Polling Place Matters: How Voting Location Influences Election Fraud

Tannishtha Ghosh* Robert M. Gonzalez† Kelly Lifchez‡

March 4, 2025

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, mosque) influences election fraud. Combining validated fraud measures during the 2009 Afghan presidential election with a novel instrumental variable approach, we find that polling centers within schools report an 10-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 show systematic differences in voter and polling official behavior between schools and mosques. Compared to mosques, the number of filed complaints and the share of complaints from individual voters and women is higher in schools, while polling officials are less likely to be the subject of complaints.

Keywords: election fraud, polling location, insurgent violence

JEL codes: D72, D73, K42

*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.

This paper studies whether polling location type affects the likelihood of election fraud. We study this question in the context of the 2009 Afghan presidential election, a setting where electoral fraud was widespread and systematically documented. Using validated fraud measures, we compare polling centers located in schools and mosques, two of the most common polling locations.

We explore this question by relying on several unique sources of data. First, we combine data on the location and characteristics of polling centers collected during a security assessment done 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 for our study because polling center classification was more accurate in Kabul—likely due to higher quality reporting of center names—and it had fewer ambiguous center names compared to other large provinces.

Second, we explore mechanisms using declassified significant actions (SIGACTs) data on insurgent attacks recorded by Afghan and ISAF forces. These data provide precise timestamps, geolocations, and incident types (e.g., IED explosion, direct fire event). We merge SIGACTs data with polling center geolocation data to identify centers attacked around election day. Finally, we incorporate data on all polling-center-level complaints received during the 2010 election in order to assess differences in voter and polling official behavior across polling center types.

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

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

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 leverage primary data from security assessments by ISAF teams of each polling center. These data provide information on the two key determinants of polling center choice: security environment in the area (security and how defensible the location was) 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 IEC allocated one voting station for every 600 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 likely driven by logistical considerations rather than political manipulation–schools have more rooms that can be used as polling stations relative mosques. We provide evidence that other polling center characteristics do not change significantly at this threshold.

Our findings show that the choice of polling center matters. We find evidence that there is less fraud occurring in schools than in mosques. Specifically, we find that polling centers located within schools exhibit an 8 percentage points lower likelihood of fraud compared to polling centers located within mosques. Our findings are robust to different model specifications, including various functional forms and covariates. Importantly, our results are 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, and the findings from a Poisson specification suggest that the likelihood of Category C fraud is reduced by 70% when polling centers are located in schools, highlighting the substantial impact of location on fraudulent activity.

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 locations are associated with a 10 to 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 underlying the relationship between polling center location and fraud. First, we consider the impact of violence on fraud. Election-related violence can affect fraud through a vote-substitution mechanism. If a certain type of polling center is more likely to be targeted by violence, then turnout will decrease in that center. Corrupt officials can respond by substituting the lost votes with fraudulent votes. Using daily incidents data from SIGACTS, we document a significant increase in attacks following the disclosure of polling sites, suggesting strategically targeted attacks by insurgent groups. However, we note that the pattern of these attacks vary across polling center types, with mosques experience attacks a higher rate than schools and other locations. To formalize these insights, we also examine the incidence of overall violence by polling center category in the year leading up to the election. Our findings show that in general, the likelihood of a school being attacked 2 percentage points lower than a mosque. However, this effect grows weaker and loses significance once we include controls and district fixed effects. Nevertheless, the discrepancy in exposure to violence aligns with our findings on fraud, indicating that reduced turnout due to attacks may have influenced fraudulent vote inflation in targeted locations.

Second, we consider the fact that differences in electoral fraud across polling centers may stem from variations in the composition and the behavior of voters and polling center officials at these centers. While we lack data on individual voters and polling officials, we leverage 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 a higher number of 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 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 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 studies how contextual factors affect electoral outcomes and the behavior of voters, our paper answers the question of how contextual factors impact fraud. To the best of our knowledge, this is the first paper studying whether contextual factors such as the features of the voting location (e.g., whether it is a school or a place of worship) can affect fraud.

2 Data

Our primary sources of data are information on the geolocation and names of polling places gathered by International Security Assistant Force (ISAF) inspection teams shortly after the election, as well as information different election fraud metrics at the polling center level gathered by a UN-sponsored audit following the 2009 election.

2.1 Measures of Fraud

We follow Weidmann and Callen [2013] and Gonzalez [2021] for our definitions of fraud during the 2009 Afghan presidential election. Specifically, our key measures of fraud are defined using the ECC fraud categories and the list of polling stations that were audited. The six categories that the ECC uses to initiate an audit are first combined into three more general categories: *Category A*, which includes stations with 600 or more votes cast; *Category B*, which includes stations where one candidate received 95 percent or more of the total votes cast; and *Category C*, which includes stations that meet categories A and B above. Consequently, the two fraud outcomes that we define are: (i) The percentage of votes cast by stations falling under *Category C* at the polling station level. In particular, if a polling station c has s stations overall, of which $n \leq s$ fall under *Category C*, 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 C*. The second is (ii), which indicates if there is at least one disqualified or *Category C* fraud station at polling center c.

A station's eligibility for one of the categories does not always indicate that fraud occurred there, even if the factors listed above are known as "measures of fraud." For example, there might be stations with exceptionally high voter turnout or significant partiality toward a particular candidate. In light of this, this measure should be interpreted as a substitute for fraud. However, considering the election's design and setting, these proxies offer an accurate indication of real fraud. For example, almost 96% of the ballots examined in stations that met the criteria for *Category C* fraud showed verifiable proof of tampering (Electoral Complaints Commission 2010) Additionally, in the same Afghan setting, these measures have been cross-validated with other fraud measures in the economics literature (Weidmann and Callen [2013]).

2.2 Classifying Polling Location Type

We classify polling center location type using rule-based and natural language processing (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, other school, university) 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 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.

2.3 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 selected polling places' accessibility and security condition 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 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. Out of the remaining 6,160 polling center observations, 5,904 (95.8%) had coordinates that were taken straight from the 2010 evaluation. The following coordinates were imputed for the remaining 256 centers: Six (0.1 percent) centers, simply used the coordinates of the district capital where the center was located, 169 (2.7 percent) used the centroid coordinates of the village or settlement where the center was located, and 81 (1.3 percent) used the coordinates of the center with the closest identifier code. We focus on Kabul province for our study because polling center classification was more accurate in Kabul—likely due to higher quality reporting of center names—and it had fewer ambiguous center names compared to other large provinces. This leaves us with 517 observations in the final sample.

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 makeup around the polling center's location. Between 2012 and 2013, the MISTI project (MISTIs 2013) gathered demographic information and physical coordinates for over 37,000 villages in Afghanistan from a variety of data sources. In the village nearest to the polling station, we use these data to generate variables that represent the population number and the languages spoken there ("Pashto", "Dari," and "Other").

3 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 whether there was at least one suspected fraudulent station in a polling center. See section 2 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 C 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, other. We leave mosque 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 occurring the year of the election and up until election day and indicators for accessibility of the center (accessible, limited, no accessible) obtained from the security assessment performed following the election. 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 versions of equation (1) that use 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 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, limiting the ability of officials or voters 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.

3.1 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 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 2, 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 sharply, while the probability of being assigned to a mosque falls to near zero.

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:

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

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

where $\mathbb{1}\{\text{Voters}_{id} > 3000\}$ 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 4.3, 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).

4 Results

4.1 Polling Center Location

Table 1 reports our estimated effects on election fraud. Columns 1-3 show the corresponding results for equation 1, which measures the share of fraudulent votes in a polling center. Column 1 is a baseline estimate; we add polling center and demographic controls in column 2 and district fixed effect in column 3. Our estimates suggest that the use of a school as a polling center is associated with a decrease in fraud. As a whole, the negative sign on each coefficient in Table 1 indicates that schools, houses, and other polling locations experience lower levels of fraud than Mosques, the comparison category. With the exception of the coefficient on the "other" category shown in column 2, which is estimated with controls and without district fixed effects, the coefficients reported for the "house" and "other" categories are not statistically significant at any level. In contrast, the estimated effects of the "school" category are large in magnitude relative to the control mean and highly statistically significant. Columns 1 and 2 show results for our OLS specifications first without and then including controls: both indicate a 7.3 percentage point reduction in the share of fraudulent votes when polling centers are located in a school. Both estimates are significant at the 1% level. We include district fixed effects in the specification reported in column 3, which shows school-based polling centers reducing the share of fraud by 5.3 percentage points, significant at the 5% level. Overall, these results suggest that there are particular aspects/features of schools that discourage or limit opportunities for fraud, rather than the alternative explanation of there being a characteristic of mosques that contributes to more fraud.

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 at schools are associated with a lower share of fraudulent votes compared to 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 those in mosques.

Columns (5) - (6) of Table 1 report estimated effects on the probability of Category C fraud in a polling center based on equations **??** and **??**, described in Section 3. The estimates are comparable to the results obtained 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 4 suggests that the probability of fraud is reduced by 8.2% 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 -0.086, meaning that a polling center being in a school decreases the probability of fraud by 8.63 percentage points.

4.2 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 1-3 and the probability of at least one station displaying category C fraud within a center as the outcome in columns 4-6. Our estimates suggest that all types of schools contribute to the reduced levels of fraud seen 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 4. The signs of the coefficients across each equation specification and school category 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%. Columns 3, 4, and 6 report a loss of statistical significance, especially on 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.[2]

4.3 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).

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

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 A3 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.

5 Mechanisms

5.1 Election-related Violence

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 explore whether violence played a part during the 2009 election by looking at how violence near polling centers changes after 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 prior to announcement to one week after announcement, i.e., election day) within a radius of 5 kilometers 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:

$$V_{idt} = \beta_0 + \sum_{\substack{j=-8 \\ j\neq -1}}^{7} \beta_j \mathbf{1}(t=j) + \mathbf{X}_i\theta + \gamma_d + \varepsilon_{idt} \tag{4}$$

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 prior to the announcement) as the reference category. $\mathbf{X}_i$ is a vector of controls at the polling center level.[3] γ_d is a district fixed effect. ε_{it} is the error term, clustered at the district level. [4]

Figure 3 shows that attacks go up right after 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.

$$\mathrm{V}_{idt} = \beta_0 + \sum_{\substack{j=-8 \\ j\neq -1}}^{7} \beta_j \mathbf{1}(t=j)\cdot \mathrm{School}_i + \sum_{\substack{j=-8 \\ j\neq -1}}^{7} \delta_j \mathbf{1}(t=j)\cdot \mathrm{Other}_i + \mathbf{X}_i\theta + \gamma_d + \lambda_t + \varepsilon_{idt} \tag{5}$$

where all terms are defined as in equation (4) 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. [5]

Figure 4 plots the estimates of β_j for all attacks and for IED-specific attacks. We find that,

[3] 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).

[4] Refer to Table A8 in the Appendix where we present the results for equation 4.

[5] Refer to Table A9 in the Appendix where we present the results for equation 5.

following the announcement of polling center locations, schools are less likely to be targeted by the Taliban relative to mosques (the same holds for other polling center types). This can explain the differential results on fraud. More attacks in mosques (less attacks in schools) implies turnout will be more affected in mosques. Therefore, officials are likely to substitute the affected turnout with fraudulent votes.

These findings provide empirical support for the violence channel. The sharp increase in attacks following the announcement of polling centers aligns with the idea that targeted violence by insurgents was influential in shaping electoral results. Moreover, the differential impact of violence across polling centers is consistent with a fraud-compensation mechanism: since attacks disproportionately targeted mosques, these locations likely experiences lower voter turnout, and thus higher levels of electoral fraud to substitute low turnout.

5.2 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 good voters or more honest individuals decide to vote in schools or manage polling centers in schools, then this can explain why fraud levels tend to be lower in schools relative to 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 give some idea of 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).[6] 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 save costs, but disagreements led to the different dates [Faiez, 2008]. In spite of this caveat, we believe that if polling center type drives selection of officials, voters, or community engagement, it will hold in similar elections. Therefore, insights are still relevant to our results even if they are from a different but quite related election.

Table 4 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),

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

the number of complaints filed by women (column 2), 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 4** 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 relative to 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 relative to 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 is 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: they are more likely to raise complaints, 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 relative to their mosque counterparts. Instead, candidates are more likely to be accused in schools.

6 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.

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.

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

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.

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.

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

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.

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

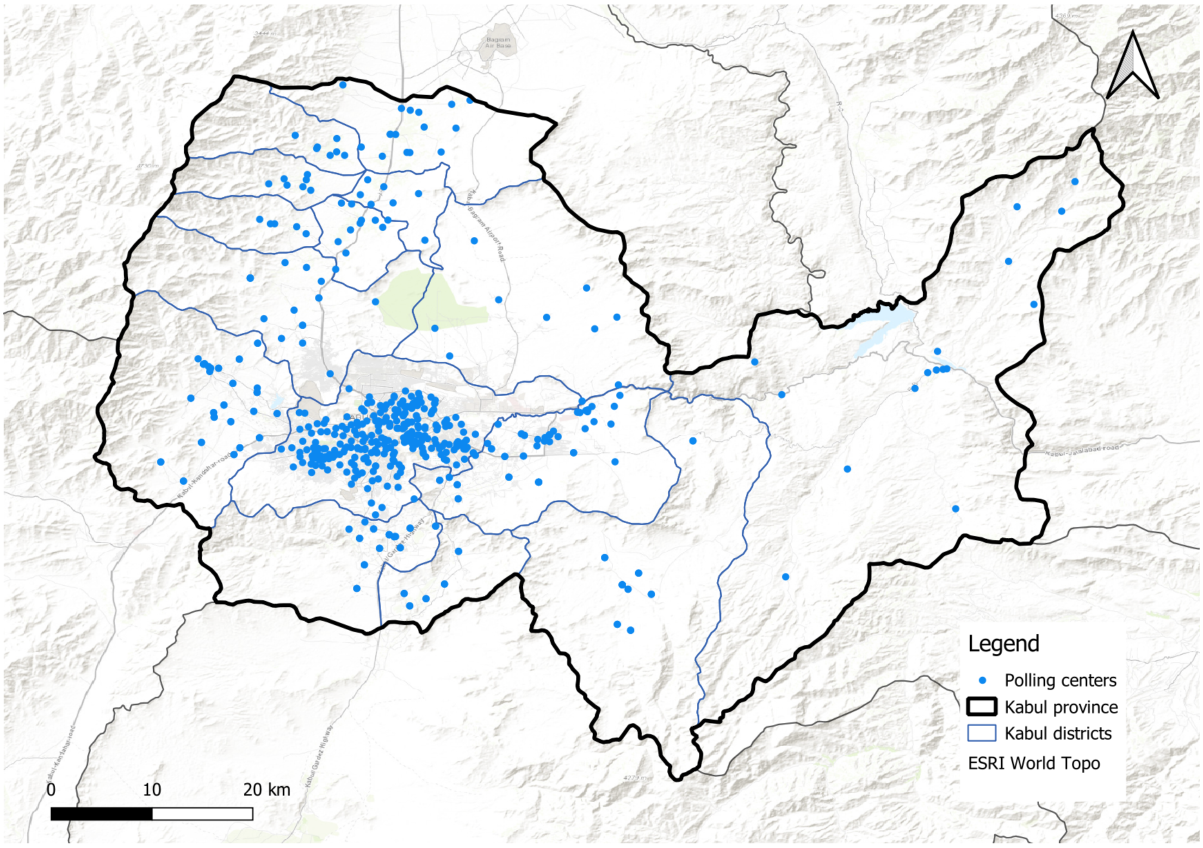


Figure 1: Polling Centers, Kabul Province

Notes: Polling centers within Kabul province.

Tables and Figures

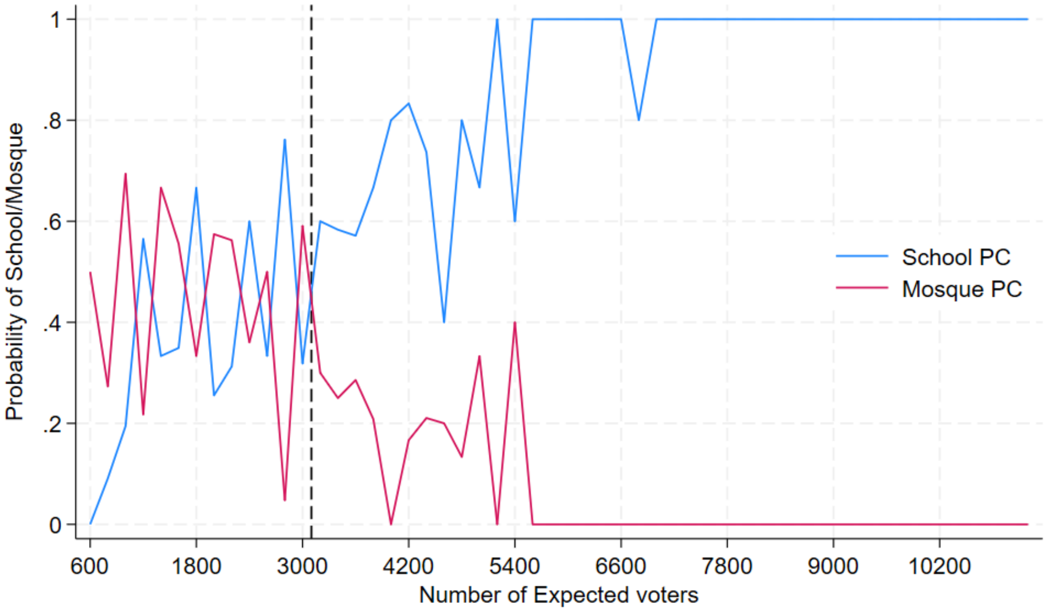


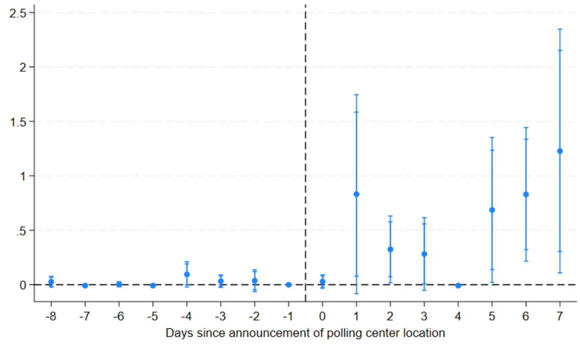
Figure 2: Probability of Polling Location and Number of Expected voters

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.

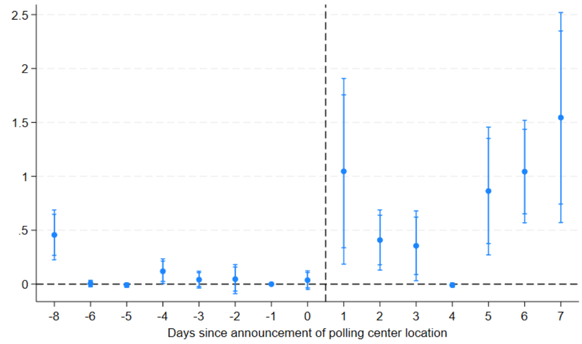
Table 1: Polling Center Type and Election Fraud

	Share of fraudulent votes (β)				$\mathbb{1}\{$Fraudulent stations $> 0\}$ (dy/dx)	
	(1)	(2)	(3)	(4)	(5)	(6)
Schools	-0.073***	-0.072***	-0.053**	-1.208***	-0.081**	-0.086***
	(0.019)	(0.021)	(0.019)	(0.297)	(0.028)	(0.017)
House	-0.031	-0.063	-0.034	-0.785	-0.045	-0.065
	(0.054)	(0.044)	(0.046)	(0.631)	(0.063)	(0.039)
Other	-0.026	-0.049*	-0.031	-0.514	-0.032	-0.029
	(0.038)	(0.027)	(0.029)	(0.553)	(0.059)	(0.058)
Control mean	0.091	0.091	0.092	0.091	0.144	0.143
Observations	512	512	511	512	511	512
Districts	18	18	17	18	17	18
Model	OLS	OLS	OLS	Poisson	OLS	Probit
Controls	No	Yes	Yes	Yes	Yes	Yes
District FE	No	No	Yes	Yes	Yes	No

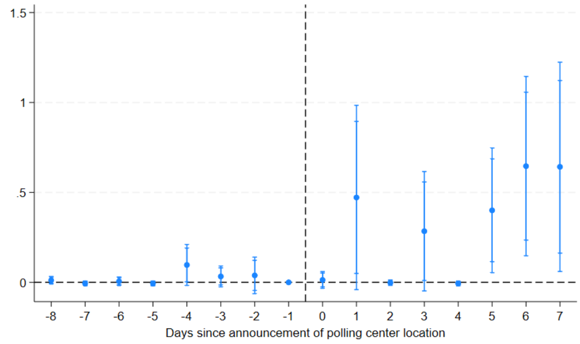
Notes: Columns 1-3 present the results of the OLS model specified in equation 1, while Column 4 presents the results of the Poisson model specified in equation **??**, using the share of fraudulent votes within the center as the outcome. Columns 5 and 6 present the results of the linear probability model specified in equation **??** and the Probit model specified in equation **??**, 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 -0.666 with standard error (0.133) 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.



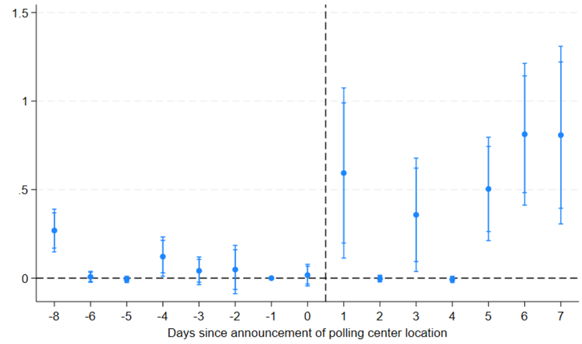
(a) All attacks



(b) All attacks + Controls



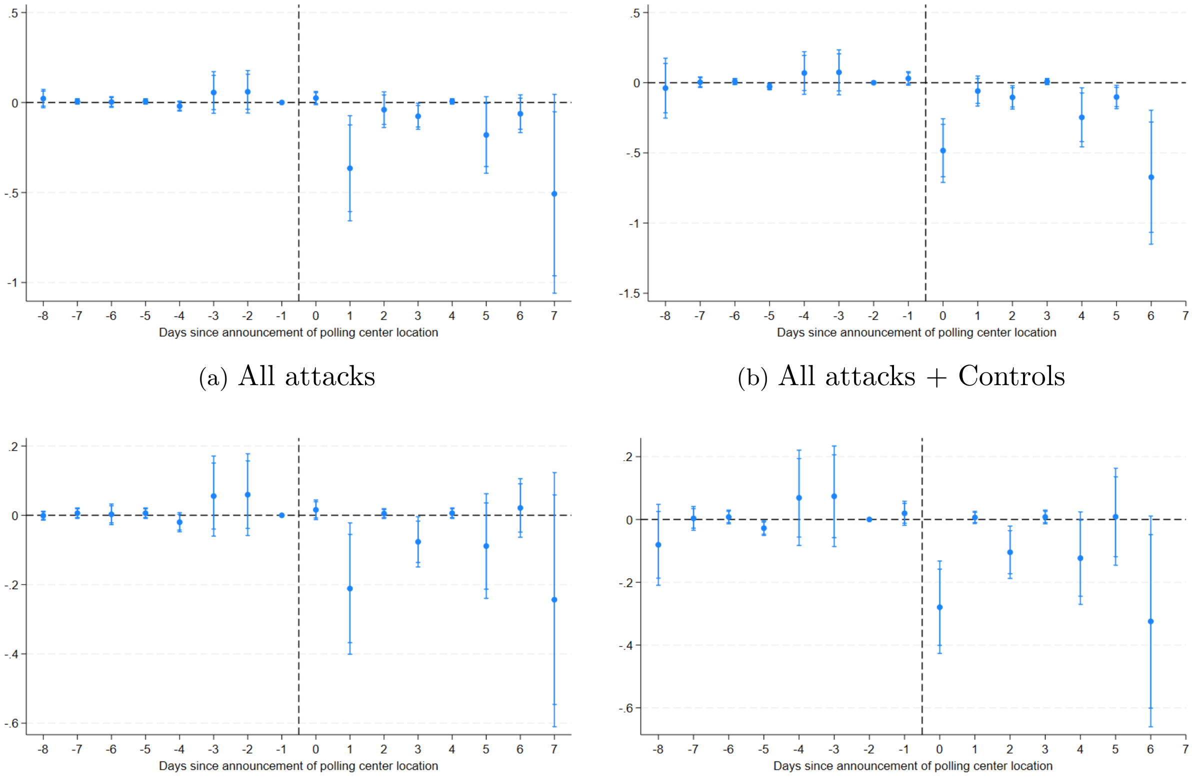
(c) Number of IEDs



(d) Number of IEDs + Controls

Figure 3: Violent Attacks after Announcement of Polling Center Location

Notes: Each dot provides an estimate of coefficient β_j from regression (4). Panels a and b use number of all violent events within a 5-km radius of a polling center as outcome. Panels c and d restricts 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 4: Violent Attacks after Announcement of Polling Center Location, Schools relative to Mosques

Notes: Each dot provides an estimate of coefficient β_j from regression (4). Panels a and b use number of all violent events within a 5-km radius of a polling center as outcome. Panels c and d restricts 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 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	-0.062**	-0.076***	-0.043	-0.543	-0.067	-0.079***
	(0.024)	(0.020)	(0.027)	(0.697)	(0.048)	(0.029)
High	-0.067***	-0.059**	-0.043*	-0.854**	-0.064**	-0.063**
	(0.020)	(0.028)	(0.023)	(0.398)	(0.030)	(0.028)
Secondary	-0.087***	-0.082***	-0.069***	-2.934***	-0.108***	-0.109***
	(0.024)	(0.022)	(0.021)	(0.774)	(0.033)	(0.023)
University	-0.091***	-0.073***	-0.055***	-18.200***	-0.099***	0.00
	(0.023)	(0.011)	(0.014)	(1.002)	(0.020)	(.)
House	-0.031	-0.063	-0.034	-0.841	-0.044	-0.64*
	(0.055)	(0.044)	(0.046)	(0.665)	(0.062)	(0.037)
Other	-0.026	-0.048*	-0.030	-0.574	-0.031	-0.028
	(0.038)	(0.027)	(0.029)	(0.497)	(0.061)	(0.058)
Control mean	0.091	0.091	0.092	0.091	0.144	0.143
Observations	512	512	511	512	511	505
Districts	18	18	17	18	17	18
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, Columns 1-3 present the results of the OLS model specified in equation 1, while Column 4 presents the results of the Poisson model specified in equation **??**, using the share of fraudulent votes within the center as the outcome. Columns 5 and 6 present the results of the linear probability model specified in equation **??** and the Probit model specified in equation **??**, 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 using equation **??**.
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. 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	0.415***	0.505***	0.532***	0.315***	0.353***
	(0.119)	(0.097)	(0.088)	(0.090)	(0.088)
Panel B: Share of fraudulent votes (Second Stage)					
School	-0.103***	-0.102***	-0.102***	-0.154**	-0.171**
	(0.025)	(0.019)	(0.020)	(0.062)	(0.060)
Control mean	0.091	0.091	0.092	0.110	0.110
Observations	445	445	444	223	223
Clusters	18	18	17	16	16
K-P F-stat	13.1	28.3	37.4	12.4	16.0
AR p-val	0.014	0.000	0.000	0.010	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. 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

Table 4: Polling Center Type and Complaints

		Complainant		Respondent	
	Complaints	Female	Individual	Polling official	Candidate
	(1)	(2)	(3)	(4)	(5)
Panel A: Number of complaints					
Schools	0.536**	0.821***	0.502**	0.115	0.829**
	(0.231)	(0.174)	(0.206)	(0.198)	(0.370)
House	-0.017	1.841**	-14.033***	0.160	0.023
	(0.269)	(0.808)	(1.047)	(0.306)	(1.005)
Other	0.002	0.632**	0.684***	-0.231***	0.490
	(0.217)	(0.305)	(0.265)	(0.085)	(0.449)
Control mean	0.561	0.224	0.224	1.239	0.149
Observations	517	236	236	236	236
Districts	18	17	17	17	17
Panel B: Share of complaints					
Schools		0.516	0.678***	-0.187***	0.770***
		(0.376)	(0.248)	(0.053)	(0.275)
House		1.689**	-11.945***	0.275	-0.093
		(0.815)	(0.946)	(0.349)	(0.797)
Other		0.581	1.115***	-0.060	0.257
		(0.552)	(0.204)	(0.077)	(0.249)
Control mean		0.131	0.095	0.738	0.092
Observations		236	236	236	236
Districts		17	17	17	17

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 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 frequently identified across different categories of polling center locations. All models are estimated using a Poisson specification as in equation **??** 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

Appendix A Additional Figures and Tables

Table A1: Polling Center Type and Election Fraud: Exact Sample Sizes

	Share of fraudulent votes (β)				$\mathbb{1}\{$Fraudulent stations $> 0\}$ (dy/dx)	
	(1)	(2)	(3)	(4)	(5)	(6)
Schools	-0.091**	-0.076**	-0.057*	-1.208***	-0.089*	-0.764***
	(0.034)	(0.027)	(0.026)	(0.308)	(0.038)	(0.200)
House	0.002	-0.075	-0.027	-0.785	-0.033	-0.748
	(0.099)	(0.089)	(0.088)	(0.656)	(0.124)	(0.503)
Other	-0.038	-0.049	-0.033	-0.514	-0.035	-0.320
	(0.039)	(0.030)	(0.033)	(0.574)	(0.070)	(0.344)
Control mean	0.113	0.113	0.113	0.177	0.177	0.177
Observations	413	413	413	413	413	413
Districts	8	8	8	8	8	8
Model	OLS	OLS	OLS	Poisson	OLS	Probit
Controls	No	Yes	Yes	Yes	Yes	Yes
District FE	No	No	Yes	Yes	Yes	No

Notes: The table shows the results of the main analysis when the sample is restricted to be identical across models. Columns 1-3 present the results of the OLS model specified in equation 1, while Column 4 presents the results of the Poisson model specified in equation **??**, using the share of fraudulent votes within the center as the outcome. Columns 5 and 6 present the results of the linear probability model specified in equation **??** and the Probit model specified in equation **??**, 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 -0.764 with a standard error (0.200) for schools. See table 1 for details about the controls used in columns 2-6.

Table A2: Polling Center Type and number of stations with 600+ votes

	(1)	(2)	(3)	(4)
Schools	-0.268**	-0.346***	-0.211***	-1.284***
	(0.096)	(0.111)	(0.068)	(0.258)
House	-0.146	-0.316	-0.108	-0.671
	(0.279)	(0.227)	(0.196)	(0.621)
Other	-0.159	-0.301**	-0.151	-0.581
	(0.134)	(0.137)	(0.095)	(0.640)
Control mean	0.364	0.364	0.365	0.364
Observations	517	517	516	517
Districts	18	18	17	18
Model	OLS	OLS	OLS	Poisson
Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes

Notes: Columns 1-3 present the results of equation 1, while Column 4 presents the results of equation **??**, using 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). Column 3 loses one observation due to a singleton district. Standard errors clustered at district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table A3: 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 A4: Polling Center Type (Disaggregated) and Violence

	Number of Attacks			$\mathbb{1}\{\text{Attacks} > 0\}$	
	(1)	(2)	(3)	(4)	(5)
Primary	-0.262	0.338***	0.301***	-0.209	0.077
	(0.171)	(0.117)	(0.107)	(0.206)	(0.054)
High	-0.040	-0.144	-0.102	-0.173*	-0.071***
	(0.100)	(0.093)	(0.081)	(0.090)	(0.026)
Secondary	-0.365***	-0.196**	-0.165**	-0.311***	-0.078**
	(0.111)	(0.077)	(0.069)	(0.112)	(0.033)
University	-14.876***	-14.657***	-18.385***	0.000	0.000
	(1.149)	(0.954)	(1.014)	(.)	(.)
House	-0.561	-0.075	-0.027	-0.098	0.091
	(0.534)	(0.571)	(0.541)	(0.461)	(0.117)
Other	0.158	-0.010	-0.019	-0.111	-0.090*
	(0.201)	(0.106)	(0.100)	(0.177)	(0.052)
Control mean	0.990	0.990	0.990	0.429	0.431
Observations	517	517	517	510	509
Districts	18	18	18	18	18
Model	Poisson	Poisson	Poisson	Probit	Probit
Controls	No	Yes	Yes	No	Yes
District FE	No	No	Yes	No	No

Notes: Columns 1-3 present the results of equation **??** using the number of violent attacks at a polling center location in the year leading up to the election as the outcome. Columns 4 and 5 present the results of equations **??** and **??**, respectively, using an indicator for whether there was at least one attack at the polling center location in the year leading up to the election. he results in columns 4 and 5 are reported as marginal effects, the discrete change from the base level. The z-scores of the probit model (5) are 0.276 with standard error (0.192) for primary schools, -0.261 (0.093) for high school, and -0.289 (0.126) for secondary schools. There are too few university observations to estimate effects for this category using equation **??**.
Columns 2-5 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 4 and 5 omit language, and access due to no variation within some districts. Columns 4 excludes 7 observations due to data separation as the 'University' category was dropped for perfectly predicting failure. Columns 5 excludes 8 observations due to data separation as the 'University' and 'Other' language categories were dropped for perfectly predicting failure. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table A5: 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	-0.032***	-0.028***	-0.025***	0.004	0.000	0.006
	(0.007)	(0.007)	(0.006)	(0.022)	(0.020)	(0.016)
House	0.028	0.019	0.015	0.097	0.038	0.057
	(0.029)	(0.025)	(0.028)	(0.075)	(0.068)	(0.059)
Other	-0.015	-0.012	-0.010	0.047	0.053	0.068*
	(0.021)	(0.020)	(0.021)	(0.041)	(0.033)	(0.038)
Control mean	0.091	0.091	0.092	0.091	0.091	0.092
Observations	517	517	516	517	517	516
Districts	18	18	17	18	18	17
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. Columns 3 and 6 lose one observation due to a singleton district. Standard errors clustered at district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table A6: Polling Center Type and Complaints (controlling for number of complaints)

	Complainant		Respondent	
	Female	Individual	Polling official	Candidate
	(1)	(2)	(3)	(4)
Schools	0.608**	0.343*	0.008	-0.101
	(0.257)	(0.176)	(0.043)	(0.118)
House	1.742**	-16.196***	-0.075	0.172
	(0.871)	(1.005)	(0.092)	(0.319)
Other	0.845**	1.041***	-0.166***	-0.130
	(0.329)	(0.250)	(0.064)	(0.098)
Control mean	0.224	0.224	1.433	1.239
Observations	236	236	236	236
Districts	17	17	17	17

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 as in equation **??** 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 A7: 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	0.337*	0.236	0.600***	-0.486**	0.607***
	(0.176)	(0.230)	(0.173)	(0.227)	(0.234)
House	0.110	1.175	0.000	0.078	-0.518
	(0.235)	(0.804)	(.)	(0.675)	(0.687)
Other	0.032	0.401	0.680**	0.117	0.405
	(0.130)	(0.542)	(0.278)	(0.213)	(0.323)
Control mean	0.340	0.246	0.153	0.787	0.148
Observations	507	204	190	224	224
Districts	15	10	9	11	11

Notes: This table examines the relationship between 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 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 as in equation **??** 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 A8: Violent attacks after announcement of polling center locations

	All Attacks		IED Attacks	
	(1)	(2)	(3)	(4)
8 Days Before	0.027	0.456***	0.012	0.269***
	(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)	(0.011)	(0.015)
5 Days Before	-0.008	-0.010	-0.006	-0.007
	(0.007)	(0.010)	(0.006)	(0.008)
4 Days Before	0.095	0.119**	0.097*	0.122**
	(0.055)	(0.054)	(0.054)	(0.052)
3 Days Before	0.033	0.041	0.033	0.041
	(0.027)	(0.037)	(0.027)	(0.037)
2 Days Before	0.037	0.046	0.039	0.049
	(0.048)	(0.064)	(0.048)	(0.064)
1 Day Before	0.000	0.000	0.000	0.000
	-	-	-	-
Day of Announcement	0.029	0.036	0.014	0.017
	(0.029)	(0.041)	(0.022)	(0.029)
1 Day After	0.832*	1.046**	0.472*	0.594**
	(0.433)	(0.406)	(0.243)	(0.227)
2 Days After	0.325**	0.409***	-0.002	-0.002
	(0.145)	(0.132)	(0.007)	(0.009)
3 Days After	0.282*	0.355**	0.284*	0.358**
	(0.158)	(0.153)	(0.157)	(0.151)
4 Days After	-0.008	-0.010	-0.006	-0.007
	(0.007)	(0.010)	(0.006)	(0.008)
5 Days After	0.687**	0.864***	0.400**	0.504***
	(0.315)	(0.280)	(0.164)	(0.138)
6 Days After	0.830**	1.044***	0.646**	0.813***
	(0.291)	(0.224)	(0.236)	(0.189)
Day of Election	1.228**	1.545***	0.642**	0.808***
	(0.531)	(0.460)	(0.276)	(0.237)
Prior-Year Mean	0.033	0.033	0.031	0.031
Observations	8272	5859	8272	5859
Districts	18	17	18	17
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 (4) that correspond with the point estimates in Figure 3. Columns 1 and 2 of this table correspond with panels a and b of Figure 3, using number of all violent events within a 5-km radius of a polling center as outcome. Columns 2 and 3 of this table correspond with panels c and d of Figure 3, 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 A9: 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	0.022	-0.039	-0.001	-0.081
	(0.024)	(0.101)	(0.006)	(0.061)
7 Days Before	0.006		0.006	
	(0.007)		(0.007)	
6 Days Before	0.003	0.003	0.003	0.003
	(0.014)	(0.018)	(0.014)	(0.018)
5 Days Before	0.006	0.008	0.006	0.008
	(0.007)	(0.010)	(0.007)	(0.010)
4 Days Before	-0.020	-0.028**	-0.020	-0.028**
	(0.013)	(0.011)	(0.013)	(0.011)
3 Days Before	0.056	0.069	0.056	0.069
	(0.055)	(0.072)	(0.055)	(0.072)
2 Days Before	0.060	0.074	0.060	0.074
	(0.056)	(0.076)	(0.056)	(0.076)
1 Days Before	0.000	0.000	0.000	0.000
	-	-	-	-
Day of Announcement	0.025	0.030	0.016	0.020
	(0.017)	(0.023)	(0.013)	(0.018)
1 Day After	-0.365**	-0.484***	-0.211**	-0.280***
	(0.138)	(0.107)	(0.090)	(0.069)
2 Days After	-0.040	-0.060	0.005	0.006
	(0.047)	(0.050)	(0.007)	(0.009)
3 Days After	-0.076**	-0.104**	-0.076**	-0.104**
	(0.034)	(0.039)	(0.034)	(0.039)
4 Days After	0.006	0.008	0.006	0.008
	(0.007)	(0.010)	(0.007)	(0.010)
5 Days After	-0.180*	-0.247**	-0.089	-0.123*
	(0.101)	(0.099)	(0.072)	(0.069)
6 Days After	-0.062	-0.102**	0.021	0.009
	(0.050)	(0.039)	(0.040)	(0.073)
Day of Election	-0.507*	-0.673***	-0.244	-0.325*
	(0.262)	(0.225)	(0.174)	(0.158)
Prior-Year Mean	0.033	0.033	0.031	0.031
Observations	8272	5859	8272	5859
Districts	18	17	18	17
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 (4) that correspond with the point estimates in Figure 4. Columns 1 and 2 of this table correspond with panels a and b of Figure 4, using number of all violent events within a 5-km radius of a polling center as outcome. Columns 2 and 3 of this table correspond with panels c and d of Figure 4, 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