manipulation. The strategic allocation of monitoring resources is particularly valuable for international election observation operations in challenging security environments.

The violence-fraud compensation mechanism identified in this study likely operates in other conflict-affected elections where insurgent groups have both the motive and capacity to disrupt voting. Election administrators and international organizations should consider these dynamics when designing electoral systems in post-conflict or active conflict environments. The framework developed here, which examines how polling location characteristics interact with targeted violence to create fraud opportunities, can be adapted to analyze electoral vulnerability in diverse conflict settings worldwide.

Our findings also hold in non-conflict settings. Violence acts as a shock that affects the prices of legal and fraudulent votes in our model. Other types of shocks can have the same effect, and there are other contexts where violence is not prevalent, but where other shocks, such as public health crises or severe weather events, can differentially affect the price of legal and/or fraudulent votes at different polling center locations.[24]

[24] Consider, for example, a public health event such as COVID-19, where voter turn out may be lower at venues that are typically smaller and more confined, such as at schools made up of small classroom buildings (or, conversely, a context where places of worship have a higher health risk than a school gymnasium). Regarding severe weather events, consider a context where a certain type of polling location is more likely to be susceptible to the elements and is less accessible on an election day after a storm or a long period of rain.

9 Discussion and Conclusion

Election fraud, which affects many developing countries, undermines critical functions of democracy. Given its prevalence, it is important to understand the contextual factors and mechanics of election fraud to inform policy decisions and future election design. In this paper, we explore the relationship between polling center location and fraudulent votes in Afghanistan's 2009 presidential election. We build on the existing electoral fraud literature by focusing on the type of polling site location, i.e., if it is a school, mosque, house, or other category, and by exploring the differential effects of known causal pathways for fraud, such as violence, across this categorical dimension. Our results suggest that polling center location has a significant relationship with electoral fraud. Specifically, we find that polling centers located in schools are substantially less likely to display serious instances of fraud than polling centers located in mosques. We are able to identify two primary mechanisms for this difference. First, we see insurgent violence increase more in mosques after the locations of the polling centers were announced than in schools. This, combined with the findings of our main results, suggests that a disproportionate amount of violence in mosques contributed to disproportionate levels of fraud in mosques. Second, the number of complaints filed is higher in schools than in mosques. This could indicate a psychosocial effect in which voters and election officials experience "contextual priming" that leads them to behave differently in different locations. Overall, our results point to the importance of contextual factors, such as polling center location, when considering electoral fraud.

References

Nicolás Ajzenman and Ruben Durante. Salience and accountability: School infrastructure and last-minute electoral punishment. *The Economic Journal*, 133(649):460–476, 2023.

World Bank. *World Development Report: Governance and the Law.* World Bank, Washington, DC, 2017. URL https://www.worldbank.org/en/publication/wdr2017.

Jonah Berger, Marc Meredith, and S Christian Wheeler. Contextual priming: Where people vote affects how they vote. *Proceedings of the National Academy of Sciences*, 105(26):8846–8849, 2008.

Sarah Birch and Carolien Van Ham. Getting away with foul play? the importance of formal and informal oversight institutions for electoral integrity. *European Journal of Political Research*, 56(3):487–511, 2017.

Human Rights Bureau of Democracy and Labor. Afghanistan: International religious freedom report 2008, Sep 2008. URL https://2001-2009.state.gov/g/drl/rls/irf/2008/108497.htm.

Human Rights Bureau of Democracy and Labor. 2009 country reports on human rights practices: Afghanistan, Mar 2010. URL https://2009-2017.state.gov/j/drl/rls/hrrpt/2009/sca/136084.htm.

M. Callen and James Long. Institutional corruption and election fraud: Evidence from a field experiment in afghanistan†. *The American Economic Review*, 105:354–381, 2015. doi: 10.1257/AER.20120427.

Luke N Condra, James D Long, Andrew C Shaver, and Austin L Wright. The logic of insurgent electoral violence. *American Economic Review*, 108(11):3199–3231, 2018.

Christian Dippel, Andreas Ferrara, and Stephan Heblich. Causal mediation analysis in instrumental-variables regressions. *The Stata Journal*, 20(3):613–626, 2020.

Electoral Complaints Commission. Final report: 2009 presidential and provincial council elections. Technical report, Electoral Complaints Commission, Kabul, 2010. URL https://www.dropbox.com/s/zq4eo4tlsp6uy0y/ECC%20Final%20Report%202009.pdf?dl=0. Accessed June 13, 2025.

Rahim Faiez. Afghanistan to hold separate presidential, parliamentary elections. *Associated Press*, 2008. Accessed: 2025-02-03.

Dexter Filkins. Threats by taliban may sway vote in afghanistan. *The New York Times*, 2009.

Carlotta Gall. Violence roils afghanistan days before election. *The New York Times*, 2009.

Germán Gieczewski and Mehdi Shadmehr. An institutional design approach to preventing election fraud. *Available at SSRN 4722709*, 2024.

Robert M Gonzalez. Cell phone access and election fraud: evidence from a spatial regression discontinuity design in afghanistan. *American Economic Journal: Applied Economics*, 13(2): 1–51, 2021.

Yana Gorokhovskaia and Cathryn Grothe. Freedom in the world 2025: The uphill battle to safeguard rights. Technical report, Freedom House, February 2025.

Bill Greer. Dataset of the day: Impact of violence on the afghanistan elections, August 2009. URL https://blog.geoiq.com/2009/08/26/dataset-of-the-day-impact-of-violence-on-the-afghanistan-elections/.

Democracy International. Democracy international report on 2009 presidential elections, Aug 2010.

Toby S. James and Alistair Clark. Electoral integrity, voter fraud and voter id in polling stations: lessons from english local elections. *Policy Studies*, 41:190 – 209, 2019. doi: 10.1080/01442872.2019.1694656.

Sandra Khadhouri. A review of suspected electoral fraud: 2009 afghan presidential and provincial council elections. *Democracy International*, 2010. URL https://www.dropbox.com/s/a9qevpqv2yxr932/ElectionsReportENGFINAL2.pdf?dl=0. Accessed June 13, 2025.

National Democratic Institute. The 2009 presidential and provincial council elections in afghanistan, 2010. This report is no longer available online.

Independent Election Commission of Afghanistan. Afghanistan: Press release of the independent election commission with reference to announcement of preliminary results of 2009 presidential elections - afghanistan, Sep 2009. URL https://reliefweb.int/report/afghanistan/afghanistan-press-release-independent-election-commission-reference-announcement.

OHCHR. Institute on religion and public policy report: Religious freedom in afghanistan, 2009.

Jamie Pearce. Techniques for defining school catchment areas for comparison with census data. *Computers, Environment and Urban Systems*, 24(4):283–303, 2000.

Shereena Qazi. Who is afghanistan's abdullah abdullah?, Mar 2020. URL https://www.aljazeera.com/news/2020/3/9/who-is-afghanistans-abdullah-abdullah.

Daniel Stockemer, B. Lamontagne, and L. Scruggs. Bribes and ballots: The impact of corruption on voter turnout in democracies. *International Political Science Review*, 34:74 – 90, 2013. doi: 10.1177/0192512111419824.

Martine Van Bijlert. What to watch in the elections (2): Polling station lists, Mar 2020. URL https://www.afghanistan-analysts.org/en/reports/political-landscape/what-to-watch-in-the-elections-2-polling-station-lists/.

Nils B Weidmann and Michael Callen. Violence and election fraud: Evidence from afghanistan. *British Journal of Political Science*, 43(1):53–75, 2013.

Tables and Figures

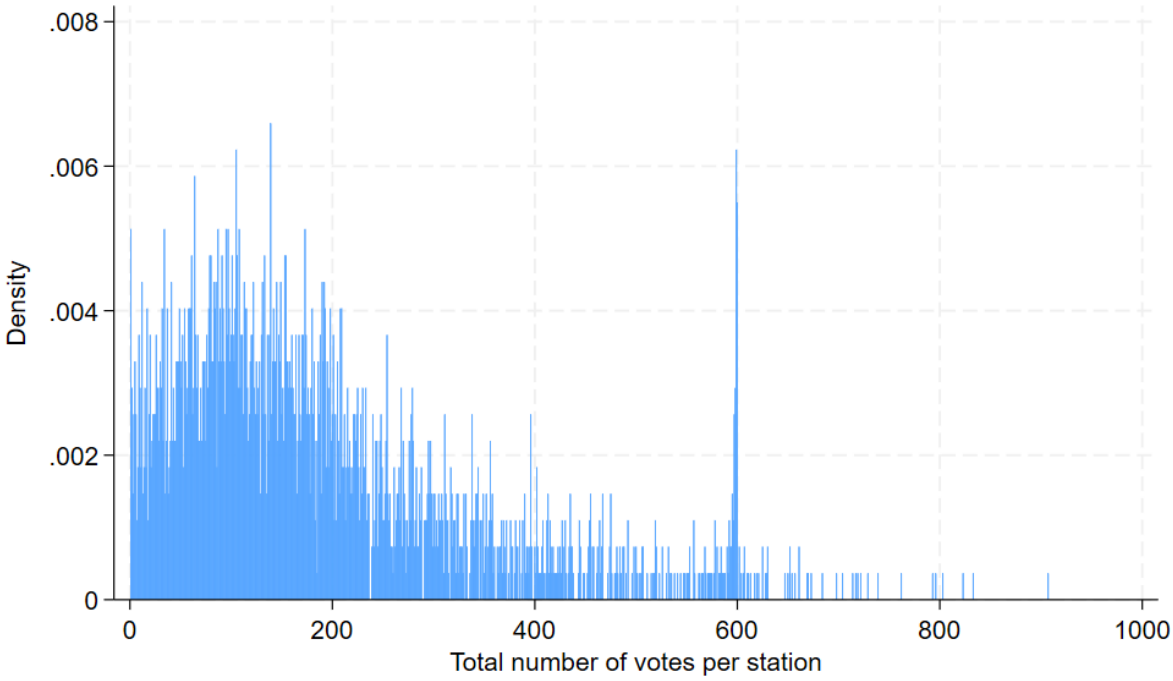


Figure 1: Total Votes per Station, Kabul Province

Notes: Total votes reported per station in Kabul province. There are a total of 2,786 polling stations in Kabul province. Sample excludes stations with zero reported votes (36 stations). Bar width is set to 1 to show the density of stations for each total vote count.

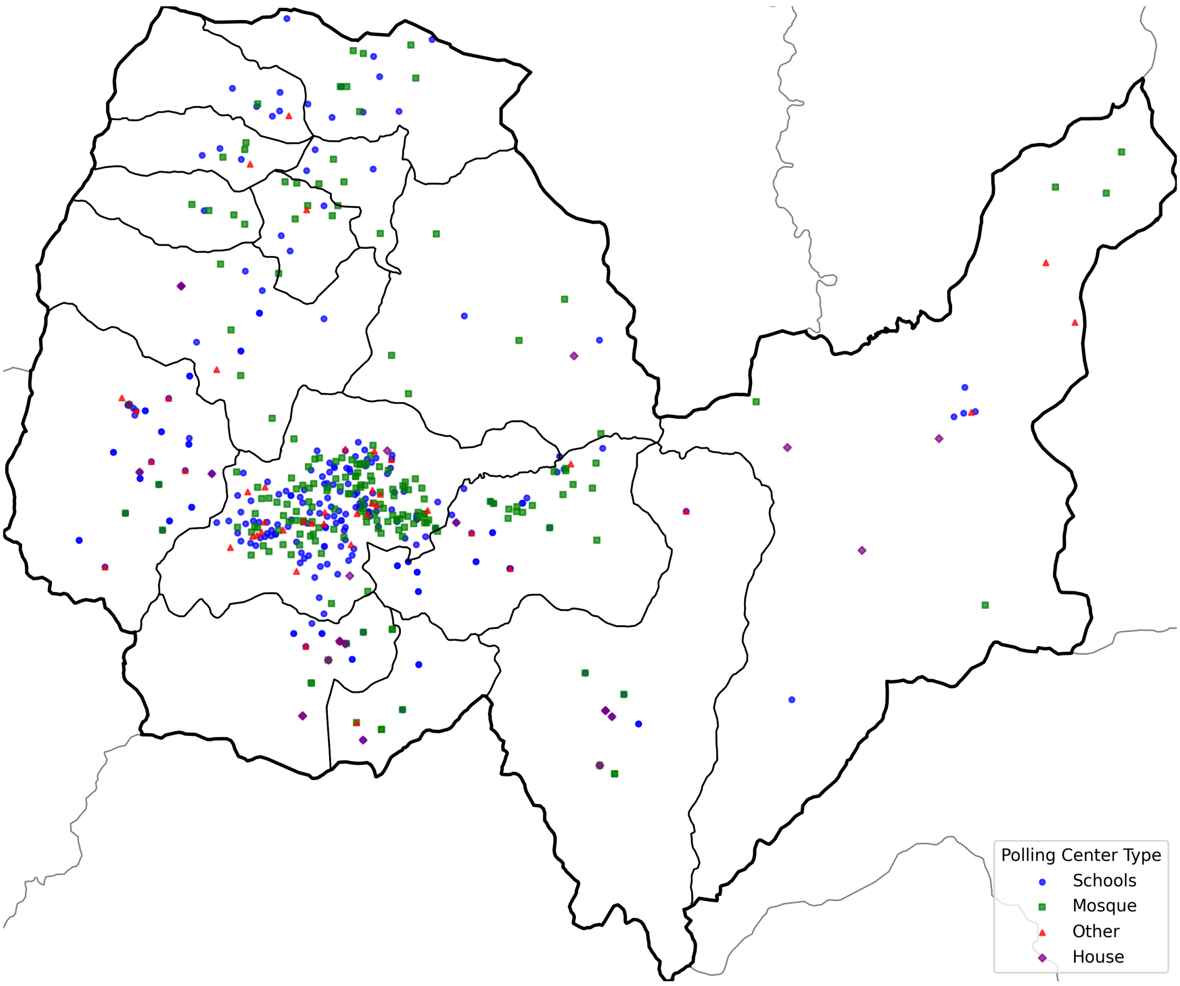


Figure 2: Polling Centers, Kabul Province

Notes: Polling centers within Kabul province.

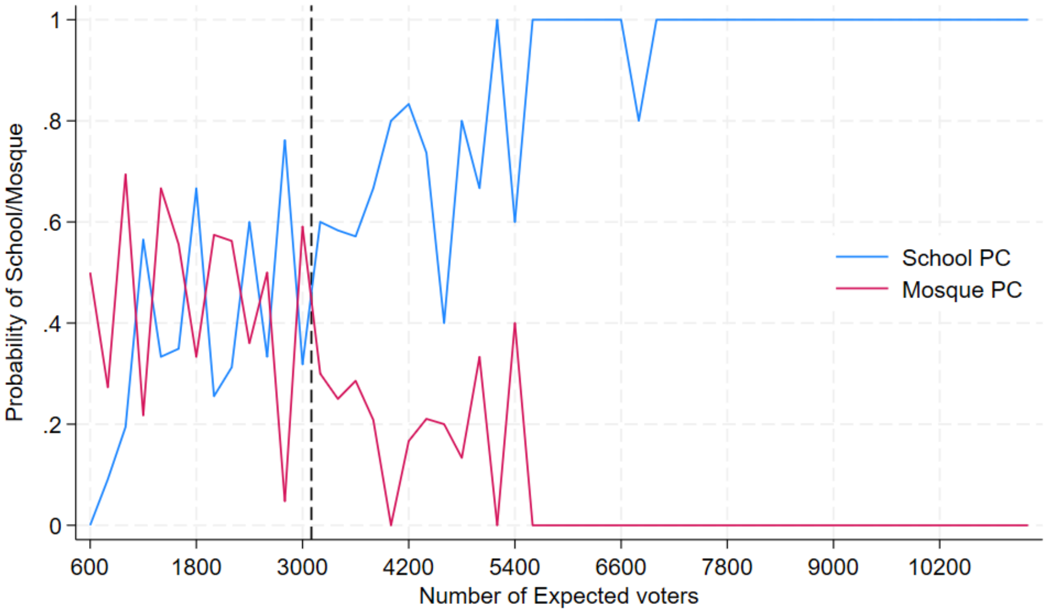


Figure 3: Probability of Polling Location and Number of Expected voters

Notes: Expected voters refers to the number of voters the Independent Election Commission (IEC) estimated to be served by a given polling center. The minimum number was 600.

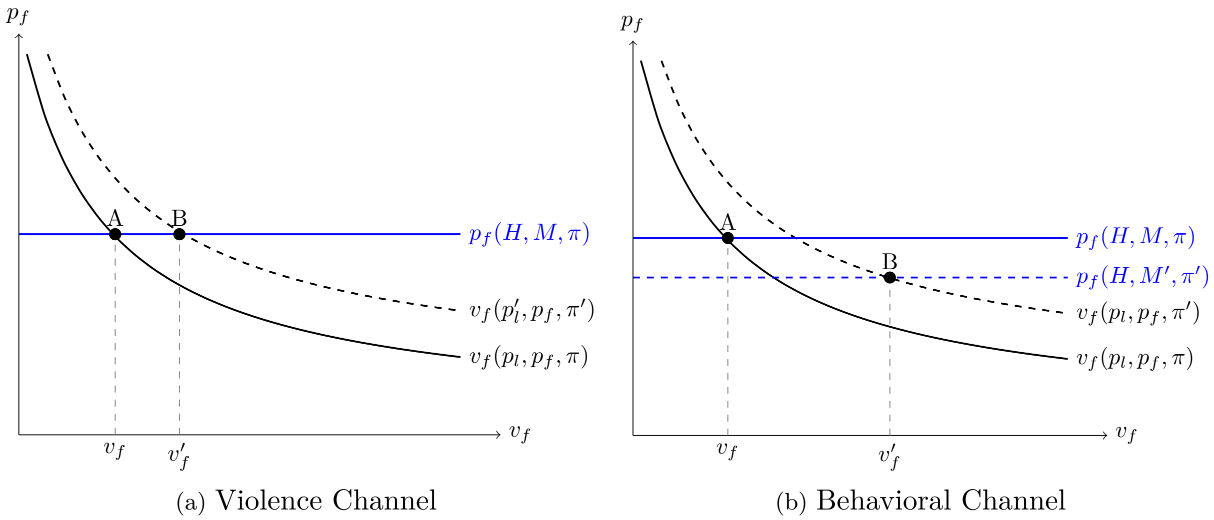
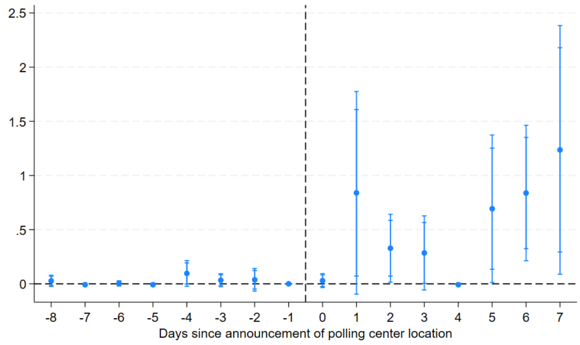
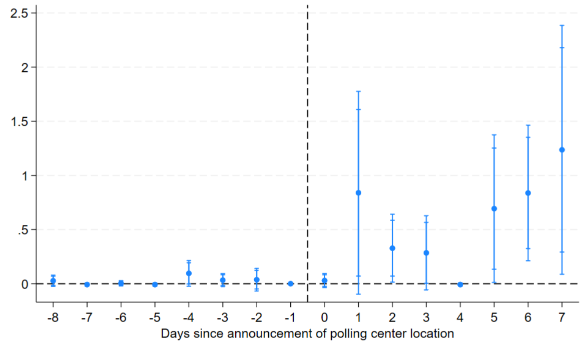


Figure 4: Graphical Representation of Mechanisms

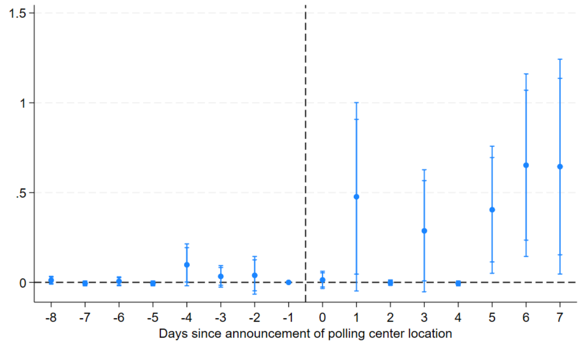
Notes: Black lines graph sample demand curves for fraudulent votes. Blue lines graph a sample price curve obtained from the election official's problem. v_f denotes fraudulent votes, p_f denotes the price of fraudulent votes, p_l denotes the price of legal votes, H denotes election-related violence, π denotes the probability of fraud being detected, and M denotes the disutility from engaging in fraud for the election official. Points A and B denote the equilibrium levels of fraud. Refer to section 6 for more information.



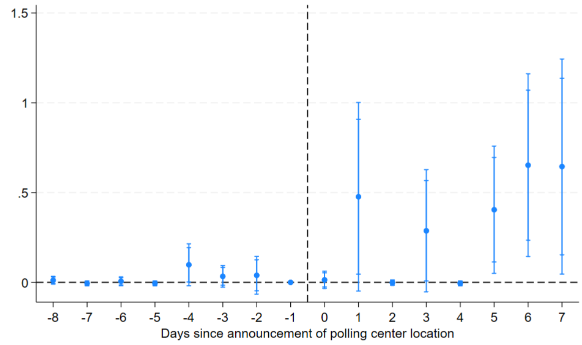
(a) All attacks



(b) All attacks + Controls



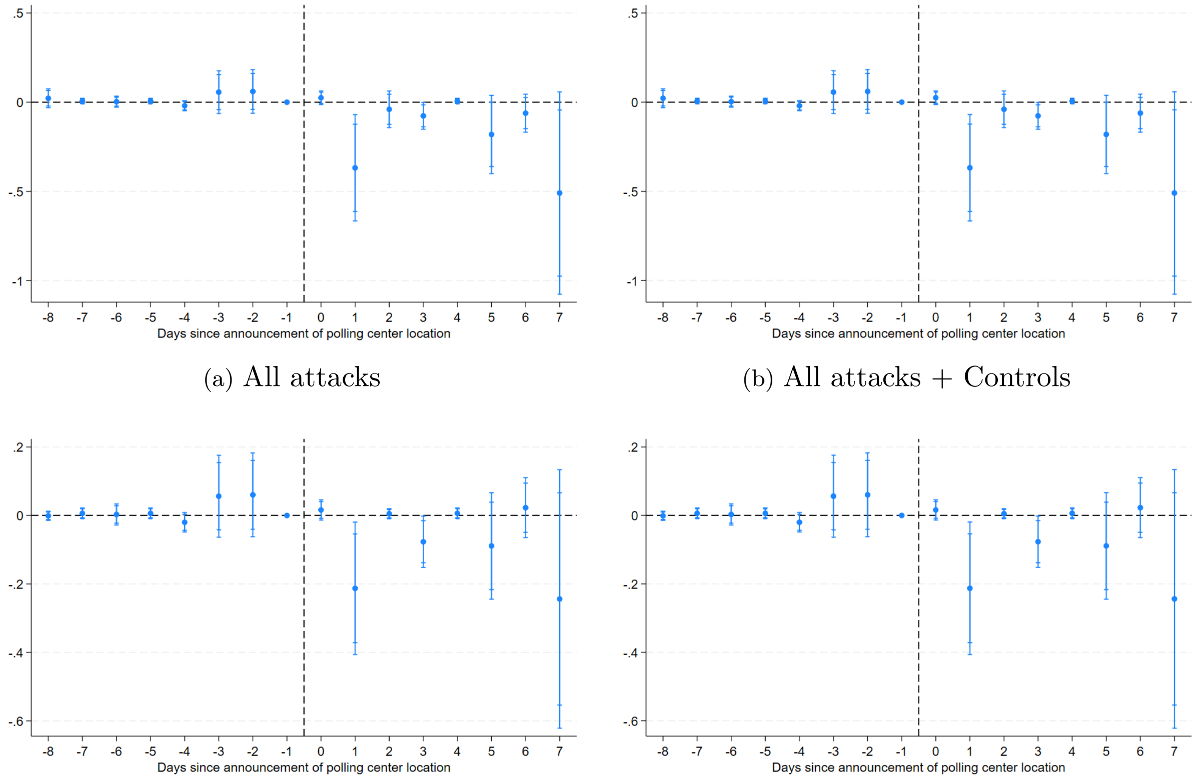
(c) Number of IEDs



(d) Number of IEDs + Controls

Figure 5: Violent Attacks after Announcement of Polling Center Location

Notes: Each dot provides an estimate of coefficient β_j from regression (9). Panels a and b use the number of all violent events within a 5-km radius of a polling center as the outcome. Panels c and d restrict this outcome to improvised explosive devices (IEDs). Panels b and d add controls to this regression. Included controls are distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, slope, indicators for language (Pashto and Other), population, and indicators for access (Limited and Other). Spikes indicate 90% and 95% confidence intervals for each estimated coefficient.



(a) All attacks

(b) All attacks + Controls

(c) Number of IEDs

(d) Number of IEDs + Controls

Figure 6: Violent Attacks after Announcement of Polling Center Location, Schools relative to mosques

Notes: Each dot provides an estimate of coefficient β_j from regression (10). Panels a and b use the number of all violent events within a 5-km radius of a polling center as the outcome. Panels c and d restrict this outcome to improvised explosive devices (IEDs). Panels b and d add controls to this regression. Included controls are distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, slope, indicators for language (Pashto and Other), population, and indicators for access (Limited and Other). Spikes indicate 90 and 95% confidence intervals for each estimated coefficient.

Table 1: Polling Center Type and Election Fraud

	Share of fraudulent votes (β)				$\mathbb{1}\{\text{Fraudulent stations} > 0\}$ (dy/dx)	
	(1)	(2)	(3)	(4)	(5)	(6)
Schools	-0.074***	-0.071***	-0.053**	-1.208***	-0.082**	-0.088***
	(0.020)	(0.020)	(0.020)	(0.299)	(0.029)	(0.016)
House	-0.032	-0.063	-0.034	-0.785	-0.045	-0.076**
	(0.055)	(0.045)	(0.047)	(0.635)	(0.063)	(0.038)
Other	-0.027	-0.049*	-0.031	-0.514	-0.032	-0.037
	(0.038)	(0.026)	(0.029)	(0.556)	(0.060)	(0.053)
Control mean	0.092	0.092	0.092	0.092	0.144	0.144
Observations	507	507	507	507	507	507
Districts	15	15	15	15	15	15
Model	OLS	OLS	OLS	Poisson	OLS	Probit
Controls	No	Yes	Yes	Yes	Yes	Yes
District FE	No	No	Yes	Yes	Yes	No

Notes: Columns 1-3 present the results of the OLS model specified in equation 1, while Column 4 presents the Poisson model variation, using the share of fraudulent votes within the center as the outcome. Columns 5 and 6 present the results of the Linear Probability and Probit models, respectively, using an indicator for whether there is at least one station within the center reporting fraud as the outcome. The results in columns 5 and 6 are reported as marginal effects, the discrete change from the base level. The z-score of the Probit model (6) is -0.676 with standard error (0.129) for schools. Columns 2-6 include controls for: distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, slope, indicators for language (Pashto and Other), population, and indicators for access (Limited and Other). Columns 6 omits language, and access due to no variation within some districts. Standard errors clustered at district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table 2: Polling Center Type (Disaggregated) and Election Fraud

	Share of fraudulent votes (β)				$\mathbb{1}\{\text{Fraudulent stations} > 0\}$ (dy/dx)	
	(1)	(2)	(3)	(4)	(5)	(6)
Primary	-0.063**	-0.074***	-0.043	-0.543	-0.068	-0.597*
	(0.024)	(0.020)	(0.027)	(0.701)	(0.048)	(0.328)
High	-0.068***	-0.058**	-0.044*	-0.854**	-0.064*	-0.454**
	(0.020)	(0.027)	(0.023)	(0.400)	(0.030)	(0.223)
Secondary	-0.088***	-0.080***	-0.070***	-2.934***	-0.110***	-1.073***
	(0.025)	(0.021)	(0.022)	(0.778)	(0.034)	(0.319)
University	-0.092***	-0.072***	-0.055***	-17.326***	-0.100***	0.000
	(0.023)	(0.010)	(0.014)	(1.007)	(0.020)	(.)
House	-0.032	-0.064	-0.034	-0.842	-0.045	-0.535
	(0.056)	(0.044)	(0.046)	(0.669)	(0.063)	(0.331)
Other	-0.027	-0.049*	-0.031	-0.574	-0.032	-0.214
	(0.038)	(0.026)	(0.029)	(0.500)	(0.061)	(0.363)
Control mean	0.092	0.092	0.092	0.092	0.144	0.144
Observations	507	507	507	507	507	500
Districts	15	15	15	15	15	15
Model	OLS	OLS	OLS	Poisson	OLS	Probit
Controls	No	Yes	Yes	Yes	Yes	Yes
District FE	No	No	Yes	Yes	Yes	No

Notes: This table presents the results of the main analysis when the polling center location category for school is disaggregated by type (primary school, high school, etc.). As in Table 1, all models use a variation of the baseline equation specified in Equation 1. Columns 1-4 present the results of the baseline OLS model and the Poisson variation using the share of fraudulent votes within the center as the outcome. Columns 5 and 6 present the results of a linear probability and a Probit model, respectively, using an indicator for whether there is at least one station within the center reporting fraud as the outcome. The results in columns 5 and 6 are reported as marginal effects, the discrete change from the base level. The z-scores of the probit model (6) are -0.592 with standard error (0.317) for primary schools, -0.429 (0.220) for high school, and -1.082 (0.329) for secondary schools. There are too few university observations to estimate effects in column 6.
Columns 2-6 include controls for: distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, slope, indicators for language (Pashto and Other), population, and indicators for access (Limited and Other). Columns 6 omits language, and access due to no variation within some districts. Column 6 excludes 7 observations due to data separation, as the “University” category was dropped for perfectly predicting failure. Standard errors clustered at district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table 3: Polling Center Type and Election Fraud: Instrumental Variable Estimates

	(1)	(2)	(3)	(4)	(5)
Panel A: Polling center is a school (First Stage)					
Expected voters > 3000	0.418***	0.515***	0.532***	0.322***	0.353***
	(0.121)	(0.094)	(0.089)	(0.089)	(0.088)
Panel B: Share of fraudulent votes (Second Stage)					
School	-0.105***	-0.101***	-0.102***	-0.153**	-0.171***
	(0.024)	(0.018)	(0.019)	(0.061)	(0.058)
Control mean	0.092	0.092	0.092	0.110	0.110
Observations	440	440	440	221	221
Clusters	15	15	15	15	15
K-P F-stat	13.0	31.6	35.8	13.1	14.9
AR p-val	0.011	0.000	0.000	0.008	0.000
Controls	No	Yes	Yes	No	No
District FE	No	No	Yes	No	Yes
Restricted sample	No	No	No	Yes	Yes

Notes: This table examines the relationship between reported fraud complaints and the type of polling center using an IV approach. Panel A presents the first stage results. Panel B presents the second stage results. The estimation sample is restricted to only mosques and schools. When indicated, the specification controls for distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, slope, population, attacks, number of female and Kuchi stations. Control mean refers to the mean of the outcome variable (Share of fraudulent votes) for polling centers located within mosques. K-P refers to the Kleibergen-Paap F-statistic for weak instruments. AR p-val refers to the p-value for the coefficient on School under weak instrument robust inference. Restricted sample refers to specifications that restrict observations to polling centers within 1,200 expected voters from the 3,000 voters cutoff. Standard errors, reported in parentheses, are clustered at the district level.
* p<.1, ** p<.05, ***p<.01

Table 4: IV Mediation Analysis of the Effect of Polling Center Type on Election Fraud

	Mediator: Post-announcement violence		Placebo mediator: Pre-announcement violence	
	All attacks (1)	IEDs (2)	All attacks (3)	IEDs (4)
Total effect	-0.102***	-0.102***	-0.102***	-0.102***
	(0.019)	(0.019)	(0.019)	(0.019)
Direct effect	-0.036**	-0.043**	-0.054**	-0.076**
	(0.018)	(0.018)	(0.021)	(0.031)
Indirect effect	-0.065**	-0.059**	-0.048	-0.026
	(0.026)	(0.025)	(0.080)	(0.066)
Control mean	0.092	0.092	0.092	0.092
Observations	440	440	440	440
Districts	15	15	15	15
Mediator effect (%)	64	58	47	25
F-stat (T on Z)	35.71	35.71	35.71	35.71
F-stat (M on Z | T)	9.22	10.52	0.80	0.78
Controls	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes

Notes: The table reports results from IV mediation analysis of the effect of polling center type (school) on the share of votes above 600 at the polling center. We use the mediation approach proposed in Dippel et al. [2020]. The estimation sample is restricted to only mosques and schools. Columns 1 and 2 use the number of violent events within 5 km of the polling center in the 7-day period *after* polling center locations were publicly announced until election day as the mediator. Columns 3 and 4 implement a placebo test in which the mediator is defined for the *pre-announcement* 7-day period before polling center locations were published. Columns 1 and 3 use all violent attacks. Columns 2 and 4 use improvised explosive device (IED) incidents. *Total effect* is the estimated causal effect of treatment on the outcome reported on Column 3 of Table 3. *Direct effect* reports the portion of the effect not operating through the mediator. *Indirect effect* reports the portion operating through the mediator. "Control mean" refers to the mean of the outcome variable for polling centers located within mosques. "Mediator effect" is the estimated proportion of the total effect mediated through violence. "F-stat (T on Z)" is the first-stage F-statistic from the regression of the treatment on the instrument. "F-stat (M on Z|T)" is the first-stage F-statistic from the regression of the mediator on the instrument controlling for treatment. All models include district fixed effects and the full set of controls: distance from the polling center to the closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, slope, population, attacks, and number of female and Kuchi stations. Standard errors (in parentheses) are clustered at the district level.
* p<.1, ** p<.05, ***p<.01

Table 5: Polling Center Type and Complaints

		Complainant		Respondent	
	Complaints	Female	Individual	Polling official	Candidate
	(1)	(2)	(3)	(4)	(5)
Panel A: Number of complaints					
Schools	0.613***	0.824***	0.502**	0.118	0.836**
	(0.185)	(0.174)	(0.208)	(0.198)	(0.376)
House	0.032	1.816**	-13.997***	0.138	0.039
	(0.251)	(0.822)	(1.055)	(0.307)	(1.004)
Other	0.053	0.629**	0.684**	-0.230***	0.502
	(0.214)	(0.305)	(0.266)	(0.087)	(0.452)
Control mean	0.526	0.231	0.215	1.169	0.154
Observations	507	230	230	230	230
Districts	15	14	14	14	14
Panel B: Share of complaints					
Schools		0.522	0.678***	-0.184***	0.762***
		(0.379)	(0.250)	(0.051)	(0.287)
House		1.640**	-11.857***	0.254	-0.032
		(0.826)	(0.953)	(0.350)	(0.774)
Other		0.577	1.115***	-0.062	0.270
		(0.553)	(0.206)	(0.077)	(0.248)
Control mean		0.135	0.096	0.733	0.095
Observations		230	230	230	230
Districts		14	14	14	14

Notes: This table examines the relationship between reported fraud complaints and the type of polling center from which they originate. Panel A uses the total number of complaints as the dependent variable, while Panel B considers the share of complaints. Column 1 reports estimates using the total number of complaints as the outcome variable. Columns 2 and 3 further differentiate complaints by the complainant's identity, specifically whether the complaint was filed by a female or submitted by an individual. Columns 4 and 5 analyze the type of suspect most frequently identified across different categories of polling center locations. All models are estimated using a Poisson specification of Equation (1) and include controls for distance from the polling center to the closest town, primary road, secondary road, tertiary road, health facility, primary and secondary rivers, elevation, and slope, along with district fixed effects. Standard errors, reported in parentheses, are clustered at the district level.

* p<.1, ** p<.05, ***p<.01

Table 6: School Assignment Based on Violence in Catchment Areas

	𝟙{Polling Center Assigned to School} (dy/dx)				
	(1)	(2)	(3)	(4)	(5)
All Attacks	-0.007	-0.002	-0.001	-0.007	-0.002
	(0.007)	(0.005)	(0.007)	(0.007)	(0.005)
IED Attacks	-0.004	-0.002	-0.000	-0.004	-0.002
	(0.008)	(0.007)	(0.008)	(0.008)	(0.007)
Observations	507	507	507	507	506
Districts	15	15	15	15	15
Model	OLS	OLS	OLS	Probit	Probit
Controls	No	Yes	Yes	No	Yes
District FE	No	No	Yes	No	No

Notes: This table presents the marginal effects of violence in a catchment area between March 01, 2009, and August 13, 2009, on the probability that the assigned polling center in the catchment area will be a school. The outcome is a binary indicator equal to 1 if the polling center in a catchment area is a school. When indicated, the specification controls for distance from the polling center to the closest town, primary road, secondary road, tertiary road, health facility, primary and secondary rivers, elevation, slope, population, language, and the number of female and Kuchi stations.
* p<.1, ** p<.05, ***p<.01

Table 7: School Assignment and Dominant Language Spoken

	𝟙{Polling Center Assigned to School} (dy/dx)				
	(1)	(2)	(3)	(4)	(5)
Panel A: In Nearest Village					
Pashto	-0.066	-0.058	-0.026	-0.065	-0.057
	(0.040)	(0.066)	(0.071)	(0.040)	(0.065)
Panel B: In Catchment Area					
Pashto	-0.036	-0.014	0.025	-0.036	-0.012
	(0.040)	(0.058)	(0.063)	(0.040)	(0.058)
Observations	507	507	507	507	507
Districts	15	15	15	15	15
Model	OLS	OLS	OLS	Probit	Probit
Controls	No	Yes	Yes	No	Yes
District FE	No	No	Yes	No	No

Notes: This table presents the marginal effects of language on school assignment. Panel A shows the marginal effects on the probability of a polling center being assigned to a school when Pashto is the dominant language spoken in the nearest village. Panel B presents the marginal effects of language on the probability of school assignment when Pashto is the most common language spoken in the village(s) of the catchment area of a polling center. The language of the nearest village is used for polling centers in urban areas where there are no villages. When indicated, the specification controls for distance from the polling center to the closest town, primary road, secondary road, tertiary road, health facility, primary and secondary rivers, elevation, slope, population, attacks, and the number of female and Kuchi stations.
* p<.1, ** p<.05, ***p<.01

Appendix A Additional Figures and Tables

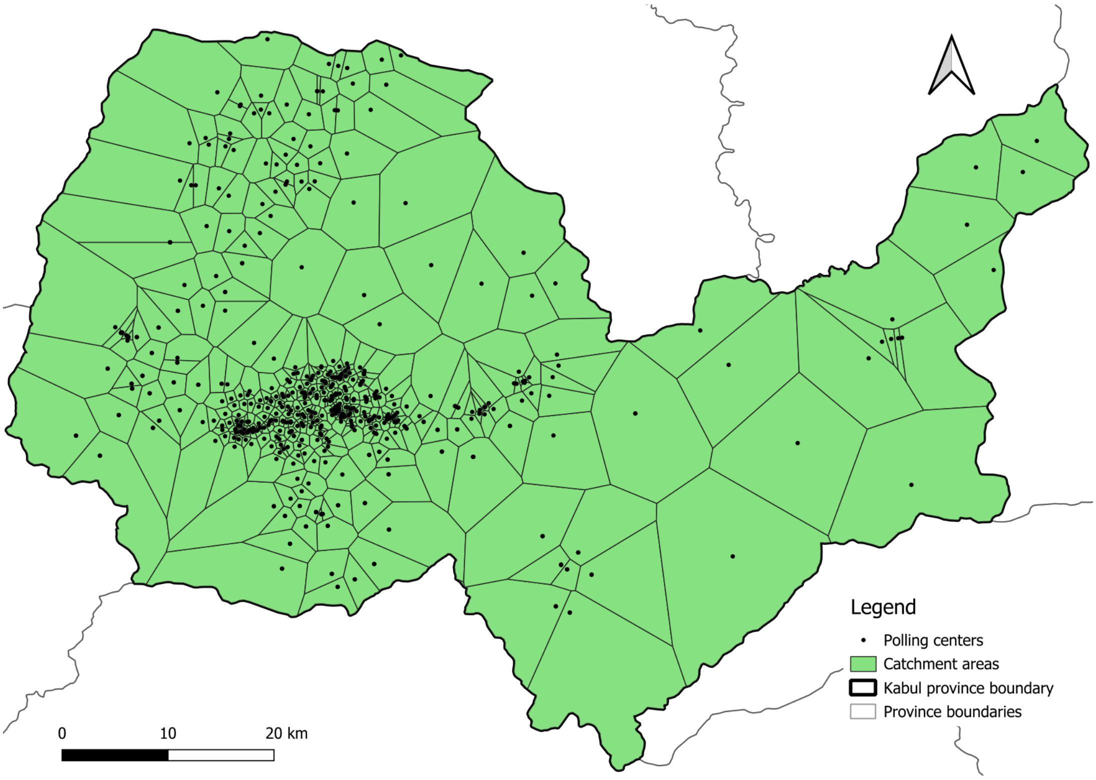


Figure A1: Polling Center Catchment Areas, Kabul Province

Notes: Catchment areas calculated as Voronoi polygons. Any point within a given catchment area is closest to the polling center associated with that catchment area.

Table A1: Descriptive Statistics

Variable	(1) Count	(2) Mean	(3) SD	(4) Min	(5) Max
Share of Fraudulent Votes	507	0.05	0.19	0.00	1.00
Indicator of Fraud	507	0.09	0.28	0.00	1.00
Total Polling Stations	507	5.50	2.94	2.00	19.00
Share of Kuchi Polling Stations	507	0.07	0.20	0.00	1.00
Share of Female Polling Stations	507	0.38	0.09	0.00	1.00
Distance to Village Center (km)	507	0.62	0.47	0.02	4.33
Elevation (km)	507	1.85	0.19	0.97	2.48
Terrain Slope	507	3.33	3.27	0.09	18.38
Distance to Primary Road (km)	507	2.75	4.39	0.00	25.42
Distance to Secondary Road (km)	507	4.02	4.74	0.01	25.83
Distance to Tertiary Road (km)	507	0.83	1.36	0.00	13.59
Distance to Basic Health Facility (km)	507	4.55	4.18	0.20	23.89
Distance to District Health Facility (km)	507	6.09	6.42	0.15	33.81
Distance to Primary River (km)	507	24.37	9.80	0.04	46.57
Distance to Secondary River (km)	507	1.19	1.25	0.00	8.94
Pashto	507	0.37	0.48	0.00	1.00
Dari	507	0.63	0.48	0.00	1.00
Other Languages	507	0.00	0.04	0.00	1.00
District Capital	507	0.07	0.26	0.00	1.00
Road Access	507	0.94	0.23	0.00	1.00
Limited Access	507	0.02	0.15	0.00	1.00
Other Access	507	0.03	0.17	0.00	1.00
Pre-election attacks within 5km	507	0.89	1.55	0.00	8.00
Local Population	507	4401.87	4033.48	8.87	10602.52

Notes: Descriptive statistics for polling centers within the Kabul province.

Table A2: Polling Center Type and Election Fraud-Category C fraud

	Share of fraudulent votes (β)				$\mathbb{1}\{\text{Fraudulent stations} > 0\}$ (dy/dx)	
	(1)	(2)	(3)	(4)	(5)	(6)
Schools	-0.031	-0.041*	-0.026	-1.744**	-0.030*	-4.341***
	(0.022)	(0.020)	(0.016)	(0.855)	(0.015)	(1.667)
House	-0.001	-0.047	-0.021	-0.177	-0.018	0.452
	(0.057)	(0.048)	(0.045)	(1.484)	(0.053)	(0.819)
Other	0.002	-0.026	-0.013	0.015	-0.014	1.479
	(0.027)	(0.026)	(0.023)	(0.678)	(0.025)	(1.779)
Control mean	0.036	0.036	0.036	0.036	0.046	0.046
Observations	507	507	507	507	507	507
Districts	15	15	15	15	15	15
Model	OLS	OLS	OLS	Poisson	LPM	Probit
Controls	No	Yes	Yes	No	Yes	Yes
District FE	No	No	Yes	Yes	Yes	No

Notes: This table replicates Table 1 using Category C fraud as the outcome. Refer to Section 3 for the distinction between Category C fraud and the fraud measure used in Table 1. Columns 1-3 present the results of the OLS model specified in equation (1), while Column 4 presents the Poisson model variation, using the share of fraudulent votes within the center as the outcome. Columns 5 and 6 present the results of the Linear Probability and Probit models, respectively, using an indicator for whether there is at least one station within the center reporting fraud as the outcome. The results in columns 5 and 6 are reported as marginal effects, the discrete change from the base level. The z-score of the Probit model (C6) is -0.676 with standard error (0.129) for schools.
Columns 2-6 include controls for: distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, slope, indicators for language (Pashto and Other), population, and indicators for access (Limited and Other). Columns 6 omits language, and access due to no variation within some districts. Standard errors clustered at district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table A3: Polling Center Type and number of stations with 600+ votes

	(1)	(2)	(3)	(4)
Schools	-0.274**	-0.335***	-0.215***	-1.285***
	(0.098)	(0.102)	(0.069)	(0.260)
House	-0.154	-0.315	-0.107	-0.668
	(0.284)	(0.226)	(0.196)	(0.627)
Other	-0.167	-0.301**	-0.152	-0.583
	(0.136)	(0.128)	(0.095)	(0.644)
Control mean	0.371	0.371	0.371	0.371
Observations	507	507	507	507
Districts	15	15	15	15
Model	OLS	OLS	OLS	Poisson
Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes

Notes: Results use Equation (1) with the number of stations with 600+ votes within the center as the outcome variable. Columns 2-4 include controls for: distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, slope, indicators for language (Pashto and Other), population, and indicators for access (Limited and Other). Standard errors clustered at district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table A4: Polling Center Type and Election Fraud with Parsimonious Controls

	Share of fraudulent votes (β)				$\mathbb{1}\{\text{Fraudulent stations} > 0\}$ (dy/dx)	
	(1)	(2)	(3)	(4)	(5)	(6)
Schools	-0.074***	-0.071***	-0.063***	-1.512***	-0.096***	-0.115***
	(0.020)	(0.020)	(0.017)	(0.314)	(0.020)	(0.023)
House	-0.032	-0.038	-0.010	-0.177	-0.026	-0.082
	(0.055)	(0.054)	(0.048)	(0.795)	(0.058)	(0.061)
Other	-0.027	-0.030	-0.016	-0.216	-0.017	-0.035
	(0.038)	(0.036)	(0.038)	(0.627)	(0.072)	(0.071)
Control mean	0.092	0.092	0.092	0.092	0.144	0.151
Observations	507	507	507	507	507	479
Districts	15	15	15	15	15	15
Model	OLS	OLS	OLS	Poisson	OLS	Probit
Controls	No	Yes	Yes	Yes	Yes	Yes
District FE	No	No	Yes	Yes	Yes	No

Notes: This table replicates the results of Table 1 using a parsimonious set of controls in columns 2-6. Specifically, columns 2-6 include controls for the reported levels of security and accessibility that were used by the IEC to assess the feasibility of a location being a polling center. The results in columns 5 and 6 are reported as marginal effects, the discrete change from the base level. The z-score of the Probit model (6) is -0.750 with standard error (0.107) for schools. Column 6 loses 28 observations because Limited access and Moderate Security predict zero perfectly in some districts. Standard errors clustered at district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table A5: Balance in Polling Center Characteristics Around 3,000-Expected Voters Threshold

Variable	(1) Coefficient	(2) S.E.	(3) Control mean
Share female stations	-0.050	(0.00)	0.390
Share Kuchis stations	0.035	(0.03)	0.060
Distance to village center (km)	-0.016	(0.03)	0.620
Elevation (meters)	-16.69	(21.19)	1886
Terrain slope	0.212	(0.37)	3.540
Distance to nearest primary road (km)	-0.272	(0.27)	3.450
Distance to nearest secondary road (km)	-0.231	(0.23)	4.760
Distance to nearest tertiary road (km)	-0.114	(0.16)	0.940
Distance to basic health facility (km)	-0.059	(0.20)	5.070
Distance to district health facility (km)	-0.108	(0.25)	6.710
Distance to primary river (km)	-0.057	(0.27)	24.44
Distance to secondary river (km)	-0.074	(0.07)	1.270
Pre-election attacks within 5km	0.046	(0.06)	0.800
Local population	80.05	(42.36)	3718
Pashto language	-0.003	(0.04)	0.420

Notes: Each row reports the coefficient from a regression of the indicated variable on an indicator for whether the polling center had more than 3,000 expected voters. All regressions include district fixed effects, and standard errors are clustered at the district level. The control mean is the average of each variable among centers with 3,000 or fewer expected voters. The sample is restricted to polling centers located in either schools or mosques.

Table A6: Violent attacks after announcement of polling center locations

	All Attacks		IED Attacks	
	(1)	(2)	(3)	(4)
8 Days Before	0.027	0.458***	0.012	0.270***
	(0.024)	(0.109)	(0.010)	(0.057)
7 Days Before	-0.008		-0.006	
	(0.007)		(0.006)	
6 Days Before	0.004	0.005	0.006	0.007
	(0.011)	(0.014)	(0.012)	(0.015)
5 Days Before	-0.008	-0.010	-0.006	-0.007
	(0.007)	(0.010)	(0.006)	(0.008)
4 Days Before	0.096	0.120**	0.098*	0.122**
	(0.055)	(0.054)	(0.054)	(0.053)
3 Days Before	0.033	0.042	0.033	0.042
	(0.028)	(0.037)	(0.028)	(0.037)
2 Days Before	0.037	0.046	0.039	0.049
	(0.049)	(0.064)	(0.049)	(0.065)
1 Day Before	0.000	0.000	0.000	0.000
	-	-	-	-
Day of Announcement	0.029	0.037	0.014	0.017
	(0.030)	(0.041)	(0.022)	(0.029)
1 Day After	0.840*	1.051**	0.477*	0.597**
	(0.436)	(0.407)	(0.245)	(0.227)
2 Days After	0.328**	0.411***	-0.002	-0.002
	(0.146)	(0.132)	(0.007)	(0.009)
3 Days After	0.285*	0.357**	0.287*	0.359**
	(0.159)	(0.153)	(0.158)	(0.151)
4 Days After	-0.008	-0.010	-0.006	-0.007
	(0.007)	(0.010)	(0.006)	(0.008)
5 Days After	0.693**	0.868***	0.404**	0.506***
	(0.317)	(0.280)	(0.165)	(0.138)
6 Days After	0.838**	1.049***	0.652**	0.817***
	(0.292)	(0.224)	(0.237)	(0.188)
Day of Election	1.236**	1.548***	0.645**	0.807***
	(0.535)	(0.462)	(0.279)	(0.239)
Prior-Year Mean	0.034	0.034	0.031	0.031
Observations	8192	5829	8192	5829
Districts	15	15	15	15
Model	OLS	OLS	OLS	OLS
Controls	No	Yes	No	Yes
District FE	Yes	Yes	Yes	Yes

Notes: This table provides estimates of coefficient β_j from regression (9) that correspond with the point estimates in Figure 5. Columns 1 and 2 of this table correspond with panels a and b of Figure 5, using number of all violent events within a 5-km radius of a polling center as outcome. Columns 3 and 4 of this table correspond with panels c and d of Figure 5, restricting the attack outcome to improvised explosive devices (IEDs). Columns 2 and 4 add controls to the regression. Included controls are distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, slope, indicators for language (Pashto and Other), population, and indicators for access (Limited and Other). Standard errors are clustered at the district level.
* p<.1, ** p<.05, ***p<.01

Table A7: Violent attacks after the announcement of polling center locations in schools compared to mosques

	All Attacks		IED Attacks	
School	(1)	(2)	(3)	(4)
8 Days Before	0.010	0.471***	0.005	0.287***
	(0.010)	(0.126)	(0.006)	(0.073)
7 Days Before	-0.010		-0.010	
	(0.012)		(0.012)	
6 Days Before	0.005	0.006	0.005	0.006
	(0.016)	(0.021)	(0.016)	(0.021)
5 Days Before	-0.010	-0.013	-0.010	-0.013
	(0.012)	(0.016)	(0.012)	(0.016)
4 Days Before	0.112*	0.142**	0.112*	0.142**
	(0.062)	(0.057)	(0.062)	(0.057)
3 Days Before	-0.000	-0.000	0.000	-0.000
	(0.007)	(0.009)	(0.007)	(0.009)
2 Days Before	0.000	-0.000	0.000	-0.000
	(0.015)	(0.019)	(0.015)	(0.019)
1 Days Before	0.000	0.000	0.000	0.000
	-	-	-	-
Day of Announcement	0.015	0.019	0.000	-0.000
	(0.024)	(0.032)	(0.015)	(0.019)
1 Day After	1.087**	1.374***	0.622**	0.787***
	(0.484)	(0.390)	(0.280)	(0.227)
2 Days After	0.357**	0.452***	-0.005	-0.006
	(0.154)	(0.124)	(0.012)	(0.015)
3 Days After	0.342*	0.432***	0.342*	0.432***
	(0.163)	(0.137)	(0.163)	(0.137)
4 Days After	-0.010	-0.013	-0.010	-0.013
	(0.012)	(0.016)	(0.012)	(0.016)
5 Days After	0.832**	1.052***	0.474**	0.600***
	(0.341)	(0.259)	(0.185)	(0.134)
6 Days After	0.893**	1.129***	0.648**	0.819***
	(0.304)	(0.198)	(0.223)	(0.147)
Day of Election	1.582**	2.000***	0.806**	1.019***
	(0.636)	(0.482)	(0.348)	(0.275)
Prior-Year Mean	0.034	0.034	0.031	0.031
Observations	8192	5829	8192	5829
Districts	15	15	15	15
Model	OLS	OLS	OLS	OLS
Controls	No	Yes	No	Yes
District FE	Yes	Yes	Yes	Yes

Notes: This table provides estimates of coefficient β_j from regression (10) that correspond with the point estimates in Figure 6. Columns 1 and 2 of this table correspond with panels a and b of Figure 6, using number of all violent events within a 5-km radius of a polling center as outcome. Columns 3 and 4 of this table correspond with panels c and d of Figure 6, restricting the attack outcome to improvised explosive devices (IEDs). Columns 2 and 4 add controls to the regression. Included controls are distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, slope, indicators for language (Pashto and Other), population, and indicators for access (Limited and Other). Standard errors are clustered at the district level.

* p<.1, ** p<.05, ***p<.01

Table A8: Share of Female and Kuchi Stations in a Polling Center

	Share of female stations			Share of Kuchi stations		
	(1)	(2)	(3)	(4)	(5)	(6)
Schools	-0.029***	-0.024***	-0.023***	0.001	-0.002	0.005
	(0.007)	(0.007)	(0.007)	(0.023)	(0.019)	(0.016)
House	0.029	0.017	0.015	0.095	0.035	0.055
	(0.030)	(0.026)	(0.029)	(0.077)	(0.068)	(0.059)
Other	-0.013	-0.010	-0.009	0.046	0.052	0.066
	(0.021)	(0.020)	(0.020)	(0.041)	(0.033)	(0.038)
Control mean	0.092	0.092	0.092	0.092	0.092	0.092
Observations	507	507	507	507	507	507
Districts	15	15	15	15	15	15
Model	OLS	OLS	OLS	OLS	OLS	OLS
Controls	No	Yes	Yes	No	Yes	Yes
District FE	No	No	Yes	No	No	Yes

Notes: Columns 1-3 use the share of female stations within the center as the outcome. Columns 4-6 use the share of Kuchi stations within the center as the outcome. Columns 2-3 and 5-6 include controls for distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, slope, indicators for language (Pashto and Other), population, and indicators for access (Limited and Other). Columns 1-3 include a control for the share of Kuchi polling stations in the center, and columns 4-6 control for the share of female polling stations in the center. Standard errors clustered at district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table A9: Polling Center Type and Complaints (controlling for number of complaints)

	Complainant		Respondent	
	Female	Individual	Polling official	Candidate
	(1)	(2)	(3)	(4)
Schools	0.613**	0.343*	-0.096	0.703**
	(0.259)	(0.177)	(0.116)	(0.316)
House	1.706*	-19.681***	0.154	0.105
	(0.882)	(1.012)	(0.321)	(0.836)
Other	0.841**	1.041***	-0.131	0.430
	(0.327)	(0.252)	(0.097)	(0.382)
Control mean	0.231	0.215	1.169	0.154
Observations	230	230	230	230
Districts	14	14	14	14

Notes: This table examines the relationship between number of reported fraud complaints and the type of polling center from which they originate while controlling for the number of complaints in a polling center. Columns (1) and (2) differentiate complaints based on the identity of the complainant, specifically whether the complaint was filed by a female or submitted by an individual. Columns (3) and (4) analyze the type of suspect most frequently identified across different categories of polling center locations. All models are estimated using a Poisson specification of Equation (1) and include controls for distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, and slope, along with district fixed effects. Standard errors, reported in parentheses, are clustered at the district level.
* p<.1, ** p<.05, ***p<.01

Table A10: Polling Center Type and Likelihood of Complaints

		Complainant		Respondent	
	$\mathbb{1}\{\text{Complaints} > 0\}$	$\mathbb{1}\{\text{Female} > 0\}$	$\mathbb{1}\{\text{Individual} > 0\}$	$\mathbb{1}\{\text{Polling Official} > 0\}$	$\mathbb{1}\{\text{Candidate} > 0\}$
	(1)	(2)	(3)	(4)	(5)
Schools	0.340**	0.241	0.600***	-0.464**	0.602**
	(0.172)	(0.231)	(0.173)	(0.227)	(0.241)
House	0.127	1.147	0.000	0.008	-0.477
	(0.242)	(0.819)	(.)	(0.769)	(0.677)
Other	0.038	0.395	0.680**	0.137	0.406
	(0.128)	(0.541)	(0.278)	(0.217)	(0.322)
Control mean	0.340	0.246	0.153	0.783	0.148
Observations	500	203	190	209	222
Districts	14	10	9	9	11

Notes: This table examines the relationship between the likelihood of reported fraud complaints and the type of polling center from which they originate. Column 1 reports estimates using the likelihood of at least one complaint as the outcome variable. Columns 2 and 3 further differentiate the probability of complaints based on the identity of the complainant, specifically whether the complaint was filed by a female or submitted by an individual. Columns 4 and 5 analyze the type of suspect most likely to be identified across different categories of polling center locations. All models are estimated using a Probit specification and include controls for distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, and slope, along with district fixed effects.Standard errors, reported in parentheses, are clustered at the district level.
* p<.1, ** p<.05, ***p<.01

Table A11: Polling Center Type and Election Fraud (Wild Bootstrap)

	Share of fraudulent votes (β)				$\mathbb{1}\{\text{Fraudulent stations} > 0\}$ (dy/dx)	
	(1)	(2)	(3)	(4)	(5)	(6)
Schools	-0.074***	-0.071***	-0.053**	-1.208***	-0.082**	-0.676***
	(0.020)	(0.020)	(0.020)	(0.299)	(0.029)	(0.129)
House	-0.032	-0.063	-0.034	-0.785	-0.045	-0.541
	(0.055)	(0.045)	(0.047)	(0.635)	(0.063)	(0.347)
Other	-0.027	-0.049*	-0.031	-0.514	-0.032	-0.219
	(0.038)	(0.026)	(0.029)	(0.556)	(0.060)	(0.357)
Wild Bootstrap p-value (School)	0.003	0.022	0.069	0.073	0.070	0.011
Control mean	0.092	0.092	0.092	0.092	0.144	0.144
Observations	507	507	507	507	507	507
Districts	15	15	15	15	15	15
Model	OLS	OLS	OLS	Poisson	OLS	Probit
Controls	No	Yes	Yes	Yes	Yes	Yes
District FE	No	No	Yes	Yes	Yes	No

Notes: This table presents the results of the main analysis along with the wild-bootstrap p-values for the coefficient of Schools. Columns 1-3 present the results of the OLS model specified in equation 1, while Column 4 presents the Poisson model variation, using the share of fraudulent votes within the center as the outcome. Columns 5 and 6 present the results of the Linear Probability and Probit models, respectively, using an indicator for whether there is at least one station within the center reporting fraud as the outcome. The results in columns 5 and 6 are reported as marginal effects, the discrete change from the base level. The z-score of the Probit model (6) is -0.676 with standard error (0.129) for schools.

Columns 2-6 include controls for: distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, slope, indicators for language (Pashto and Other), population, and indicators for access (Limited and Other). Columns 6 omits language, and access due to no variation within some districts. Columns 3 and 5 lose one observation due to a singleton district. Standard errors clustered at district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table A12: Polling Center Type and Election Fraud: Instrumental Variable Estimates with Wild Bootstrap

	(1)	(2)	(3)	(4)	(5)
Panel A: Polling center is a school (First Stage)					
Expected voters > 3000	0.418***	0.515***	0.532***	0.322***	0.353***
	(0.121)	(0.094)	(0.089)	(0.089)	(0.088)
Panel B: Share of fraudulent votes (Second Stage)					
School	-0.105***	-0.101***	-0.102***	-0.153**	-0.171***
	(0.024)	(0.018)	(0.019)	(0.061)	(0.058)
Wild Bootstrap p-value (School)	0.094	0.007	0.027	0.113	0.050
Control mean	0.092	0.092	0.092	0.110	0.110
Observations	440	440	440	221	221
Clusters	15	15	15	15	15
K-P F-stat	13.0	31.6	35.8	13.1	14.9
AR p-val	0.011	0.000	0.000	0.008	0.000
Controls	No	Yes	Yes	No	No
District FE	No	No	Yes	No	Yes
Restricted sample	No	No	No	Yes	Yes

Notes: This table presents the results of the IV analysis along with the wild bootstrap p-values for the coefficents of School. Panel A presents the first stage results. Panel B presents the second stage results. Estimation sample restricted to only mosques and schools. When indicated, the specification controls for distance from polling center to closest town, primary road, secondary road, tertiary road, health facility, primary and secondary river, elevation, slope, population, attacks, number of female and Kuchi stations. Control mean refers to the mean of the outcome variable (Share of fraudulent votes) for polling centers located within mosques. K-P refers to the Kleibergen-Paap F-statistic for weak instruments. AR p-val refers to the p-value for the coefficient on School under weak instrument robust inference. Restricted sample refers to specifications that restricts observations to polling centers within 1,200 expected voters from the 3,000 voters cutoff. Standard errors, reported in parentheses, are clustered at the district level.

* p<.1, ** p<.05, ***p<.01