Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 26-1401

Capturing Voter Turnout at the School District Level: Validating a Geospatial Strategy

Adam Shepardson
University at Albany, SUNY

Melissa Arnold Lyon
University at Albany, SUNY

Beth E. Schueler
Stanford University

Joshua Bleiberg
University of Pittsburgh

School boards are critical sites of education policymaking, yet scholarship on these institutions is scarce because of severe data limitations. We introduce a geospatial strategy and open-source R package, called "Query, Overlay, Recover" (QOR), that generates high-quality estimates of voter turnout in school board elections overall and for voter subgroups. We describe this process and gather validity evidence using voter files from North Carolina and Washington (2013-2022). We illustrate the value of these data by addressing two substantive research questions that would otherwise not be answerable. We show voter turnout and the representativeness of voters of color are significantly lower in districts serving less advantaged students. The QOR process can now be used to generate district-level elections data nationwide.

VERSION: February 2026

Suggested citation: Shepardson, Adam, Melissa Arnold Lyon, Beth E. Schueler, and Joshua Bleiberg. (2026). Capturing Voter Turnout at the School District Level: Validating a Geospatial Strategy. (EdWorkingPaper: 26-1401). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/bmh1-xr32

Capturing Voter Turnout at the School District Level:
Validating a Geospatial Strategy

Adam Shepardson, University at Albany, SUNY
Melissa Arnold Lyon, University at Albany, SUNY
Beth E. Schueler, Stanford University
Joshua Bleiberg, University of Pittsburgh

Abstract: School boards are critical sites of education policymaking, yet scholarship on these institutions is scarce because of severe data limitations. We introduce a geospatial strategy and open-source R package, called "Query, Overlay, Recover" (QOR), that generates high-quality estimates of school board election voter turnout at the school district level, overall and for voter subgroups. We describe this process and gather validity evidence using voter files from North Carolina and Washington (2013-2022). We illustrate the value of these data by addressing two substantive research questions that would otherwise not be answerable. We show voter turnout and the representativeness of voters of color are significantly lower in districts serving less advantaged students. The QOR process can now be used to generate district-level elections data nationwide.

Keywords: Politics of Education, School Board Elections, Voter Turnout, Political Representation, Geospatial Methods

Note: The authors thank participants at the Association for Education Finance and Policy (AEFP) 2025 Annual Conference, as well as Vladimir Kogan, Karin Kitchens, Lucy Sorensen, Melissa Marschall, John Holbein, Jon Valant, and NaLette Brodnax for helpful feedback, and Kylie Cooper, Orobo E. Obaro-Ogbovoh, Erica Whitman and Chris Kunze, among others, for research assistance. This work has been funded by the William T. Grant Foundation.

Introduction

School boards are key actors responsible for significant policymaking activity in the United States' decentralized education space, yet limited evidence exists on these boards and the voters that select the board members who make policy. The evidence is particularly scarce regarding how voter participation in school board elections relates to the educational policies boards adopt and the academic outcomes school systems produce. These holes in the literature are in large part due to notable data limitations. There is no national data source on voter turnout in school board races, and subnational data is often not made public. Local election results are often reported at the county or candidate level rather than the school district level. This makes it difficult to study education policy-relevant questions given K-12 school policy is made at the school district rather than county level by school boards and the superintendents they hire and oversee. Therefore, the field currently has no credible estimate of average voter turnout in school board elections nationwide, nor a good way to study the relationships between voter participation and education policy or academic outcomes.

We solve this problem using a novel method we call "Query, Overlay, and Recover" (QOR) for which we link individual voter residential addresses to school districts and then generate school district level estimates of voter turnout in school board races. We use freely available data from roughly 75 million registered voters in 410 districts and 3,288 elections in North Carolina and Washington state from 2013 to 2022. We describe this process, show evidence of its validity, and provide an R package, *QOR*, that researchers can use to implement the QOR method themselves (Shepardson et al., 2026).[1] Using this method, we show that average voter turnout in school board elections is 46% in North Carolina and 41% in Washington state (see Table 1, further discussed in Results). There is also substantial variation by subgroups:

average turnout in school board elections for Hispanic voters in North Carolina, for example, is only 28.4%, whereas average turnout for white voters is 50%. Scholars can now use the QOR method to generate similar overall and subgroup estimates of voter turnout in school board elections across the United States.

The QOR method also allows researchers to study a host of questions related to the nature of public participation in local elections, the effects of education policies on voter turnout, the effects of electoral policies on participation, the effects of voter participation on who gets elected and what education policies they adopt, and much more. Indeed, we use the resulting data from our two states to answer two substantive research questions about the extent to which (1) school board election turnout varies depending on the demographic characteristics of the student populations served by the district and (2) local electorates voting in school board races are demographically representative of their broader student population. Our analysis reveals that school districts with higher proportions of economically disadvantaged students and students of color face a dual challenge: not only do these districts see lower overall voter participation, but they also demonstrate more pronounced racial disparities in electoral engagement. These results highlight important questions about equity in school governance representation and demonstrate the value of the QOR method for advancing scholarship, informing policy decisions, and improving practice.

Motivation

Why Study Voter Turnout in School Board Elections?

School board elections are consequential given the United States' tradition of local control over education. School boards are directly responsible for a wide range of policy decisions, impacting large numbers of citizens and stakeholder groups, including parents, teachers, children, homeowners, and taxpayers (Hess & Leal, 2005; Howell, 2005; Kirst & Wirt, 2009; Manna & McGuinn, 2013). Over 90% of the U.S. population attends public K-12 schools at some point in their lives. Board policies affect parents who rely on schools for their custodial functions, as well as taxpayers who contribute to the system and whose property values are impacted by the perceptions of local schools (Black, 1999). School boards have authority over hiring and firing superintendents, budgets, contracts, scheduling and curricular decisions, and operational decisions such as whether and when to re-open schools during crises such as a global pandemic (Blazar & Schueler, 2024). Boards have the capacity to determine whether and to what extent schools focus on building engaged citizens, preparing students for the workforce, and developing social cohesion. At least in part as a result of their policy influence, a non-trivial amount of variation in student outcomes resides at the school district level (Chingos et al., 2015).

Previous research further demonstrates that the composition of school boards (i.e., who serves on these boards) matters for the policies these boards adopt and the student outcomes school districts produce. For instance, descriptive representation (i.e., the extent to which elected officials demographically reflect their constituents) for citizens of color leads to the passage of education policies that benefit non-White students (Fischer, 2023; Grissom, 2010; Kogan et al., 2021a). Further research establishes that the characteristics of school board members beyond their race/ethnicity affect the policies these boards pursue, such as their partisanship (Macartney & Singleton, 2018) and teaching experience (Shi & Singleton, 2023).

The characteristics of the officials who are ultimately elected to office are in large part a function of who turns out to vote in elections. The literature establishes, for example, that more diverse electorates are more likely to prefer and elect more diverse candidates and that people of color tend to express more satisfaction with elected officials of their own race/ethnicity. Lublin and Wright (2024) use data from 33 states between 2011-2020 and show Asian Americans expressed a preference for legislators of Asian ethnicities. Arnold and Carnes (2012) conduct a time-series analysis of mayoral approval in New York City between 1984-2009 and find race-based preferences for candidates among voters. Burnett and Kogan (2022) also show that race predicts voters' judgement of candidates, as do other scholars (Hill et al., 2001; Kaufmann, 2003). Therefore, understanding who votes in school board elections is critical for understanding who gets elected as a board member and the education policies they pursue.

Even for those readers more interested in political dynamics broadly than in education policies or outcomes, boards are an important topic of study. The more than 80,000 school board members make up the largest group of elected officials in the U.S. (Berry, 2024). Additionally, elementary and secondary education accounts for roughly 40% of local government spending, and boards typically direct these funds (Urban Institute, 2023). School systems also are some of the largest local employers, whose salary policies and working conditions can affect the ecosystem of local education directly and indirectly (Henig et al., 2001; Lyon, 2023). School board elections therefore provide a valuable source of information for scholars of local policy and politics more broadly.

The limited evidence that does exist on the topic of voter participation in school board elections suggests a grim picture that motivates the need for attention to this area. Scholars have consistently found that turnout in school board races falls well short of our democratic ideals, often in the single digits or low teens, with some evidence suggesting turnout is 20 percentage points lower than state or national elections (Anzia, 2014; Berry & Gersen, 2011; Hajnal, 2010; Holbrook & Weinschenk, 2014; Kogan et al., 2018; Marschall & Lappie, 2018). Not only is turnout in board elections low overall but there are signs that the electorates who select these decision-makers are not representative of the broader communities these officials go on to serve. Evidence from a handful of states shows "considerable demographic differences between voters who participate in board elections and students attending the schools that boards oversee," such that White voters, older voters, and affluent voters are overrepresented; gaps are most pronounced in majority non-White communities (Kogan et al., 2021b). Turnout is especially low in places where school board elections are held "off-cycle"—in a different month or year—from national elections (Anzia, 2014; Dynes et al., 2021; Hajnal et al., 2022; Hartney, 2021). However, the data on which these understandings are built, and the field's understanding of school board elections, are not commensurate with the importance of these institutions.

What Are the Limitations of Existing Data Sources for Studying School Board Elections?

The data on voter turnout in school board elections are limited in several notable ways, particularly for those studying education policy. First, turnout rates and/or the data needed to calculate them in local elections—if and when they are reported publicly—are often reported at the county level rather than the school district level. This is an unfortunate limitation given that the school district, not the county, is typically the unit of interest for education policymaking. As described earlier, a substantial amount of variation in education policy occurs at the district level given the nation's decentralized approach to policymaking in the education space. The school district is also the unit for which school board members are elected, and the unit for which academic and other educational outcomes of interest are reported. Therefore, without data at the school district level, it is not possible for researchers or community advocates to link school board election data to other relevant sources of data that are reported out at the school district level, for example, data on the demographic characteristics or academic performance of districts, or data on variation in policy adoption across districts.

This lack of data on school board elections at the district level is a significant problem because, for the vast majority of school districts, district boundaries do not line up with county boundaries. We use the National Center for Education Statistics (NCES)'s school district shapefiles and Geographic Relationship Files, as well as the Census TIGER/LINE County shapefiles, to illustrate the extent of the problem. In Supplemental Table A1, we find that only 6.9% of districts nationally have boundaries that are the same as county boundaries. There are two sources of this misalignment: (1) school districts crossing into more than one county and (2) multiple school districts within the same county. Figure 1 visually shows that only four states have perfect alignment between county and school district boundaries, and in a majority of states, 0% of districts have boundaries that perfectly align with counties. Furthermore, the states where there is greater alignment between districts and counties are not representative of states in the U.S. In particular, states with alignment tend to be located in the South where countywide districts are more common. Even in the South, only 25.51% of districts exactly match counties. This means that findings from studies for which researchers have been able to calculate school board election statistics from county-level data may not generalize to the country as a whole.

In part because of the challenge of calculating information about school board elections at the school district level, there is no comprehensive national data source on school board election turnout. Several datasets cover multiple states, but each come with meaningful limitations for researchers studying school board elections. In some cases, data are only available in even numbered years or selected years, such as in the case of the American National Election Studies (2024), the Current Population Survey Voting Supplement (Census Bureau, 2024), the General Social Survey (NORC, 2024), the MIT Elections Data and Science Lab (2025), and the Vote Database (O'Donnell, 2025). This is a particularly problematic limitation for scholars of school boards because many school board elections are held “off-cycle”—in different years or months from national elections held in November of even-numbered years—and election timing is one of the strongest predictors of voter turnout, with turnout rates substantially depressed in off-cycle races (Anzia, 2014; Dynes et al., 2021; Hajnal et al., 2022; Hartney, 2021).

In other cases, data are available for some local, state, or federal elections but not for school board races. For example, the Cooperative Congressional Election Study, the U.S. Elections Project focuses on national races, as do others (Barber & Holbein, 2022), whereas the Local Elections in America Project (Marshall & Ruhil, 2011) covers a variety of local races but does not include comprehensive data on school boards, as do others (Hajnal et al., 2022). Other sources of data that researchers have used to study turnout in school board races are not publicly available or can only be obtained for a fee, such as the L2 Voter File Data, which does not include identifiers that would allow for the calculation of turnout rates at the school district level (L2, 2025). Other data sources provide only a subset of states or subset of districts within a state, such as the American Local Government Elections Database (de Benedictis-Kessner et al., 2023), among others (Kogan et al., 2025).

In still other cases, the data are available only at the candidate level, rather than the election level, such as for the American Local Government Elections Database. This poses a problem in the many instances where voters can cast votes for more than one candidate for school board in the same election. When there are two or more open seats on the school board, voters can typically cast votes for multiple candidates. In these situations, the number of votes cast are much greater than the number of voters who turned out to vote on a given day, making it difficult to accurately calculate a voter turnout rate.

Additionally, it is not possible to use candidate-level data to estimate an accurate measure of turnout in districts that have ward-based elections. Ward-based election systems restrict voters' candidate choices based on sub-district areas (i.e., wards). School district wards are geographically coherent sub-districts where voters are represented by a smaller number of school board members than the sum of all those who make up the entire school board. These governance structures pose severe challenges when calculating voter turnout from official election returns, rather than individual voter files, for three main reasons. In brief: (1) many (though not all) wards prevent their residents from voting in other wards' contests; (2) turnout rates or registered voter counts (that would allow users to calculate a turnout rate) are almost never estimated at the ward level; and (3) wards may sometimes elect their representatives on different schedules than either at-large seats or other wards. Ward-based systems make it impossible to estimate the denominator in the school board turnout equation using candidate-level results, because researchers do not observe the percentage of total registered voters in each ward and eligible to vote on a specific date. Prior efforts at creating school board election datasets using election returns have therefore simply dropped ward based elections (Kogan et al., 2021b).

Other sources of data that may seem promising also come with limitations. Data from Ballotpedia are often crowdsourced, not easily downloadable, and do not include all states, districts, or years (Jacob, 2024). Some school board member survey data are based on members' self-reported estimates of turnouts and collected only in a single year (Hess, 2002). The A New Nation Votes resource is limited to pre-civil war elections, not allowing for research on more contemporary education policy.

Researchers have also used subnational sources of data to study school board election turnout. However, many of these sources are not publicly available, either focus on a limited number of years or states or districts (Anzia, 2012, 2014; Berry & Gersen, 2011; Holbein & Hillygus, 2016; Kitchens, 2021, 2023; Moe, 2006, 2011; Payson, 2017), rely on intensive manual data collection efforts that would be very resource-intensive to replicate annually on a large scale (Anzia, 2012; Carlson et al., 2025), capture related constructs but not voter turnout in school board races (e.g., school bond referenda), or use outdated years and/or provide limited information on the methods for capturing turnout (Kohfeld & Sprague, 2002). A method that would allow for calculating information about school board elections at the school district level would facilitate the creation of a national dataset on school board elections that could allow the field to learn about these races and their relationship to education policies and outcomes.

A final limitation of many existing sources of data on school board elections is that it is rarely possible to get information on voter turnout for subgroups of voters. This makes it impossible to—for example—compare turnout rates across different groups (e.g., Are older citizens more likely to participate in school board races than younger voters? Are Democratic voters more likely to vote in school board races than Republican voters? What about potential differences across racial or ethnic groups?). To the extent that some groups are more likely to be represented among school board voters than others, this has important implications for studying questions with equity implications and understanding possible sources of political and educational inequalities. Subgroup statistics would also allow for a deeper understanding of the mechanisms through which other dynamics operate. For example, if higher levels of voter turnout in school board races result in higher levels of educational spending on non-White students (e.g., Kogan et al., 2021a), subgroup results would help researchers understand whether this is due to higher turnout rates among non-White voters.

These limitations make it impossible to study a range of important policy-relevant questions that fall into at least the following four categories: (1) How do school board elections vary depending on school district characteristics? (e.g., Do districts serving more economically disadvantaged student populations have lower voter participation in school board races? Is voter turnout higher in districts that are higher achieving academically?) (2) How do school district educational policies impact school board elections? (e.g., Do school closures increase voter turnout in school board elections? What about integration efforts, teacher pay schemes, book bans, COVID school closures?) (3) How do electoral policies impact school district policies and/or the performance of school systems? (e.g., Does holding school board elections on-cycle with national elections increase educational spending and/or the academic achievement of students in the district?) (4) How does voter turnout in school board races impact school district policies and/or academic achievement? (e.g., Do higher levels of voter participation increase teacher salaries and/or improve student achievement outcomes?). With the availability of longitudinal district level election data, researchers could address these questions both correlationally and potentially causally if combined with data on district level variation in policies and/or performance over time. Unfortunately, without a comprehensive dataset of school board election turnout, it is not just difficult but sometimes impossible to address these types of research questions.

What Are the Contributions of This Paper?

We introduce a new process we call, "Query, Overlay, Recover (QOR)" that allows us to address three research questions: (1) How can researchers accurately measure voter turnout in school board elections at the school district level, both overall and for subgroups of voters? (2) Do turnout rates in school board elections vary depending on the demographic characteristics of the student populations served by the districts? (3) Does the demographic representativeness of the electorates in school board races vary depending on the demographic characteristics of the student populations served by the districts?

To answer RQ1, we present and describe the new QOR method for linking individual voters to school districts, resulting in data on voter turnout calculated both at the school district level overall and for subgroups of voters within districts. We implement the QOR method in two states, one of which has a high degree of overlap between county and school district boundaries (North Carolina) and another with limited alignment between county and school district lines (Washington). We validate our method by comparing our resulting data to other credible sources, demonstrating that it improves upon other options and replicating well-established substantive findings from prior literature. Though we use a two-state panel, researchers can easily adapt our process for any other state where individual level voter files are available or with proprietary voter data (e.g., Catalist, L2), either for a subset of states or all states.

The QOR method allows for the creation of a data source on school district-level voter turnout for a national census of school board elections that can be linked to a host of other data sources on school district characteristics and policies. It also provides the ability to calculate voter turnout rates for subgroups of voters. While we demonstrate the process using voter turnout, the QOR method we developed could also be used to connect any data with individual addresses to corresponding school districts. These data open the possibility of addressing a host of research questions related to the relationships between school board elections and district characteristics and/or public policies.

We then demonstrate the value of the QOR method and resulting data by addressing two substantive research questions (RQs 2 and 3) that would not be answerable without the kind of data produced by the QOR method. To address these questions, we link our QOR data on school board votes cast and voter turnout in both North Carolina and Washington with NCES data on school district characteristics. We find that average school board election turnout and the representativeness of voters of color declines as districts serve more low-income and non-White students. This analysis illustrates one example of how our approach opens new opportunities to study questions about the relationship between political participation at the local level and educational inequality.

Study Contexts

The data for this paper are drawn from North Carolina and Washington state—two states we have purposefully sampled for gathering evidence of validity for the QOR method. Both states produce comprehensive individual level voter files and yearly voter registration files that include voter's residential addresses and are either freely available to anyone at any time (i.e., North Carolina; North Carolina Board of Elections, 2025) or easily accessible and no-cost to researchers (i.e., Washington; Washington Department of State, 2025). As we detail below, the two states also offer useful variation in their school district structures, election systems, geography, political leaning, and demographics.

Data from North Carolina have several important strengths. First, in North Carolina, there is a high degree of overlap between school districts and counties (100 out of North Carolina's 115 school districts are county-based). Of these county school districts, only 12 have a city school district within them, meaning that county turnout is synonymous with at-large (not necessarily ward-based) school board election turnout in 88 districts (76%). The other 15 districts assign individuals to school districts based on their city. Additionally, North Carolina consistently produces and publishes county election returns for their school board elections.

These election returns include both county-level turnout and candidate-level vote counts, meaning that for many races, we have verified records on turnout and the number of votes cast in school board elections (the numerator in turnout). This allows us to directly compare QOR estimates to those based on the publicly reported county level returns (we refer to these as benchmark estimates) for some races. Second, the freely accessible voter registration files provide rich demographic information on individual voters in terms of their age, gender, race, and partisan affiliation. Third, North Carolina does not mandate a specific day for school board elections, so there is considerable within-state and within-district variation in election timing. Fourth, North Carolina is a demographically diverse, politically competitive Southeastern swing state that often leans Republican and allows districts to hold partisan school board elections.

Election data from Washington state are a helpful complement to North Carolina. First, Washington offers a useful contrast to North Carolina in terms of school district construction because zero Washington districts are county-based, and the school districts are much smaller, on average, than North Carolina districts. This allows us to test our method in two distinct contexts that reflect the diversity of U.S. school district construction. Importantly Washington is the more typical context in terms of the lack of alignment between county and school district boundaries, and this is the context where the QOR method is most valuable. In addition, like in North Carolina, the public data are freely accessible. Washington state employees provided us with yearly voter registration and vote history information upon request. These files contain demographic information on Washington voters in terms of their age and gender, but not race and partisan affiliation, as was the case in North Carolina. Third, Washington mandates a specific day for school board elections, which are all non-partisan, offering another useful contrast to North

Carolina. Fourth, Washington state is less racially diverse and more politically liberal than North Carolina and is situated in the western region of the U.S., demonstrating generalizability.

Data and Methods

We develop and undertake a three-stage geospatial strategy for transforming individual level voter files into functional, school district level election datasets. The three stages are Query, Overlay, and Recover (QOR). First, we *Query* a Census geocoding tool to turn individual voter residential addresses into geocoded coordinates. In stage 2, we *Overlay* the geocoded coordinates (converted into point geometries) on top of school district shapefiles, which allows us to assign voters an accurate NCES identification number for their local school district (NCES, 2024). Finally, we *Recover* unlocated voters and match them to school districts using their registration zip codes. Below, we describe the data we use for this analysis, then the methods for each of the three steps of the QOR method, and finally the methods used to calculate voter turnout.

Data Sources

Statewide Individual Level Voter Data

We obtain publicly available statewide voter registration records and voter history files for North Carolina and Washington state (North Carolina Board of Elections, 2025; Washington Department of State, 2025). Our raw data encompass all individuals registered to vote during every school board election year between 2013-2022 (even and odd years in North Carolina and odd years only in Washington state). We merge voter registration files containing voter addresses and individual characteristics with vote history files containing dates on which these individuals voted using unique voter identification numbers in each year for each state. The merged file details individual voter-by-election date records that flag participation in any state, federal, or local election that occurred during a specific year.

Unfortunately, these data offer no consistent indication of which election dates correspond to school board races. We therefore develop original data on school board election dates according to each state's unique election schedule. In North Carolina, we manually collect school board election dates for its 115 public school districts from the state Board of Elections website, reducing the final list to all consequential school board elections (North Carolina Board of Elections, 2025).[2] North Carolina's consequential school board elections occur in odd and even years 2013 through 2022 in our panel. Washington employs uniform, odd-year November school board general elections across districts, resulting in a second statewide panel covering 2013 through 2021. Given the uniformity of timing for general elections, we include the August primary elections in some analyses related to election timing.

Importantly, voter registration files (which we use as the denominator for turnout rates) are regularly updated and must be pulled at specific "snapshot dates." We choose snapshot dates tailored to each state, given differences in data availability across states. Specifically, Washington only allows access to its voter registration snapshots via direct requests. Metadata sent to us from the Washington state government alongside some of the files suggested that snapshot dates were collected at the end of November each year. We use registration data from shortly after the election to capture any last-minute registrations leading up to school board contests. In North Carolina, we similarly select the date that occurs after the relevant elections: January 1st of the following year because it is after all school board elections from the previous year.[3]

Using the voter registration files and original election date lists, we retain all voters on each school board election date who can vote without (re)registering according to state election laws. This intuitive restriction retains North Carolina voters coded by state officials as active, inactive, or temporary (military) voters given that even the inactive voters can all become active again merely by showing up to vote (North Carolina Legislature, 2023). We only retain voters with an active registration status in the Washington data, as the state utilizes mail elections and does not send inactive voters a ballot (Washington Legislature, 2011).

Across both states, the merged voter registration and vote history files tie every school board election date on which individuals voted to their registration status, residential home address (i.e., city, county, street, zip code), official precinct codes, age, and gender. Home addresses are split across several string and numeric variables (e.g., house numbers, street names, apartment numbers) that we combine into coherent street addresses following U.S. Postal Service conventions. Precinct codes contain two components, which we treat as a unique precinct identifier string when combined. The voter ages in North Carolina are specified as integer values, whereas we manually construct Washington voter ages from differences between registered birthdates and school board election dates. Gender categories in both states overwhelmingly feature Male or Female responses but officials record some voters as unknown or other.

North Carolina's voter files additionally contain individuals' party affiliation and race/ethnicity. Here, voters could register as Republican, Democrat, Unaffiliated, or Libertarian in all years, and could select White, Black, Asian, and several other race designations, as well as indicate Hispanic ethnicity. Our race/ethnicity categories code all race designations as mutually exclusive with Hispanic ethnicity, collapse Asian and Pacific Islander categories together, and pool the remaining set of smaller race categories as Other Race because in the 2020 Census, over 80% of North Carolinians were either White non-Hispanic or Black non-Hispanic, and Hispanic was the next largest group (America Counts Staff, 2021).

Spatial and Federal Data Sources

Our geospatial strategy draws upon yearly school district boundary shapefiles and zip code shapefiles sourced, respectively, from the NCES and the U.S. Census Topologically Integrated Geographic Encoding and Referencing (TIGER) system (NCES, 2024).[4] We also use state boundary shapefiles from government sources: the NC OneMap system (North Carolina Department of Public Safety, 2025) and the Washington Geospatial Open Data Portal (WaTech Geospatial Program Office, 2025). Furthermore, we clean and merge a variety of student population variables from the NCES Common Core of Data (U.S. Department of Education, 2024), as well as child poverty and population data from the Census Small Area Income and Poverty Estimates (SAIPE) program (Census Bureau, 2025b).

We draw from federal school district shapefiles to link NCES identifiers with school districts' physical boundaries. Similarly, the Census TIGER/LINE zip code shapefiles link 5-digit identifiers for zip code tabulation areas with zip code boundaries. We then use district-year information from the NCES Common Core of Data on the total number of students, students qualifying for free or reduced price lunch, and students classified as White, Black, Hispanic, and Asian. Longitudinal Small Area Income Poverty Estimates (SAIPE) of adult population, child population, total population, and child poverty rates complement the NCES data by considering communities within school district boundaries but not necessarily within the schools themselves.

The QOR Method for Matching Voters to School Districts

We integrate individual voter and geospatial data sources through the multi-stage QOR method. In short, QOR converts voter addresses into longitude and latitude coordinates (point geometries) in its first stage and then identifies the school district in which that voter is located either through the *Overlay* or *Recover* stage. Ideally, a voter's address is present within a validated database such as the Census TIGER/Line system, in which case submitting a *Query* to the Census allows our software to *Overlay* the point and school district geometries. We then *Recover* any observations missing from Census records using distances between registration zip codes and school districts. We provide software to implement these stages via an original R package with three functions that directly correspond to each sub-method (Shepardson et al., 2026). Although the software is intended as a scalable technique for placing voters within school districts at many different points in time, users could easily modify the program to place any unit of interest with an address inside any school district or other special district government. For brevity, we provide a conceptual description of each step below and include technical details about the software in Online Supplement C.

Step 1: Query

The initial *Query* stage exports text strings of each voter's registration address to the Census Geocoder API by default (Census Bureau, 2025a). We extract addresses and use geocoding (matching addresses against records in a spatial database; Cambon et al., 2021) to return longitude/latitude coordinates from the same TIGER database that provides our school district and zip code shapefiles. The accuracy of the match is improved through basic string variable cleaning procedures such as removing excess white space from the original voter address records, and through taking advantage of the Census tool's several "vintages," or options for temporal snapshots. To this end, Census Geocoder documentation suggests that the address database can be switched out with data reflecting a small set of different time periods. While vintages do not exist for all years, we select the closest vintage to a data year among the time periods where vintages do exist *after* the start of that year.[5] For example, since the Census has yearly vintages from 2017 onward, data years 2013-2017 in North Carolina are all assigned the 2017 Census snapshot during geocoding; this allows us to estimate voter locations at the plausibly correct time period while accommodating the presence of real estate trends. In sum, *Query* takes in the voter registration list for a specific state in a specific year and returns voters as point geometries.

Step 2: Overlay

The second stage in our process is *Overlay*, which sorts individual voters into school districts. We intake the returned, successfully geocoded voter coordinates from *Query* as point geometries and then match individual voters with the NCES identifiers corresponding to the school district whose polygon (based on NCES spatial data) contains their coordinates. Our assignment process is conceptually illustrated in Panel A of Figure 2. The voter point geometry represents the physical space occupied by a North Carolina voter along a 2-dimensional plane, where only two other school districts are rendered for simplicity (although the remainder are visible in the statewide view). Logically, the voter is clearly located in Columbus County Schools rather than Whiteville City Schools, and our process is sufficiently granular to mechanically understand where exactly Columbus ends, and Whiteville begins. This voter is therefore matched to Columbus Schools based upon the spatial boundaries of each district.

The first two stages facilitate voter-district matching across almost all observations. Supplemental Table C1 shows that the percentage of registered voters who are matched to a coordinate after *Query* is above 95.5% for all years in North Carolina, and above 97.4% for all years in Washington.

Step 3: Recover

In a third stage that we call *Recover*, we fill in the remaining 3-5% of voters that the Census tool could not locate. The most important of such voters are those whom (1) the Census tool placed outside the focal state, or (2) the Census tool could not locate. These phenomena could arise, for one, because the geocoding tool assigns input addresses to their *most likely* coordinates rather than guaranteeing a city-state-street address match. Alternately, it is also possible that the addresses in state voter records may contain errors or be affected by the slightly different times in which the spatial data sources were created. We implement the *Recover* stage to address these issues by assigning unmatched voters (who are returned from the *Query* function as their own dataset) the closest school district to their registration zip code based on distances between zip code and district center points.

Figure 3, Panel B shows a prototypical example of a voter whose school district is only matched in the *Recover* stage. This North Carolina voter is located somewhere inside a zip code that intersects three different school districts: Duplin County Schools, Onslow County Schools, and Pender County Schools. Because we cannot geocode the voter's actual address, we do not know their exact location. However, voters nearly always provide zip codes[6] during voter registration, allowing us to locate their respective zip code's center point on the 2-dimensional plane. Similarly, we can approximate center points for any given school district (estimated as internal points). We then calculate ellipsoidal distances between each voter's zip code centroid and internal points for every school district in the state for all unmatched voters (from the *Query* stage). Recall that the *Query* stage isolated voters missing from Census address records into a separate dataset from the voters fed into *Overlay*. The *Recover* stage effectively selects the best approximate match for these voters by minimizing the distance between a given zip code's centroid (exact center point) and all school districts' internal points (rough center points).

Addressing Ward Elections

Some districts elect their board members through wards, and (as noted above) these have presented a unique challenge to prior efforts at estimating school board election turnout using election returns. We address this challenge by calculating turnout rates at the school district-by-precinct-by-date level and removing registered voters in district-by-precinct areas on dates with turnout below a 0% threshold (see details below). We assume that registered voters in these areas were likely ineligible to vote due to ward structures. This process also helps us to identify cancelled primaries in Washington state. We provide full details on the ward identification procedures, distributional analyses, and validation checks in Online Supplement D.

To determine the threshold for dropping registered voters in these district-by-precinct-by-date cells, we examined sub-district turnout distributions and found compelling evidence *in both states* that the most obvious bunching occurs between 0% and 5% district-by-precinct-by-date level turnout. We also examined observations within confirmed at-large districts and find no evidence of such bunching—suggesting that this is indeed a function of ward structures. We remove registered voters from school district-by-precinct-by-date cells with 0% turnout as our preferred method but also show it with less than 5% turnout in off-cycle elections. This removes only a small fraction of the individuals registered to vote in any given school district-by-election date observation (from the final panel) and substantially improves data validity compared to alternatives (e.g., excluding all ward-based elections or retaining all ineligible voters).

Match Rates

In Supplemental Table C1, we show that over 99% of both states' voters from the raw registration files are successfully matched to a school district, with at most around 4% or fewer in any year matched via zip code rather than exact coordinates. This coverage holds across both North Carolina and Washington. The exact reasons why a small segment of voters are not locatable by either address coordinates or zip codes remains unclear. Still, the conditions under which any voter is dropped from the data are stringent. Such voters must neither appear in the

Census database of valid addresses nor have provided, at time of voter registration, a zip code whose first 5 digits correspond to the Census zip code shapefiles from the relevant year.[7]

Calculating Voter Turnout in School Board Elections

We use the resulting data from the QOR method to calculate school district-level turnout rates both overall and for each voter subgroup of interest. We formally define school board election turnout, V_{it}, in school district *i* and on election date *t*, as 100 times the number of district residents who voted on an election date (AV_{it}, or actual voters) divided by the number of individuals registered to vote who were eligible to vote within the same district and on the same election date (RV_{it}, or registered voters):

$$V_{it} = \frac{AV_{it}}{RV_{it}} * 100.$$

Estimating Roll-Off

We define voting in a school board election as voting on the day in which a school board election takes place in one's district. However, we do not observe actual ballots that individuals cast. It is therefore possible to vote for only the "top of the ticket" (e.g., Governor, President) and not cast a vote in their school board election. This phenomenon is called "roll-off" (Meier et al., 2005). To estimate the extent to which our turnout estimates are affected by this issue, we use a subset of the official election returns we obtained through extracting school board contests from the North Carolina Board of Elections website (North Carolina Board of Elections, 2025). Specifically, we compare votes cast in at-large, vote for 1 races (where the number of voters is equivalent to the votes cast) and the total number of voters. This allows us to assess the extent to which voters participate in other same-date races while leaving school board contests blank.

Table 2 describes total roll-off in all 75 applicable North Carolina elections before breaking these elections out into 62 "On-Cycle" (Midterm or Presidential) and 13 "Off-Cycle" (all other) contests. Roll-off in this context is the difference between the total number of voters in school board elections and the total number of voters (in any contest) across the same set of district-dates. Aggregate roll-off is substantively small: roll-off peaks at 4.72 percentage points in on-cycle school board elections and 1.39 percentage points off-cycle. The shifts by election cycle timing are intuitive given that many off-cycle contests only include local elections.

Findings

Validity of the QOR-Generated School Board Elections Data

We generate evidence in support of the validity our QOR method by examining the convergent validity, incremental validity, and nomological validity of QOR voter turnout estimates. In other words, we show that QOR estimates are highly correlated with benchmark measures of turnout, add value above and beyond other commonly used measures of turnout, and adhere to patterns we would expect based on prior research.

Convergent Validity (Correlations with Benchmarks)

Convergent validity refers to the degree to which a measure correlates positively with other measures or indicators that theoretically assess the same or similar constructs, providing evidence that the measure captures its intended target. Strong convergent validity is demonstrated when a new assessment shows substantial positive associations with established measures of the same underlying construct. The unique structural features of school boards and ballot reporting procedures in North Carolina allow us to create several benchmark measures to which we can compare our QOR estimates for the purpose of examining convergent validity.

The North Carolina State Board of Elections posts official returns with county-level turnout statistics, county-level counts of registered voters, and county-level counts of “Ballots Cast” (i.e., number of unique voters)[8]. We take advantage of the fact that roughly 69% of North

Carolina school districts exactly match county boundaries to compare turnout rates, the number of residents registered to vote, and the number of voters casting ballots from the QOR dataset against the official election returns for this subset of districts where counties and districts are aligned. We are able to do this for all county-wide, at-large school board elections (but not for districts in counties with city districts or ward-based elections) because in county-wide elections that are at-large, the school board election has the same turnout, registered voters, and ballots cast as other county-level elections. Figure 3 illustrates the districts that contribute to this validation procedure in purple. These elections account for 96 of the 560 elections observed in the full North Carolina panel. We also compare the SAIPE adult population to QOR-produced total registered voters. We do this for the QOR panel as constructed with the 0% threshold (preferred) and the 5% threshold for dropping wards (see above for details).

In Table 3, Panel A, we show the correlation, percentage difference, and percentage-point difference (where applicable) between the North Carolina QOR estimates and both (1) county election returns for our benchmark sample of cases and (2) SAIPE estimates. The QOR estimates are highly correlated with the county election returns and SAIPE estimates. The Pearson's r correlation coefficients are equal to or greater than 0.995. We also present the median percentage difference between each of our QOR measures and each of the benchmark measures and the median percentage point difference between our QOR measure and the benchmark measures for turnout rates (the only measure in percentage units). On average our QOR measures underestimate turnout and the number of voters by about 5% and overestimate the number of adults registered to vote by just 0.6%. The -5% average difference in turnout amounts to a -2.5 percentage point average difference in turnout estimates.[9]

In Table 3, Panel B, we further examine evidence of convergent validity by comparing the QOR data to a panel in which we manually assigned voters in North Carolina to their school district using their city and county in the address on their voter registration. Again, this is only possible because North Carolina has a relatively small number of public school districts (115) with a clear assignment process based on established geographies: 100 districts assign individuals to school districts based on their county of residence and the other 15 based on their city. This would not be possible in other states like Washington where school district boundaries do not always overlap with other geographies. As is shown in Panel B, the panel where we manually assigned voters performs very similarly to our QOR estimates. The manual assignment correlations with variables from official county-wide, at-large election returns never drop below 0.995; on average they underestimate turnout and the number of voters by about 5%; and they overestimate the number registered to vote by 0.7%. As with the QOR panel, the -5% difference in turnout amounts to a -2.5 percentage point difference in average turnout estimates between the manual assignment panel and the benchmark estimates. In short, the high level of consistency between the QOR turnout estimates and the benchmark estimates in county-district aligned communities provides evidence of convergent validity for the QOR-generated estimates overall.

Incremental Validity (Improvements on Alternatives)

We also examine our QOR estimates in terms of incremental validity. Incremental validity refers to the degree to which a new measure explains additional variance in an outcome or provides unique predictive power beyond what is already accounted for by existing measures. It demonstrates that the new measure adds meaningful information above and beyond what other assessments already capture, justifying its use alongside or in place of established instruments. We do this by comparing turnout, number of individuals registered to vote, number of voters, and number of districts included in the QOR panel to that which we would observe if we used zip codes to match voters to school districts instead of the more granular QOR method. The alternative zip code-only strategy assigns voters the school district that has the largest land area overlap with their zip code. We match small districts first because to ensure they are matched with a zip code (see Online Supplement F). We would expect zip codes to be correlated with school districts but to be an imperfect way of identifying them given that zip codes and school districts are not always aligned. We use zip codes instead of counties because zip codes are smaller than counties and typically smaller than school districts. Additionally, only 6.9% of districts nationally have boundaries that are the same as county boundaries (see details above). We use both North Carolina and Washington state panels to examine evidence of this type of incremental validity.

Figure 4 displays scatter plots for North Carolina of the district-by-election estimates from the QOR panel on the x-axis and the zip code-match panel on the y-axis. Here, the zip code-match panel performs well, particularly for turnout rates (Panel A). With individuals registered to vote and voters (Panels B and C), the zip code panel performs well in districts with larger numbers of individuals registered and voters but diverges more from QOR in smaller areas. In Panel D of Figure 4, we highlight the small districts that had to be matched first in order to be assigned at least one zip code. We also show that, under this matching process, the zip code-match panel can generate a unique estimate for every district-by-election observation that we know occurred (Panel D).

Figure 5 displays the same set of scatter plots for Washington state, where the 295 school districts tend to be smaller on average than in North Carolina. Here, the zip code-match panel performs worse. The estimates for turnout, individuals registered to vote, and voters are more varied between the QOR panel and the zip code-match panel, and not in a clearly systematic way. Additionally, the zip code-match panel cannot generate estimates for a number of district-by-election observations (represented by Xs in Panels A-C). The zip code match does not match any voters to school districts in several periods because there are too many school districts associated with each of these zip codes (see Panel D). Therefore, we find evidence that the QOR-generated estimates provide value above and beyond the use of zip codes to identify school district, particularly in contexts like Washington with relatively small school districts.

Nomological Validity (Replications of Prior Findings)

We also examine evidence of the nomological validity of turnout estimates produced with the QOR method. Nomological validity is a form of construct validity that refers to the degree to which a measure demonstrates theoretically predicted relationships with other variables (Kock et al., 2024; Lim, 2024). It describes whether a construct behaves as expected within an established conceptual framework, like when a new measure replicates findings from previous studies or when patterns of correlations with extant constructs align with theoretical expectations. To demonstrate nomological validity, we examine turnout rates using QOR estimates disaggregated by election timing, primary vs. general elections, and demographic groups. A substantial literature shows that election timing is highly predictive of turnout rates; local elections held on-cycle with national elections have significantly higher rates of voter turnout than off-cycle elections (Anzia, 2014; Dynes et al., 2021; Hajnal et al., 2022; Hartney, 2021). Similarly, previous scholarship has consistently demonstrated higher rates of turnout in general elections than primaries (Gerber et al., 2017). Finally, prior scholarship shows higher rates of turnout among older and White voters than younger and non-White voters (Barber & Holbein, 2022).

We begin by examining school board election voter turnout rates in North Carolina and Washington across different election times (even year, presidential; even year, mid-term; even year, non-November; odd year, November; and odd-year, non-November). We show in Panel A of Table 1 that 46.2% of voters participated in the typical North Carolina school board election in our panel of 560 races. However, as expected, we find notable differences in turnout rates based on election timing. Voter turnout was highest in school board elections when they coincided with presidential elections (69.3% averaged across 155 races), while turnout was much lower when school board elections coincided with mid-term elections (47.3% across 230 races). Average turnout was even lower in off-cycle elections, but there are meaningful distinctions even within this group. Turnout was lowest (12.1%) in the four races that occurred in non-November months of odd years and the 39 races held in November of odd years (15.4%). This was notably lower than turnout in the 132 off-cycle races that occurred in even years during the spring, when one in four registered voters (24.1%) turned out, on average.

We show consistent results for Washington in Panel B. Washington's school board elections all occur in odd years, and Washington exclusively uses mail-in ballots. General elections are held in November of odd years and primaries are held in August of odd years. We include primaries for these within-state comparisons across different election times. The average overall turnout rate was 41.1% in November (general) elections and 26.2% in August (primary) elections across the 2,728 elections represented in our panel.[10]

We also examine differences in turnout rates across demographic groups, made possible by our QOR method which uses individual voter-level data to generate subgroup turnout estimates. Table 1 shows that turnout rates are highest for citizens over 60, who vote at a rate of 61.4% in the average school board election in North Carolina, compared to 48.5% for registered voters aged 36-60 and 27.6% for those 35 and younger. Similarly, in Washington, turnout is highest among voters over age 60 (54.9%) and lowest among those 35 and below (15.7%). We find few differences across the two major parties and across sex: 50.3% of registered Republicans voted in the typical North Carolina school board election compared to 47.2% of registered Democrats. Similarly, men and women exhibit similar turnout rates in both North Carolina and Washington (46.1% vs. 47.2% and 33.6% vs. 34.5%, respectively). We do find notable differences across race and ethnicity. White individuals are most likely to vote (49.5%) in the average school board election, followed by Black (41.4%), Asian (35.2%), and Hispanic (28.4%) residents.[11] Overall, the differences in turnout rates based on election timing, primary vs. general elections, and demographic groups are all consistent with what one would expect based on prior research on political behavior.

School Board Election Turnout Rates by School District Characteristics

We next demonstrate the value of the QOR method and resulting data by examining substantive research questions that could not be answered without access to voter turnout data at the school district level and for subgroups of voters within school districts. More specifically, we examine how turnout rates and the demographic representativeness of electorates vary depending on the demographic characteristics of the student populations served by the school districts. We focus on two demographic characteristics of school districts: (1) the economic composition of the district as captured by the share of the student population that is eligible for free or reduced-price lunch and (2) the racial/ethnic makeup of the school district as reflected by the share of the student population that racially identifies as non-White. We chose these characteristics to illustrate the value of the QOR method as they are both commonly used in academic research and important for understanding equity in school governance. However, researchers could also examine turnout and representativeness across any school district characteristics available from NCES or any other data source that includes district identifiers.

Results demonstrate that school board election turnout is lower in districts serving higher shares of students qualifying for subsidized lunch and in districts serving higher concentrations of non-White students. Figure 6 shows scatter plots comparing school board election turnout on the y-axis based on the two district-level demographic characteristics on the x-axis in both Washington in blue and North Carolina in red (see Supplemental Table G1 for OLS regression results). In both states, districts with higher shares of students eligible for free or reduced-price lunch tend to exhibit lower voter turnout (Panel A). On average across the two states, a 10-percentage point increase in the share of students eligible for free or reduced-price lunch is associated with roughly a 0.78-percentage point decrease in turnout in school board elections. The relationship is stronger in Washington where a 10-percentage point increase in the share of students eligible for subsidized lunch is associated with roughly a 1.04-percentage point decrease in turnout (compared to a 0.5-percentage point decrease in North Carolina). Panel B of Figure 6 similarly shows that turnout decreases as the proportion of students of color served by the district increases. Specifically, a 10-percentage point increase in the share of students of color is associated with roughly a 1.73-percentage point decrease in turnout in school board elections, on average. Here, the relationship is similar across states. These are notable associations, given the low overall turnout rates in school board elections (34.0% and 46.2% in Washington and North Carolina, respectively, see Table 1). These patterns suggest that communities serving more economically disadvantaged and larger non-White student populations have less influence in school governance elections, raising concerns about representational equity and the potential responsiveness of school boards to the needs of these populations.

Panels C and D of Figure 6 present scatter plots examining the relationship between the demographics of the student population in a district and the representativeness of voters of color in school board elections. We conduct this analysis in North Carolina where the public voter registration record contains information on voter race. We define representativeness as the difference between the share of the election day electorate (voters) comprised of people of color and the share of the school population comprised of students of color. Negative values suggest that the electorate is less racially representative of the students that the district serves. Panel C shows that districts with higher shares of students eligible for free or reduced-price lunch tend to have less representative participation by voters of color. On average, a 10-percentage point increase in the share of students eligible for free or reduced-price lunch is associated with approximately a 1.3-percentage point decline in the representativeness of voters of color in school board elections. Panel D reveals a stronger relationship with racial composition: a 10-percentage point increase in the share of students of color corresponds to about a 2.7-percentage point decrease in the representativeness of the electorate, on average.

Discussion

School boards are institutions that make significant education policy decisions in the U.S. but are understudied, in large part, because of the paucity of data on the elections that determine who serves on them. Currently there is no validated procedure for estimating school board election turnout at the school district level, making it difficult to study questions related to how district level education policy affects voter turnout and how voter turnout affects school district policy and educational outcomes. In this study we develop a novel and replicable process for estimating voter turnout in school board elections at the school district level, both overall and for key subgroups of voters, by using individual voter level data and geospatial methods to link voters' residential addresses to school districts. We also demonstrate how the QOR method generates data that can be used to explore the relationship between school board turnout and district characteristics.

We provide evidence of the convergent, incremental, and nomological validity of our QOR-based estimates of turnout, registered voters, actual voters, and adult voting populations. Our estimates are strongly correlated with official "benchmark" election results in North Carolina and substantially improve upon conventional zip-code-based school district turnout estimates in contexts like Washington state which have many small districts. We provide further evidence of the validity of the QOR method by showing that QOR estimates of voter turnout are correlated with election timing, election type, and voter demographic characteristics in ways that we would expect based on prior literature.

We also illustrate the potential value of the resulting QOR data by addressing two substantive research questions that would not be answerable without school district level and subgroup turnout data. We show that the share of students eligible for subsidized lunch and share of non-White students are both negatively correlated with turnout in school board elections and non-White voter representativeness. These patterns suggest that districts serving more economically disadvantaged student populations and greater shares of students of color not only experience lower turnout overall but also exhibit greater race-based disparities in who participates, raising concerns about equitable representation in school governance.

The development of the QOR method for estimating voter turnout ameliorates the three major challenges to studying policy-relevant questions related to school boards. First, there is a lack of data on elections at the school district level. Turnout in local areas is usually reported at the county level, but only 6.9% of school district borders are conterminous with counties. The

QOR method overcomes this challenge by matching voters to districts based on their home addresses, rather than counties (or zip codes which are imprecisely related to districts when districts are small). Second, prior studies on school board elections have relied on a number of assumptions that the QOR method does not rely on. For example, scholars using candidate-level votes have assumed that every voter with the option to vote for two candidates always votes for two candidates and that ward-based elections are not systematically different from at-large elections. Because the QOR method relies on data at the individual level, it is based on estimates of both the number of actual voters and the number of voters who were eligible to vote, allowing scholars to estimate a valid estimate of turnout in these contexts. Third, prior efforts have not embedded a process that could generalize nationally or allow for subgroup effects. The QOR method offers a process by which researchers could calculate national data, conditional on obtaining national voter files, such as those that national data vendors (e.g., Catalist, L2) provide. The QOR method also allows for individual subgroup estimation based on any individual voter characteristic available in the voter file. Here, we use race, sex, party, and age, but other sources may have other individual voter characteristics. In short, prior studies involved intensive data collection and/or strong assumptions about voter behavior, but the QOR method allows researchers to easily estimate school board election turnout throughout the country.

The QOR method enables scholars to study a host of new research questions. Future scholarship can use the QOR method to examine the effects of public policies on voter participation, including the effects of both education policies (e.g., book bans, school safety policies, school-based health clinic access) and electoral policies (e.g., school board election timing) across the country. Additionally, the QOR method is useful for exploring how school board voter turnout influences policy and student achievement. It also allows for replicating previous studies on accountability, timing, and representation that were conducted in single states. This is valuable both for understanding the political processes that lead to the policies that are adopted, as well as the role of school systems in shaping civic participation.

Beyond school board elections, the QOR method can easily be adapted to connect any address to a corresponding school district and to estimate voter turnout for a variety of special district governments. For example, scholars could use the QOR method to connect non-profit organizations to their corresponding school districts. Additionally, the QOR method is useful for estimating special district government turnout because, like school districts, these boundaries often cross county and zip code borders. Nationwide there are 39,555 special district governments (Census Bureau, 2022) responsible for critical government functions including: water, sewer districts, waste disposal, fire, economic development, and pest control. Like school board races, voter turnout in special district government elections is profoundly understudied.

School boards make consequential decisions affecting millions of students, families, and communities. These decisions occur at the school district level, however, until now there has been no way to comprehensively study voter participation at the school district level. The QOR method represents an important methodological advancement for education politics and policy research, offering a scalable and validated approach to measuring voter turnout in school board elections at the unit where significant decision-making is made. Understanding who participates in the election of these critical decisionmakers is essential to understanding policy predictors and impacts across governments, and how to promote more equitable outcomes. The QOR method offers a valuable tool for researchers, practitioners, and advocates interested in enhancing political representation and strengthening school systems throughout the country.

[1] Users can easily install QOR using the R terminal command: devtools::install_github("adam-p-shepardson/QOR", dependencies=TRUE) if they have installed the *devtools* package. For more information and to check the required list of dependencies that must be installed prior to use, visit https://adam-p-shepardson.github.io/QOR.
[2] We detail this process in Online Supplement B. Our term "consequential" describes elections where at least one board candidate won a seat on the specific election date. This includes all general election races and non-partisan primary races where a candidate received enough votes to be elected to the seat (in which case, the race did not appear on a later ballot). This substantive focus has also been applied in prior work on turnout in municipal elections (Hajnal et al., 2022).
[3] North Carolina school board election schedules are unique to individual boards (Anzia, 2014), and like cities can sometimes, to our knowledge, employ a runoff structure when there is no majority winner for a seat (North Carolina Legislature, 2018). These features prompted us to use a schedule-agnostic voter extract date that accounts for all registered voters present by the year's end. For each data year (e.g., the 2013 elections), we utilize registration records from Jan. 1st of the subsequent year (e.g., 2014). This errs on the side of including voters who moved into NC or newly registered to vote between Jan. 1st of the data year and the election that occurred in the data year. We assumed that ignoring these voters would bias the eventual turnout calculation's denominator and numerator, whereas overcounting recent movers/registrants between an election and the next Jan. 1st likely only injects a smaller amount of error into the denominator.
[4] Following standard practices in geospatial data science, we ensure that all shapefiles are set to a common coordinate reference system. To this end all geometries adopt NAD83, the same reference system as the NCES school district shapefiles.
[5] New apartment or housing construction over time (i.e., 2015) might go unaddressed if we assign Census data from an earlier vintage (i.e., 2010) to later data years.
[6] We use zip codes (rather than counties) because they are typically smaller than school districts, and as noted above, only 6.9% of district boundaries align with county boundaries. However, in states with county-based school district systems, counties could be used instead.
[7] We also ensure that voters fed into the QOR method are uniquely identified via removing duplicate observations along the state voter ID variable during pre-processing. These are extremely rare occurrences assumed to be errors in state record keeping.
[8] We obtained all such election returns through web-scraping code that submitted automated requests for every school board election between 2013 and 2022. The code employed regular expression pattern matching followed by manual review of cases with missing values and manual NCES ID assignment to whittle down raw scraping results and produce a validation dataset at the school board-by-election date level (see Online Supplement E for more information).
[9] Readers may wonder why these numbers are not perfectly correlated. This pattern is plausibly driven by registered voters moving within the state of North Carolina during a given year. We count such voters as part of the voter pool for the district they occupy at the end of the year, but some small number may have moved into their end-of-year district after already having voted in a different district. These dynamics may at least partially explain the slight under (over) estimates across each turnout sub-component on average. We would not expect the magnitudes of these errors to be exactly equivalent, as voters may move into districts whose elections were not county-wide and at-large.
[10] Washington State contains 295 public school districts, and we observe both their primaries and general elections every two years.
[11] The Hispanic designation is mutually exclusive with the race designations.

Reference List

America Counts Staff. (2021). *North Carolina Gained Around 900,000 People Last Decade*. https://www.census.gov/library/stories/state-by-state/north-carolina.html

American National Election Studies. (2024). *Time Series Cumulative Data File (1948-2020)*. https://electionstudies.org/

Anzia, S. F. (2012). The election timing effect: Evidence from a policy intervention in Texas. *Quarterly Journal of Political Science*, *7*(3), 209–248.

Anzia, S. F. (2014). *Timing and Turnout: How Off-Cycle Elections Favor Organized Groups*. University of Chicago Press.

Arnold, R. D., & Carnes, N. (2012). Holding Mayors Accountable: New York's Executives from Koch to Bloomberg. *American Journal of Political Science*, *56*(4), 949–963. https://doi.org/10.1111/j.1540-5907.2012.00603.x

Barber, M., & Holbein, J. B. (2022). 400 million voting records show profound racial and geographic disparities in voter turnout in the United States. *Plos One*, *17*(6), e0268134.

Berry, C. R. (2024). The Timing of Local Elections. *University of Chicago Center for Effective Government, Chicago, IL*. https://blockyapp.s3.eu-west-2.amazonaws.com/store/74afb145c4a36f9cd2e3a31ce0421a26.pdf

Berry, C. R., & Gersen, J. E. (2011). Election Timing and Public Policy. *Quarterly Journal of Political Science*, *6*(2), 103–135.

Black, S. E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education*. *The Quarterly Journal of Economics*, *114*(2), 577–599. https://doi.org/10.1162/003355399556070

Blazar, D., & Schueler, B. (2024). Effective School District Policies and Practices: Synthesizing Theoretical Frameworks and Empirical Findings across Disciplines. *Journal of Research on Educational Effectiveness*, *17*(4), 647–686. https://doi.org/10.1080/19345747.2023.2269923

Burnett, C. M., & Kogan, V. (2022). Do nonpartisan ballots racialize candidate evaluations? Evidence from “Who Said What?” experiments. *Party Politics*, *28*(3), 541–553. https://doi.org/10.1177/1354068821990625

Carlson, D., Domina, T., Barron, N., Carter III, J., Perera, R., & Lenard, M. (2025). Backlash? Schooling Reassignments and the Politics of School Desegregation. EdWorkingPaper No. 25-1150. *Annenberg Institute for School Reform at Brown University*. https://eric.ed.gov/?id=ED672520

Census Bureau. (2022). *Census of Governments*. https://www.census.gov/programs-surveys/cog.html

Census Bureau. (2024). *Current Population Survey Voting Supplement*. https://www.census.gov/data/datasets/time-series/demo/cps/cps-supp_cps-repwgt/cps-voting.html

Census Bureau. (2025a). *Census Geocoder API*. https://geocoding.geo.census.gov/geocoder/

Census Bureau. (2025b). *Small Area Income and Poverty Estimates*. https://www.census.gov/programs-surveys/saipe.html

Chingos, M. M., Whitehurst, G. J., & Gallaher, M. R. (2015). School districts and student achievement. *Education Finance and Policy*, *10*(3), 378–398.

de Benedictis-Kessner, J., Lee, D. D. I., Velez, Y. R., & Warshaw, C. (2023). American local government elections database. *Scientific Data*, *10*(1), 912.

Dynes, A. M., Hartney, M. T., & Hayes, S. D. (2021). Off-cycle and off center: Election timing and representation in municipal government. *American Political Science Review*, *115*(3), 1097–1103.

Fischer, B. (2023). No Spending without Representation: School Boards and the Racial Gap in Education Finance. *American Economic Journal: Economic Policy*, *15*(2), 198–235. https://doi.org/10.1257/pol.20200475

Gerber, A. S., Huber, G. A., Biggers, D. R., & Hendry, D. J. (2017). Why don't people vote in US primary elections? Assessing theoretical explanations for reduced participation. *Electoral Studies*, *45*, 119–129.

Grissom, J. A. (2010). The determinants of conflict on governing boards in public organizations: The case of California school boards. *Journal of Public Administration Research and Theory*, *20*(3), 601–627.

Hajnal, Z. (2010). *America's uneven democracy: Race, turnout, and representation in city politics*. Cambridge University Press. https://books.google.com/books?hl=en&lr=&id=B9D5yQgiKgcC&oi=fnd&pg=PR7&dq=Hajnal,+Z.+(2010).+America%27s+Uneven+Democracy:+Turnout,+Race,+and+Representation+in+City+Politics.+Cambridge:+Cambridge+University+Press.&ots=zXc7U0pSlX&sig=AMeFad-STBB_dwvhWz1xLxihyN4

Hajnal, Z., Kogan, V., & Markarian, G. A. (2022). Who votes: City election timing and voter composition. *American Political Science Review*, *116*(1), 374–383.

Hartney, M. T. (2021). Revitalizing Local Democracy: The Case for On-Cycle Local Elections. *Manhattan Institute, Oct*, *14*, 9.

Henig, J. R., Hula, R. C., Orr, M., & Pedescleaux, D. S. (2001). *The color of school reform: Race, politics, and the challenge of urban education*. Princeton University Press.

Hess, F. M. (2002). *School boards at the dawn of the 21st century: Conditions and challenges of district governance.* National School Boards Association. https://eric.ed.gov/?id=ED469432

Hess, F. M., & Leal, D. L. (2005). School house politics: Expenditures, interests, and competition in school board elections. *Besieged: School Boards and the Future of Education Politics*, *249*. https://books.google.com/books?hl=en&lr=&id=m7X3B--Q1goC&oi=fnd&pg=PA228&dq=+Hess,+R.+%26+Leal,+D.+(2005).+School+House+Politics:+Expenditures,+Interests,+and+Competition+in+School+Board+Elections.+Besieged:+School+Boards+and+the+Future+of+Education+Politics.&ots=-ttdjLO_4n&sig=Xv3E9jED5ogCfWwXOWsWu_k7dMg

Hill, K. A., Moreno, D. V., & Cue, L. (2001). Racial and Partisan Voting in a Tri-Ethnic City: The 1996 Dade County Mayoral Election. *Journal of Urban Affairs*, *23*(3–4), 291–307. https://doi.org/10.1111/0735-2166.00090

Holbein, J. B., & Hillygus, D. S. (2016). Making Young Voters: The Impact of Preregistration on Youth Turnout. *American Journal of Political Science*, *60*(2), 364–382. https://doi.org/10.1111/ajps.12177

Holbrook, T. M., & Weinschenk, A. C. (2014). Campaigns, Mobilization, and Turnout in Mayoral Elections. *Political Research Quarterly*, *67*(1), 42–55. https://doi.org/10.1177/1065912913494018

Howell, W. G. (2005). *Besieged: School boards and the future of education politics*. Rowman & Littlefield. https://books.google.com/books?hl=en&lr=&id=m7X3B--

Q1goC&oi=fnd&pg=PP1&dq=Howell,+2005+besieged&ots=-ttdjLPZ2m&sig=b_FkC7AqAaVPZys_YAnPFgIhAM4

Jacob, B. A. (2024). School Board Elections Before, During, and After the COVID-19 Pandemic. *AERA Open*, *10*, 23328584241291029. https://doi.org/10.1177/23328584241291029

Kaufmann, K. M. (2003). Black and Latino voters in Denver: Responses to each other's political leadership. *Political Science Quarterly*, *118*(1), 107–126.

Kirst, M. W., & Wirt, F. M. (2009). *The Political Dynamics of American Education*. McCutchan Pub Corp.

Kitchens, K. E. (2021). Partisan politics and public education: Finding the formula for (electoral) success. *State Politics & Policy Quarterly*, *21*(1), 55–79.

Kitchens, K. E. (2023). School Characteristics and Voting: What Matters in Turnout and Passage. *Urban Affairs Review*, *59*(6), 1838–1874. https://doi.org/10.1177/10780874221119987

Kock, F., Berbekova, A., Assaf, A. G., & Josiassen, A. (2024). Developing a scale is not enough: On the importance of nomological validity. *International Journal of Contemporary Hospitality Management*, *36*(10), 3233–3244.

Kogan, V., Lavertu, S., & Peskowitz, Z. (2018). Election Timing, Electorate Composition, and Policy Outcomes: Evidence from School Districts. *American Journal of Political Science*, *62*(3), 637–651. https://doi.org/10.1111/ajps.12359

Kogan, V., Lavertu, S., & Peskowitz, Z. (2021a). How Does Minority Political Representation Affect School District Administration and Student Outcomes? *American Journal of Political Science*, *65*(3), 699–716. https://doi.org/10.1111/ajps.12587

Kogan, V., Lavertu, S., & Peskowitz, Z. (2021b). The democratic deficit in US education governance. *American Political Science Review*, *115*(3), 1082–1089.

Kogan, V., Lavertu, S., & Peskowitz, Z. (2025). High Turnover With Low Accountability: Local School Board Elections in 16 States. *AERA Open*, *11*, 23328584251349180. https://doi.org/10.1177/23328584251349180

Kohfeld, C. W., & Sprague, J. (2002). Race, space, and turnout. *Political Geography*, *21*(2), 175–193.

L2. (2025). *L2 Voter File Data*. https://www.l2-data.com/

Lim, W. M. (2024). A typology of validity: Content, face, convergent, discriminant, nomological and predictive validity. *Journal of Trade Science*, *12*(3), 155–179.

Lublin, D., & Wright, M. (2024). Diversity matters: The election of Asian Americans to US State and federal legislatures. *American Political Science Review*, *118*(1), 380–400.

Lyon, M. A. (2023). Current Perspectives on Teacher Unionization, and What They're Missing. *Educational Policy*, *37*(5), 1420–1443. https://doi.org/10.1177/08959048221103798

Macartney, H., & Singleton, J. D. (2018). School boards and student segregation. *Journal of Public Economics*, *164*, 165–182.

Manna, P., & McGuinn, P. (Eds.). (2013). *Education Governance for the Twenty-First Century: Overcoming the Structural Barriers to School Reform*. Brookings Institution Press.

Marschall, M., & Lappie, J. (2018). Turnout in Local Elections: Is Timing Really Everything? *Election Law Journal: Rules, Politics, and Policy*, *17*(3), 221–233. https://doi.org/10.1089/elj.2017.0462

Marshall, M., & Ruhil, P. (2011). *Local Elections in America Project*. https://leap.rice.edu/

Meier, K. J., Juenke, E. G., Wrinkle, R. D., & Polinard, J. L. (2005). Structural Choices and Representational Biases: The Post-Election Color of Representation. *American Journal of Political Science*, *49*(4), 758–768. https://doi.org/10.2307/3647695

MIT Elections Lab. (2025). *Precinct Level Data*. https://electionlab.mit.edu/data

Moe, T. M. (2006). Political Control and the Power of the Agent. *Journal of Law, Economics, and Organization*, *22*(1), 1–29.

Moe, T. M. (2011). *Special Interest: Teachers Unions and America's Public Schools* (Reprint edition). Brookings Institution Press.

NCES. (2024). *School Geocodes & Geoassignments*. https://nces.ed.gov/programs/edge/geographic/schoollocations

NORC. (2024). *General Social Survey*. https://gssdataexplorer.norc.org/

North Carolina Board of Elections. (2025). *Data and Results*. https://www.ncsbe.gov/results-data

North Carolina Department of Public Safety. (2025). *North Carolina State and County Boundary Polygons*. https://www.nconemap.gov/datasets/9728285994804c8b9f20ce58bae45899/explore

North Carolina Legislature. (2018). *Determination of election results in cities using the plurality method* (163-292). https://www.ncleg.gov/EnactedLegislation/Statutes/PDF/BySection/Chapter_163/GS_163-292.pdf

North Carolina Legislature. (2023). *List maintenance* (§ 163-82.14). https://www.ncleg.gov/EnactedLegislation/Statutes/HTML/BySection/Chapter_163/GS_163-82.14.html

O'Donnell, J. (2025). *Vote Database*. https://votedatabase.com

Payson, J. A. (2017). When Are Local Incumbents Held Accountable for Government Performance? Evidence from US School Districts. *Legislative Studies Quarterly*, *42*(3), 421–448. https://doi.org/10.1111/lsq.12159

Shepardson, A., Lyon, M. A., Schueler, B. E., & Bleiberg, J. (2026). *QOR: Functions to Identify Who Can Vote in What School District* (Version 0.9.0.0) [Computer software]. https://adam-p-shepardson.github.io/QOR

Shi, Y., & Singleton, J. D. (2023). School boards and education production: Evidence from randomized ballot order. *American Economic Journal: Economic Policy*, *15*(1), 438–472.

Urban Institute. (2023). *State and Local Backgrounders*. https://www.urban.org/policy-centers/cross-center-initiatives/state-and-local-finance-initiative/state-and-local-backgrounders

U.S. Department of Education. (2024). *Common Core of Data*. https://nces.ed.gov/ccd/ccddata.asp

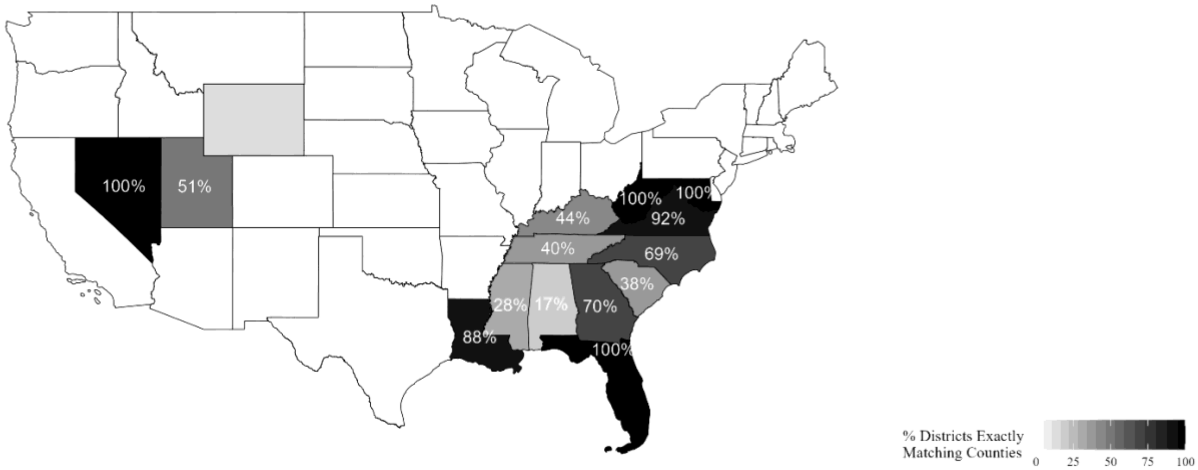
Washington Department of State. (2025). *Elections*. https://www.sos.wa.gov/elections#data_research

Washington Legislature. (2011). *Ballots by mail* (29A.40.010). https://app.leg.wa.gov/rcw/default.aspx?cite=29A.40.010.

WaTech Geospatial Program Office. (2025). *Washington Geospatial Open Data Portal*. https://geo.wa.gov/

Tables and Figures

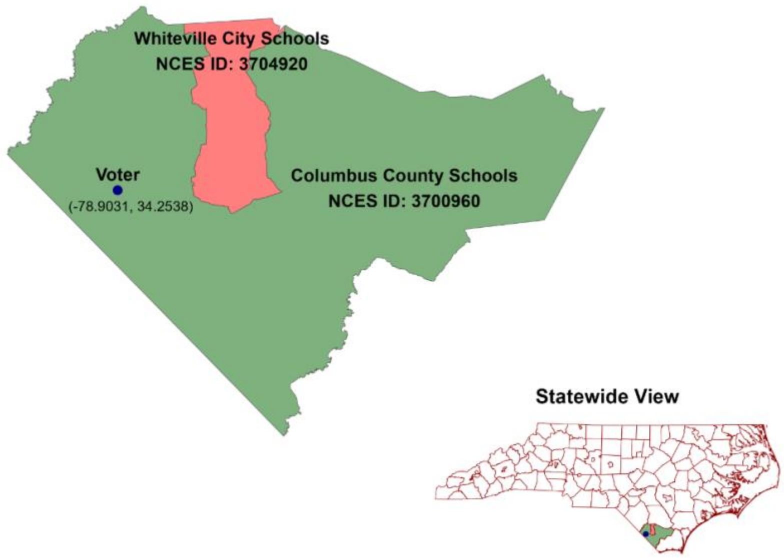
Figure 1. *School District and County Exact Matches within U.S. States*



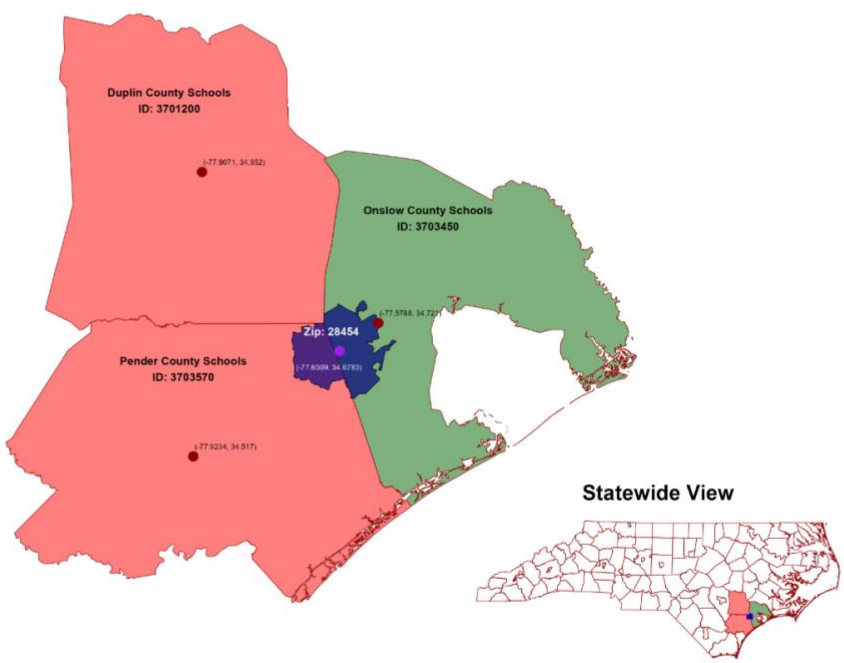
Note. The scale reports percentages of school districts within each state whose district boundaries match onto county boundaries within a 0.5% tolerance of district area (miles squared). Text labels are rounded to nearest whole number and only shown for states that have >15% district-county exact overlap. AK and HI contribute to Supplemental Table A1 but are excluded from map for simplicity.

Figure 2. *Query, Overlap, Recover (QOR) Process*

Panel A. Query and Overlay Voter Coordinates

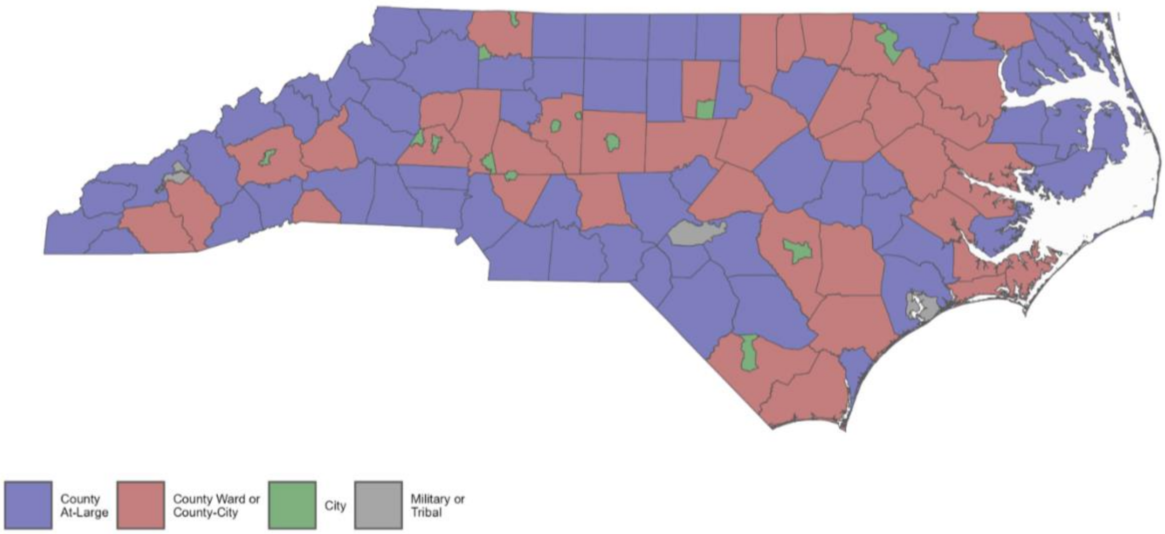


Panel B. Recover Leftover Voters using Zip Codes



Note. We color the hypothetical voter's matched district green. The "hole" in Onslow Schools is intentional, reflecting both water area and the Camp Lejeune military (federal) district.

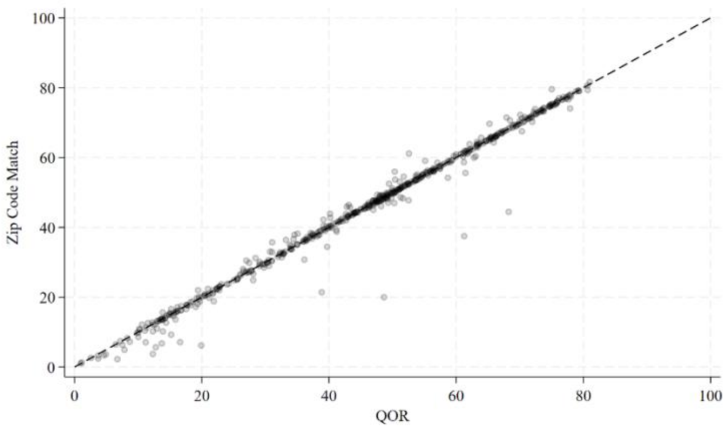
Figure 3. *North Carolina School District Types for Validation Purposes: 2013-2022*



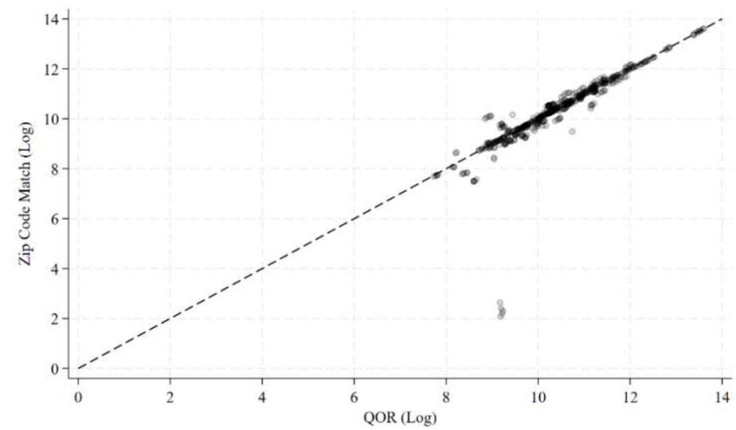
Note. School district shapefiles are sourced from NCES and reflect 2016 boundaries, and the state shapefile and election information are sourced from NC government websites. The validation is restricted to at-large countywide elections from 2013-2022 (in blue). The County At-Large category includes any counties with a least one likely at-large election.

Figure 4. *North Carolina QOR and Zip Code Voter-District Matching Comparison*

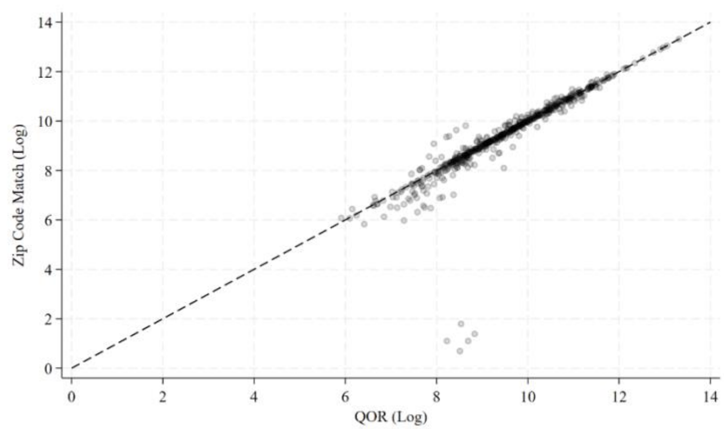
Panel A. Turnout



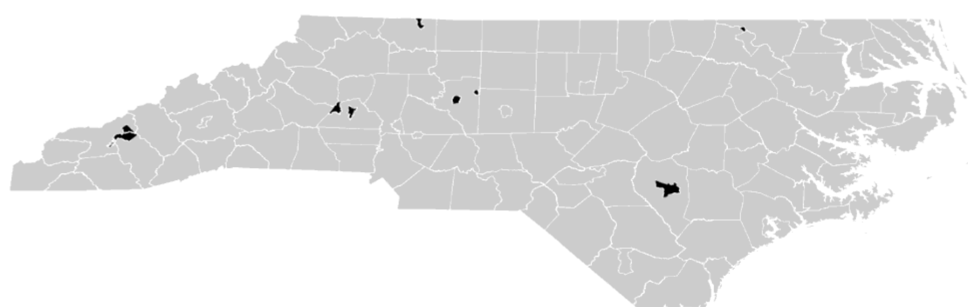
Panel B. Number Registered to Vote



Panel C. Number of Voters

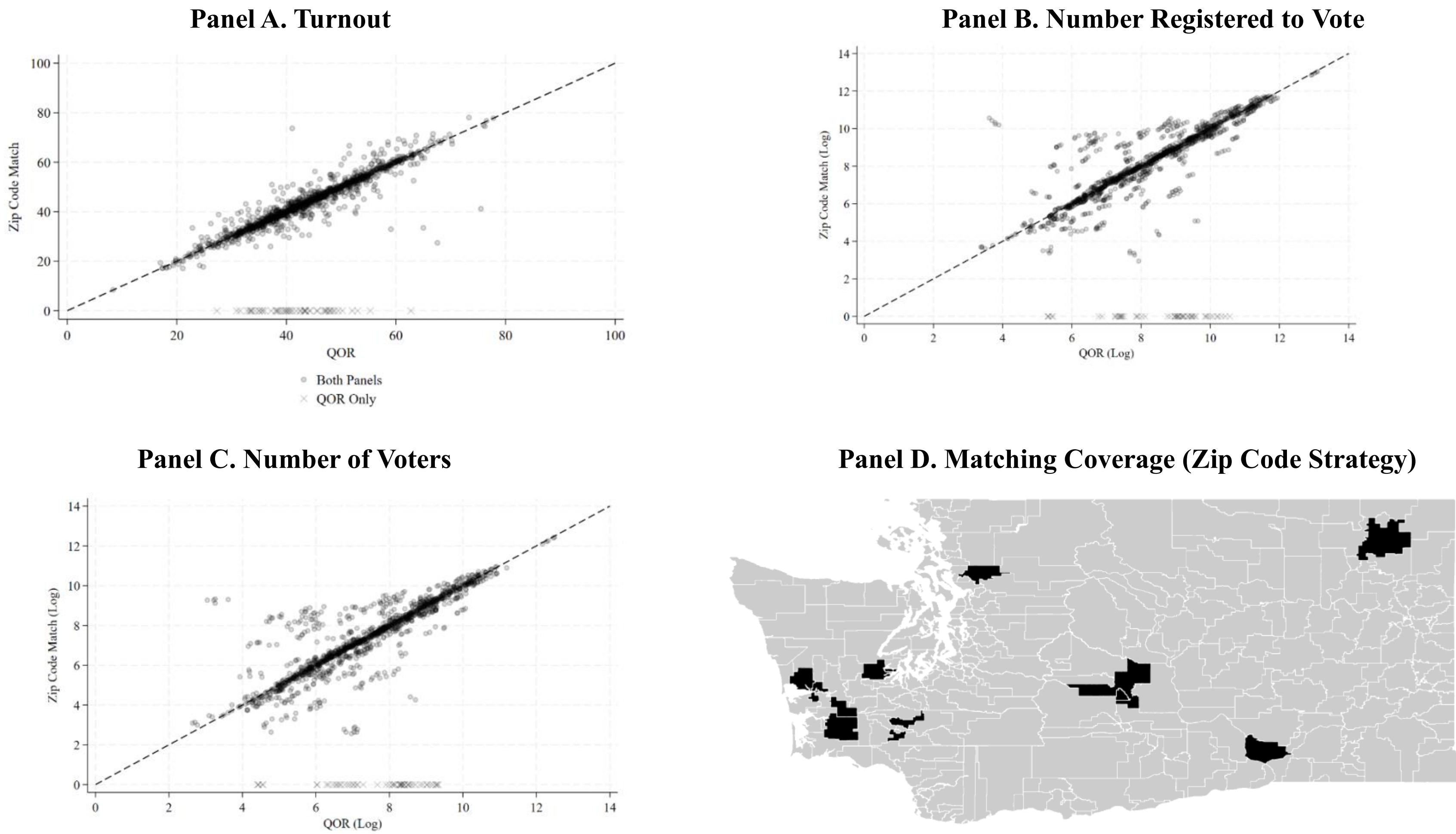


Panel D. Matching Order (Zip Code Strategy)



Note. Dotted line shows hypothetical perfect alignment between the two voter-district matching methods. Panel D shows school districts that need to be matched first during at least one year of the zip code-only district matching strategy. The grey school districts reflect those always matched after these (small) districts.

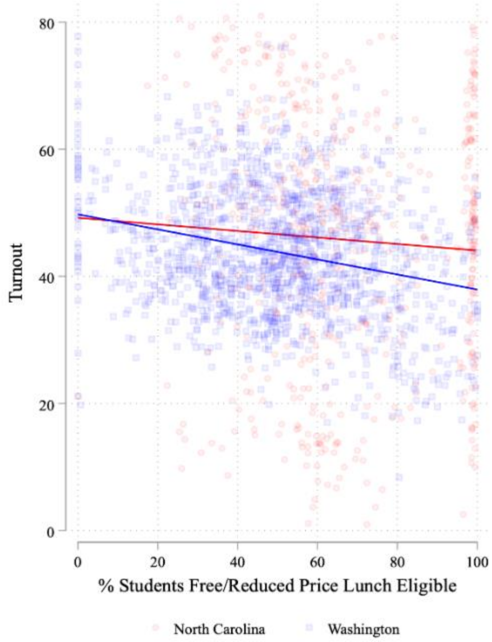
Figure 5. *Washington QOR and Zip Code Voter-District Matching Comparison*



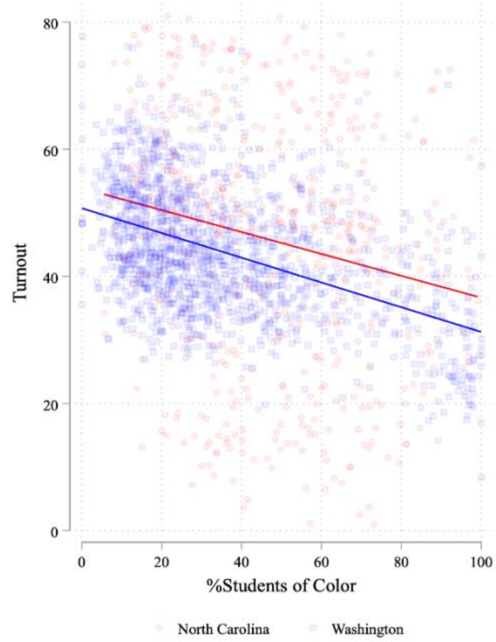
Note. Dotted line shows hypothetical perfect alignment between the two voter-district matching methods. Panel D colors black any school districts that cannot be uniquely matched to any zip codes during at least one year (due to zip codes crossing too many school districts' boundaries). Grey districts can always be matched to at least one zip code every year.

Figure 6. *Turnout and Representativeness Relationships with Select Voter Characteristics*

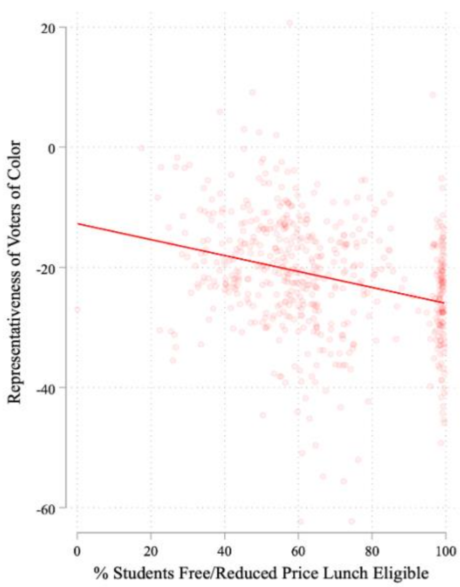
Panel A. Turnout and Income



Panel B. Turnout and Students of Color



Panel C. Representativeness and Income



Panel D. Representativeness and Students of Color

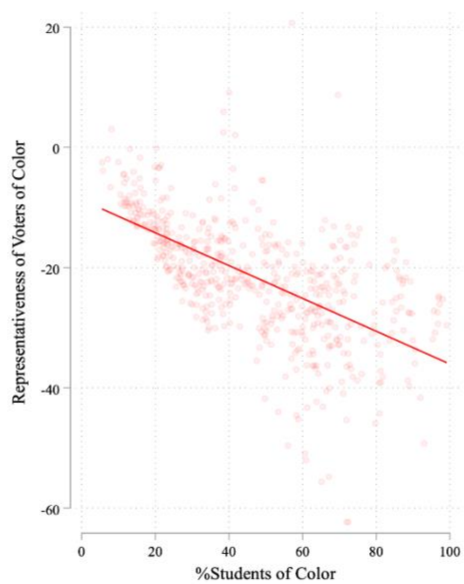


Table 1. *Voter Turnout on School Board Election Dates, by Detailed Timing, State, and Voter Characteristics*

	Overall	Even Year, Presidential	Even Year, Mid-term	Even Year, Non-November	Odd Year, November	Odd Year, Non-November
Panel A. North Carolina (2013-2022)						
Overall	46.2 %	69.3 %	47.3 %	24.1 %	15.4 %	12.1 %
	(19.8)	(6.4)	(5.9)	(8.4)	(3.4)	(4.4)
Democrat	47.2 %	69.6 %	48.3 %	26.4 %	16.7 %	14.3 %
	(19.5)	(6.3)	(6.3)	(10.1)	(3.8)	(5.0)
Republican	50.3 %	74.6 %	53.3 %	22.4 %	16.9 %	12.8 %
	(22.0)	(6.5)	(6.2)	(10.3)	(5.1)	(5.0)
Unaffiliated (No Party)	39.7 %	62.9 %	39.7 %	18.1 %	12.0 %	9.1 %
	(19.4)	(7.7)	(7.5)	(7.5)	(3.2)	(3.4)
Libertarian	29.4 %	54.0 %	27.7 %	6.4 %	5.2 %	4.6 %
	(19.2)	(8.4)	(6.3)	(4.2)	(1.9)	(2.1)
Age 18-35	27.6 %	52.0 %	23.9 %	10.2 %	4.8 %	3.5 %
	(18.2)	(8.3)	(6.0)	(5.8)	(1.7)	(1.3)
Age 36-60	48.5 %	73.3 %	49.9 %	23.6 %	16.7 %	12.9 %
	(21.1)	(6.3)	(6.1)	(9.1)	(4.6)	(5.0)
Age Over 60	61.4 %	80.4 %	66.1 %	38.4 %	29.3 %	24.5 %
	(19.2)	(4.6)	(4.9)	(10.3)	(6.3)	(9.3)
Female	47.2 %	70.8 %	48.2 %	25.3 %	16.0 %	12.2 %
	(20.1)	(6.3)	(6.2)	(9.1)	(3.7)	(4.3)
Male	46.1 %	68.2 %	47.9 %	24.0 %	15.3 %	12.4 %
	(19.6)	(6.8)	(6.2)	(8.4)	(3.3)	(4.6)
Hispanic	28.4 %	53.9 %	24.9 %	9.3 %	4.9 %	3.1 %
	(19.1)	(7.0)	(7.3)	(6.6)	(1.6)	(1.3)
White	49.5 %	72.0 %	51.6 %	26.6 %	17.8 %	14.3 %
	(20.3)	(6.9)	(7.0)	(9.3)	(4.3)	(5.5)
Black	41.4 %	63.7 %	41.5 %	21.5 %	13.7 %	8.6 %
	(18.7)	(6.6)	(5.4)	(9.3)	(4.0)	(3.1)
Asian/Pacific Islander	35.2 %	62.4 %	33.6 %	11.5 %	5.7 %	4.2 %
	(22.3)	(11.3)	(9.8)	(7.4)	(3.1)	(1.4)
Other Race	34.8 %	61.6 %	32.4 %	13.2 %	7.3 %	5.9 %
	(20.8)	(9.1)	(7.3)	(6.6)	(2.5)	(2.4)
N Registered	34,513,080	9,979,372	15,249,547	4,571,531	3,905,977	806,653
N Elections	560	155	230	132	39	4
Panel B. Washington (2013-2021)						
Overall	34.0 %	-	-	-	41.1 %	26.2 %
	(11.3)	-	-	-	(8.3)	(8.7)
Age 18-35	15.7 %	-	-	-	19.7 %	11.2 %
	(8.2)	-	-	-	(7.6)	(6.2)
Age 36-60	30.8 %	-	-	-	38.7 %	22.1 %
	(12.6)	-	-	-	(9.7)	(9.1)
Age Over 60	54.9 %	-	-	-	63.1 %	45.9 %
	(12.9)	-	-	-	(7.2)	(11.7)
Female	34.5 %	-	-	-	41.5 %	26.9 %
	(11.4)	-	-	-	(8.6)	(9.0)
Male	33.6 %	-	-	-	40.9 %	25.5 %
	(11.3)	-	-	-	(8.0)	(8.5)
N Registered	40,939,929	-	-	-	21,408,485	19,531,444
N Elections	2,728	-	-	-	1,475	1,253

Notes. Standard deviations shown in parentheses. Odd Year, Non-November refers to July, October, and December school board elections in NC and August primaries in WA. Turnout estimates are population weighted.

Table 2. *Rolloff in North Carolina School Board Election Voting*

	# Elections	# Total Voted for School Board	# Total Voted on Date	% Difference ("Rolloff")	School Board Election Turnout	School Board Election Date Turnout	Percentage Point Difference
All Vote for 1, At Large Races	75	2,454,429	2,676,624	-8.30	46.18	50.37	-4.18
On-Cycle Vote for 1, At Large Races	62	2,239,649	2,449,859	-8.58	50.29	55.02	-4.72
Off-Cycle Vote for 1, At Large Races	13	214,780	226,765	-5.29	24.93	26.33	-1.39

Notes. School board elections in this table are restricted to confirmed at-large contests where voters could only vote once (i.e., each ballot cast per voter could contain, at most, one vote for school board). Rolloff is the percentage of total voters on a given election date who did not vote in school board elections. School board election turnout refers to the number of voters reported on official returns divided by the total number of registered voters in the voter files. School board election date turnout is the number of voters reported in the individual vote history files divided by the total number of registered voters in the voter files. The percentage point difference is defined as school board election turnout minus school board election date turnout.

Table 3. *Validation Results: Convergent Validity*

	At-Large & County Only School Board Elections			SAIPE
	Turnout	Registered Voters	Actual Voters	Adult Population
Panel A. QOR Validity Checks				
QOR (0% Threshold)				
Correlation	0.99489*	0.99956*	0.99897*	0.99890*
Percentage Difference	-4.972	0.579	-5.121	-
	(3.388)	(4.072)	(4.302)	
Percentage Point Difference	-2.552	-	-	-
	(1.682)			
N Elections		96		560
QOR (5% Threshold)				
Correlation	0.99456*	0.99947*	0.99897*	0.99857*
Percentage Difference	-4.872	0.531	-5.121	-
	(4.201)	(5.209)	(4.488)	
Percentage Point Difference	-2.551	-	-	-
	(1.748)			
N Elections		96		560
Panel B. Convergent Validity Checks ("Similar To")				
Manual Assignment Panel (0% Threshold)				
Correlation	0.99488*	0.99966*	0.99902*	0.99294*
Percentage Difference	-4.917	0.696	-5.627	-
	(3.317)	(2.498)	(2.978)	
Percentage Point Difference	-2.532	-	-	-
	(1.697)			
N Elections		96		560
Manual Assignment Panel (5% Threshold)				
Correlation	0.99453*	0.99956*	0.99902*	0.99353*
Percentage Difference	-4.882	0.648	-5.627	-
	(4.372)	(3.990)	(3.121)	
Percentage Point Difference	-2.532	-	-	-
	(1.768)			
N Elections		96		560

Notes. * $p < 0.05$. Standard deviations in parentheses. In this table, the QOR and Manual panels are primarily compared against county-wide measures of unique ballots cast and total registered voters reported online by the North Carolina state Board of Elections. SAIPE adult population estimates are only compared against total registered voters and come from yearly school district-level measures produced by the Small Area Income and Population Estimates program of the U.S. Census. The QOR panel employs the QOR method, while the Manual panel leverages the fact that North Carolina school districts are embedded inside either counties or cities to match voters to a school district based on county/city of residence. The 0% and 5% thresholds refer to the precinct-by-district-by-date-level turnout rate used to prune voters assumed to be in wards with no applicable candidates on a given date.

Capturing School-District Level Voter Turnout in School Board Elections: Validating a Geospatial Strategy

Online Supplemental Appendices

The supplementary appendices provide in-depth explanations for technical details, coding decisions, and data cleaning strategies that were unnecessary for a general understanding of the QOR method as articulated in the main manuscript. Online Supplemental Appendix A describes the process by which we determine school district and county overlap using federal data sources. Supplemental Appendix B contains our search strategy to determine North Carolina school board election dates. Supplemental Appendix C encompasses technical documentation that illustrates the logic behind each function within our *QOR* package, mapping onto the corresponding sub-sections of the main manuscript. Supplemental Appendix D describes our approach to addressing ward elections, and Supplemental Appendix E contains the web-scraping procedure for obtaining North Carolina election returns. Supplemental Appendix F explains the zip code-only match to which we compare the QOR panels, and Supplemental Appendix G contains the table with OLS coefficients mentioned in the Results section.

Supplemental Appendix A: District-County Overlap

We cross-reference several federal data sources to construct the first estimates, to our knowledge, of the extent to which school districts align with county boundaries in every U.S. state. This exploration hones in on school year 2022-2023, which represents the final time period for our two-state school board election panel. We began by building two dictionaries that mapped state FIPS codes to two-letter state initials, and then mapped states to Census Regions.[1] Second, we identified every school district found in both the National Center for Education Statistics (NCES) shapefiles and the Geographic Relationship Files (GRF)[2] for school year 2022-2023, cycling through the list of school district NCES identification numbers on a state-by-state basis. Third, for each school district within each state, our algorithm undertook the following major procedures:

1. Calculating the total area (land area + water area) in miles squared and rounding the result to two decimal places.
2. Calculating a value equivalent to 0.5% of this district area as a small "tolerance" to allow for slight differences (e.g., rounding) between district areas and county areas as offered by Census or NCES records. We mainly use this tolerance level because some of our data are from the Census and some are from NCES, and we assumed that different agencies would be imperfectly (though closely) coordinated in coding decisions and calculations.

[1] We obtained FIPS-to-state name mappings from the Bureau of Labor Statistics (https://www.bls.gov/respondents/mwr/electronic-data-interchange/appendix-d-usps-state-abbreviations-and-fips-codes.html), and FIPS-to-Census Region mappings from the Census (https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf).
[2] We specifically use the NCES GRF that contains school district-county intersections, where every row in the dataset is one such intersection with district and county identifiers.

3. Filtering the GRF to just contain school district-county intersections featuring the focal school district.
    a. If there were multiple intersections, the district was immediately coded as crossing multiple counties (therefore lacking an exact county match).
4. Feeding the county identifier into the U.S. Census TIGER/LINE County Shapefiles from Jan. 1^{st} 2023, thereby filtering these shapes down to the focal county.
5. Calculating the single intersecting county's total area (land area + water area) in miles squared and rounding the result to two decimal places
6. Differencing the district and county areas, checking for an absolute value less than or equal to the calculated tolerance level unique to the district.
    a. If the difference satisfied the tolerance level, the district was coded as an exact match to the county in which it was contained.
    b. If not, the district was contained within one county alongside parts (however small) of other districts.

During this process we calculated the following quantities displayed in Supplemental Appendix Table A1: the number of total school districts per state, the number of districts in multiple counties per state, the number of districts in exactly one county per state, and the respective percentages of *total* districts per state that fell into either multiple counties or "exactly" (within the district-specific tolerance level) matched a single county's boundaries. We applied the Census Region mappings to calculate these percentage statistics for the Midwest, Northeast, South, and West regions, and then for the entire U.S. (summing totals across regions). Generally speaking, there are some important caveats to the NCES and Census data that may not be intuitive to all readers.[3] First, while the districts utilized in the analysis overwhelmingly include traditional public K-12 school districts, the NCES shapefiles also contain military and tribal districts as well as “pseudo-districts” which can be geographically non-contiguous despite sharing a common governance structure. We possess no a-priori method for distinguishing all of these districts in every state from public K-12 school districts but nevertheless think they are relevant to a discussion on the extent to which counties and school districts overlap. These districts help further illustrate the complex governance structures that implicate American education and its policy outcomes. Secondly, the NCES data contain both secondary and elementary school districts, which may have different boundaries depending on state law. Still other complexities may be found in the NCES documentation, beyond the scope of the paper.

[3] Detailed documentation can be found directly from NCES: https://nces.ed.gov/programs/edge/geographic/relationshipfiles#:~:text=As%20a%20result%2C%20school%20districts,intersect%20three%20different%20Congressional%20Districts.

Supplemental Appendix A: Tables and Figures

Table A1. *School Year 2022-2023 School District Misalignment with Counties, by U.S. State and Census Region*

Census Region	FIPS Code	State	Number of Total Districts	Num. Districts in Multiple Counties	Num. Districts Exactly Matching One County	% Districts in Multiple Counties	% Districts Exactly Matching One County
Midwest (50.57 % Multi-County) (0.55% Exactly One County)	17	IL	862	323	2	37.47	0.23
	18	IN	290	40	13	13.79	4.48
	19	IA	327	240	0	73.39	0
	20	KS	286	197	2	68.88	0.7
	26	MI	540	241	0	44.63	0
	27	MN	330	218	1	66.06	0.3
	29	MO	516	267	1	51.74	0.19
	31	NE	244	200	0	81.97	0
	38	ND	172	107	1	62.21	0.58
	39	OH	611	252	0	41.24	0
	46	SD	149	113	5	75.84	3.36
	55	WI	421	203	1	48.22	0.24
Northeast (14.23% Multi-County) (0.14% Exactly One County)	9	CT	166	3	0	1.81	0
	23	ME	267	25	0	9.36	0
	25	MA	303	11	0	3.63	0
	33	NH	179	9	0	5.03	0
	34	NJ	557	4	0	0.72	0
	36	NY	680	240	0	35.29	0
	42	PA	500	88	3	17.6	0.6
	44	RI	36	1	0	2.78	0
	50	VT	180	27	1	15	0.56
	1	AL	141	19	24	13.48	17.02

	5	AR	234	112	1	47.86	0.43
	10	DE	16	3	0	18.75	0
	12	FL	67	0	67	0	100
	13	GA	184	17	128	9.24	69.57
	21	KY	177	6	78	3.39	44.07
South (26.03% Multi-County) (25.51% Exactly One County)	22	LA	69	0	61	0	88.41
	24	MD	24	0	24	0	100
	28	MS	137	9	39	6.57	28.47
	37	NC	118	4	81	3.39	68.64
	40	OK	509	229	0	44.99	0
	45	SC	77	8	29	10.39	37.66
	47	TN	157	4	63	2.55	40.13
	48	TX	1019	399	21	39.16	2.06
	51	VA	136	2	125	1.47	91.91
	54	WV	55	0	55	0	100
	2	AK	53	8	22	15.09	41.51
	4	AZ	215	7	0	3.26	0
	6	CA	978	93	4	9.51	0.41
	8	CO	178	60	7	33.71	3.93
West (16.22% Multi-County) (3.66% Exactly One County)	15	HI	1	1	0	100	0
	16	ID	115	53	5	46.09	4.35
	30	MT	399	44	9	11.03	2.26
	32	NV	17	0	17	0	100
	35	NM	89	27	3	30.34	3.37
	41	OR	197	52	1	26.4	0.51
	49	UT	41	1	21	2.44	51.22
	53	WA	295	71	0	24.07	0
	56	WY	48	9	7	18.75	14.58

Total	-	-	13,362	4,047	922	30.29	6.9

Note. Data reflect School Year 2022-2023 (TIGER/LINE Year 2023) Geographic Relationship Files and School District Shapefiles from the National Center for Education Statistics (NCES), and County Shapefiles from the U.S. Census. The analysis includes any school districts tracked by NCES, which can cover elementary, secondary, unified, tribal, and military-run districts; some of these districts do not have their own elected school boards or are run by tribal governments and the U.S. military. We treat districts and counties as having the "exact" same boundaries if their combined land and water areas are either equal or within a tolerance bound of 0.5% of the school district's total area in miles squared. Readers should be aware that NCES makes some coding decisions when building school district geometries and classifying districts which may not be immediately intuitive. We highlight here that the federal data treat all of Hawaii as one school district, and all of New York City schools as one district, while also lumping together the boundaries of some "pseudo-districts" which are governed by the same legal entity but may be scattered throughout a given state. Additionally, NCES uses approximate county boundaries for Connecticut based on relevant administrative structures in that state. See the official documentation for additional details and other complexities in the federal data: https://nces.ed.gov/programs/edge/geographic/relationshipfiles#:~:text=As%20a%20result%2C%20school%20districts,intersect%20three%20different%20Congressional%20Districts.

Supplemental Appendix B: Obtaining Election Dates

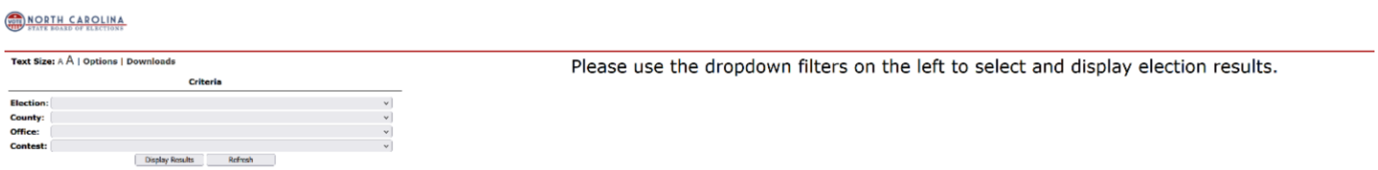
Since voters within state voter files are simply flagged as having voted on a specific election date, it is not possible to immediately know which dates correspond to *school board elections* specifically when certain states do not impose uniform board election schedules by law. Our solution in North Carolina involved manually working through the state Board of Elections website, pictured as Supplemental Appendix Figure B1, and compiling a spreadsheet listing each district and its election dates for elections where we confirmed that at least one candidate won a seat on the school board (e.g., excluding non-consequential primaries). The centralized database allows for selecting a date in the "Election" box, after which users can filter by county, the level of office that describes the election, and then specific contest names. With help from a research assistant, we exhaustively queried this database starting from the first possible date in 2013 and continuing through the last possible date in 2022. We were able to distinguish between consequential and non-consequential primaries by noting if a district had multiple board races in the same year, seeing which candidates carried over to the next race and whether their vote share plausibly would have allowed them to win in the earlier election, and then using Google to find winners in ambiguous cases. This strategy revealed that some board primary elections produced one or more winners while still triggering runoff elections among a subset of the other candidates.

The compiled dates, in practice, serve to prune the North Carolina voter files after the QOR method matches individuals to their school district each year. On a year-by-year basis, we merge information on which school districts had board elections on specific dates to the individual voter files— retaining only the individual voters who lived in a school district with an active election that year. This allows us to the aggregate individuals' behavior and demographic characteristics to the school district-by-date level, as at this point we know (1) that they were in a school district with elections that date, (2) on which date we should examine their vote history, and (3) whether or not each individual voted on the correct election date for each school board election corresponding to their district.

Supplemental Appendix B: Tables and Figures

Figure B1

North Carolina Board of Elections "Elections Dashboard"



Supplemental Appendix C: Programming Logic

Appendix C is a technical supplement to the conceptual overview of each QOR stage. We provide additional context and mechanical features from the package which may not be obvious to most readers. QOR needs to draw upon specific data structures in the R coding language, as well as several existing packages, to achieve the aims discussed throughout the main manuscript. The final QOR North Carolina and Washington panels that we present in figures and tables were constructed using scripts run from a *conda* (https://docs.conda.io/en/latest/) environment that installed R version 4.2.0, which was the most recent R version compatible with the cloud computing system of the lead author's university. We deployed QOR on the university's Linux-based systems and relied upon *sf* (Pebesma, 2018)*, magrittr* (Bache & Wickham, 2025), *tidyverse* and *dplyr* (Wickham et al., 2019), *tidygeocoder* (Cambon et al., 2021), and *tictoc* (Izrailev, 2024) as the required dependencies to achieve full functionality with the original code. We also find the *patchwork*, *ggthemes*, and *haven* packages to be useful for creating and saving maps/graphs. Our preferred *conda* environment for executing the project was installed via the following Linux terminal command after setting up the environment, practically suggesting that package versions for the dependencies should generally be selected to work around our chosen version of R to achieve the best results:

conda install conda-forge::r-base=4.2.0 conda-forge::r-tidyverse conda-forge::r-dplyr conda-forge::r-tidygeocoder conda-forge::r-ggthemes conda-forge::r-sf conda-forge::r-patchwork conda-forge::r-haven conda-forge::r-tictoc

The remainder of Appendix C provides addenda to help interpret the code behind QOR and associate the conceptual descriptions with their mechanical execution:

Step 1: Query

In practice the *Query* function acts as a wrapper around *tidygeocoder* (Cambon et al., 2021), but is modified in several important ways to tailor its operations for geocoding voter addresses. First, our function automatically divides large voter registration lists into smaller, user-determined batches compatible with several geocoding Application Programming Interfaces (API's).[4] API's serve as intermediaries between two or more different programs. Here, they functionally refer to the public-facing interface of the Census (or an alternative) geocoding service that receives our geocoding requests (batches of addresses) and returns appropriate responses (geocoded coordinates) back to our program. Secondly, we embed the geocoding process in error handling code that catches internet connection failures and waits for re-connection to the API to prevent lost progress. Third, we automatically break "tied" geocoding results by re-submitting indeterminate observations from batch geocoding as individual entries that explicitly request the Census to only return one match. Census documentation suggests that ties are broken arbitrarily by choosing the first match in the return order. Lastly, our function separates out voters who can and cannot be located within the Census database and sorts them into different objects that are each returned separately from the function.

Step 2: Overlay

Mechanically, the voter-district match employs an R-language *list of lists* built from all possible intersections between (1) a dataset containing voter addresses, unique voter identifiers, and the location coordinates identified by the Census Geocoder, and (2) a second dataset encompassing NCES school district boundary geometries and metadata associated with every school district from the focal state in a given year. The intersections of these two datasets are

[4] *Query* can be modified to work with different geocoding services, but we strongly prefer and exclusively use the Census Geocoding Tool given that it is public, free, and allows for batch geocoding jobs. If the Census Geocoding Tool becomes unavailable for some reason, users can use paid geocoding services that also work with *tidygeocoder*.

compiled in a list object, whereby each voter is represented as an item within the outer list, and each voter's item then contains a number representing the school district in which that voter is located.[5] We thus use the index that corresponds to each voter to extract a number representing the school district in which they are located. This district number corresponds to a row in the dataset of all school districts from the focal state.[6] We ultimately locate a voter's correct school district NCES identifier by returning the "TRUE" Boolean match when the re-identified row number associated with each voter's intersecting school district polygon is fed back into the original school district dataset. At the end, each voter is assigned to the school district whose boundaries contain that voter's registration address.

Step 3: Recover

Recall that our *Recover* procedure (matching voters to districts based upon zip codes) minimizes the ellipsoidal distance between district internal points and zip code centroids. We use school district internal points extracted via the *sf* st_point_on_surface() function to anticipate and accommodate highly irregular school district shapes. Some school districts may take on complex physical forms, like crescent moon shapes, where the actual center point (centroid) could fall

---

[5] We mainly take advantage of the *sf* st_intersects() function (Pebesma, 2018; Pebesma & Bivand, 2023) to identify geometry intersections while preserving the row numbers of voters and school districts from their respective original datasets. This allows us to cross-communicate between the original datasets and the new object containing numerical representations of intersections between voters and school districts.

[6] We use the row numbers to avoid the computationally inefficient process of filtering and sorting the datasets, which contain between 3.8 million and 7.4 million voters for each state-year combination. The data are saved and used in a consistent ordering without any sorting that could shift the row numbers. Our method is also robust to possible edge cases which may disrupt the ideal matching procedure. For one, some voters may be located directly on a boundary between school districts, therefore intersecting two or more different school district geometries. We accommodate such cases when they arise by minimizing the distance between the impacted voter's coordinates and each intersecting school district's internal point, selecting the closest such district. Secondly, it is possible that some voters may be successfully located by the Census Tool but still fall outside the official school district boundaries from NCES. We assume throughout the project that all government data is authoritative, but even the best available data may contain small errors (e.g., some literal corner cutting on shapefile boundaries). Thus, successfully geocoded voters not located in any school district are simply assigned the closest overall school district based on district internal points. This is functionally similar to the first accommodation strategy, but uses all school districts rather than restricting the distance minimization to districts that intersect the impacted voter's coordinates.

outside the district boundaries. This makes internal points preferred from the standpoint of face validity, as they approximate the center point while restricting the final coordinates to be within the bounds of the district polygon.[7]

Match Rates

QOR as implemented for North Carolina and Washington state matches essentially all voters fed into the yearly process with a school district. Table C1 calculates these match rates by year and state to demonstrate our ability to identify nearly all "raw" (input) voters from individual registration records as either belonging to or, in the case of *Recover*, near a relevant school district.

[7] For additional information see the following discussion about the utility of st_point_on_surface() from the *sf* package's "issues" section on Github (https://github.com/r-spatial/sf/issues/1302).

Supplemental Appendix C: Tables and Figures

Table C1. *Match Rates by State and Year*

Year	Raw Voters (Input)	Num Coordinate Matched	Percent Coordinate Matched	Num Zip Matched	Percent Zip Matched	Num With Successful Match	Percent with Successful Match
Panel A. North Carolina (Elections in 2013-2022)							
2014	6486956	6310480	97.28	162741	2.51	6473221	99.79
2015	6550942	6375901	97.33	160495	2.45	6536396	99.78
2016	6437375	6247716	97.05	173795	2.70	6421511	99.75
2017	6835298	6646884	97.24	172231	2.52	6819115	99.76
2018	6841814	6635426	96.98	189397	2.77	6824823	99.75
2019	7143136	6923994	96.93	201898	2.83	7125892	99.76
2020	6841662	6645789	97.14	189865	2.78	6835654	99.91
2021	7391091	7136112	96.55	250707	3.39	7386819	99.94
2022	7205091	6931398	96.20	268661	3.73	7200059	99.93
2023	7466092	7130621	95.51	330024	4.42	7460645	99.93
Sum	69199457	66984321	96.80	2099814	3.03	69084135	99.83
Panel B. Washington (2013-2021)							
2013	3892404	3830683	98.41	61514.00	1.58	3892197	99.99
2015	3972319	3907767	98.37	64376.00	1.62	3972143	100.00
2017	4243339	4170319	98.28	72821.00	1.72	4243140	100.00
2019	4504157	4413128	97.98	90799.00	2.02	4503927	99.99
2021	4813424	4689476	97.42	123835.00	2.57	4813311	100.00
Sum	21425643	21011373	98.07	413345	1.93	21424718	100.00

Note. North Carolina voter snapshots are taken Jan. 1st of the year after the election, which tends to be either temporally closer or equally close to Spring and Fall elections than Jan. 1st of the same year.

Supplemental Appendix D: Ward Elections

The ward issue raised in the main manuscript involves two main components: (1) understanding the extent to which voters may be confined to wards with no possible elections, and (2) identifying these cases for removal when building the final district-by-date level election panel. We chose North Carolina as the focal case for intensive manual research, consistent with the rest of the paper. Here, we turned our attention to also collecting election *structures* and reproduce, via Appendix Table D1, a school board election structure resource first compiled by North Carolina state senator Julie Mayfield's office.[8] This resource attempted to categorize school boards as entirely at-large, entirely elected by "district" (meaning ward), or a mixture of these two structures. We then manually checked the ward-based elections implicated by this list with help from a research assistant, searching the centralized North Carolina state Board of Elections website to determine if all voters in these school districts had the opportunity to vote in at least one other election on a given date (see Appendix Table D2). For the purposes of this exercise, we did not flag board elections with co-occurring city- or county-wide elections, statewide primaries, and statewide general elections as ward-based, given that all voters could participate in some election on these dates. We eventually determined that only the Hickory City Schools election in 2021 appeared to truly suffer from the issue of some ward-based voters having *no* election in which to participate.

Due to this background research, we are reasonably confident that, in North Carolina, there is at least one model "ward-based" district that can be compared against model at-large

[8] We found the list from local news (https://www.asheville.com/news/2023/06/new-law-spells-the-end-of-at-large-school-board-voting-in-buncombe-county/), and then copied the state senator's spreadsheet in September of 2023 (https://docs.google.com/spreadsheets/d/1oAwktlQU1p8861y8dxMoo_MA3zEStFtR/edit?gid=1282000746#gid=1282000746).

districts where all voters could participate in the school board election. We proceeded to combine our manual data collection and web-scraping strategies to develop a generalizable approach for addressing the ward-based election issue. We specifically extracted from our web-scraped election return data (see Online Supplemental Appendix E below) all election date-school district identifier pairs explicitly labelled "at-large" by the state of North Carolina in their contest names. Combining these flags with the individual level voter files yields a rare systematic attempt to solve the calculation problem for turnout estimates in ward-based school board elections.

In essence, we argue that voters in wards without any elections (at any level) on a specific date should be visible from the distribution of turnout rates at some sub-district level (e.g., precinct or similar unit). If many voters truly cannot vote on certain school board election dates due to a ward-based system, then their data should exhibit bunching around zero turnout that violates a normal distribution for sub-district turnout. We argue that voters in the sub-district areas that exhibit abnormally low turnout should be pruned from the data, as they likely were not eligible to vote in any election on the impacted school board election dates. Failing to prune these voters would bias turnout estimates downward by overestimating the number of registered voters eligible to vote. Our raw data provide precinct codes that we can use to calculate sub-district turnout rates conducive to examining this proposition. Practically speaking, after matching individual voters to school districts, we place individual voters in the part of each precinct that overlaps with their school district by calculating turnout rates at the school district-by-precinct-by-date level— approximating turnout below the district level, but crucially still *within* the district. Knowing the election structure of specific North Carolina races then allows us to compare the school district-by-precinct-by-date turnout across areas that operated under a ward-based structure (i.e., areas within Hickory City Schools in the year 2021), against school district-by-precinct-by-date turnout for areas that operated under a confirmed at-large structure (in this case, all elections literally named “at-large”).

Although the detailed nature of North Carolina’s election reporting is best suited to identifying wards, the notion of school district-by-precinct-by-date turnout is likewise useful for ensuring data quality in Washington. The Washington school board election panel should also contain ward-based elections likely to be noticeable from sub-district turnout distributions in both primary and general elections. However, analyzing these distributions is relevant to another facet of Washington state law, which allows school board primary elections to be cancelled if no more than two candidates file (Lindell, 2017). This poses the challenge that even one voter being misplaced by the Census geocoder into a district with a cancelled primary could generate an improper primary election observation. This is not an issue when manually collecting school board election dates, as in North Carolina, but the Washington election dates rely on a known uniform election schedule throughout the state. We argue that any “bad” primary observations should exhibit abnormally low school district-by-precinct-by-date level turnout that bunches around zero, similar to any ward-based elections that likewise lacked actual vote choices for certain district residents.

Panel A of Appendix Figure D1 analyzes the school district-by-precinct-by-date turnout rate distributions in each state, with distributions separated out by known election characteristics. We label these as “Precinct Turnout” turnout rates for brevity, as we suspect that these areas are likely still close to representing precincts, though they actually reflect school district-by-precinct-by-date observations. The North Carolina graph plots kernel densities of sub-district turnout rates for confirmed “at-large” district-by-precinct-by-date areas, the “Hickory21” district-by-precinct-by-date areas, and then all other district-by-precinct-by-date areas whose statuses we could not confirm due to either ambiguous or not "at-large" official names on the web-scraped election returns. For Washington, we plot district-by-precinct-by-date turnout distributions separately for primary and general elections to address both ward elections and geocoding error.

As noted in the main manuscript, we find compelling evidence *in both states* that the most obvious bunching occurs between 0% and 5% district-by-precinct-by-date (sub-district) level turnout as expected. The left tail of turnout in North Carolina's confirmed "at-large" sub-districts seems to exhibit no bunching, justifying our proposition that ward-based voters with no active elections can be pruned using a school district-by-precinct-by-date level turnout threshold. These findings are also supported by the Washington general election data. The more extreme left tail for primary election turnout is explainable by both ward-based voters and the presence of some primaries being cancelled due to lack of competition. Complementing these, Figure D1 Panel B demonstrates that dropping voters from sub-districts with either 0% or 5% turnout removes only a small fraction (in North Carolina, less than 5%) of the individuals registered to vote in any given school district-by-election date observation.[9]

We use the school district-by-precinct-by-date turnout rates to remove the portion of voters in each school district-by-date who live in a ward with no actual election, while keeping voters in wards that do have valid elections. Thus, we assume that dropping voters in sub-district areas with either 0% or 5% turnout is a prudent strategy and the potential to improve the validity of our data exceeds the possible risks. Other alternatives could involve fully dropping elections that have any wards at all, which unnecessarily throws out data, or ignoring the ward dilemma

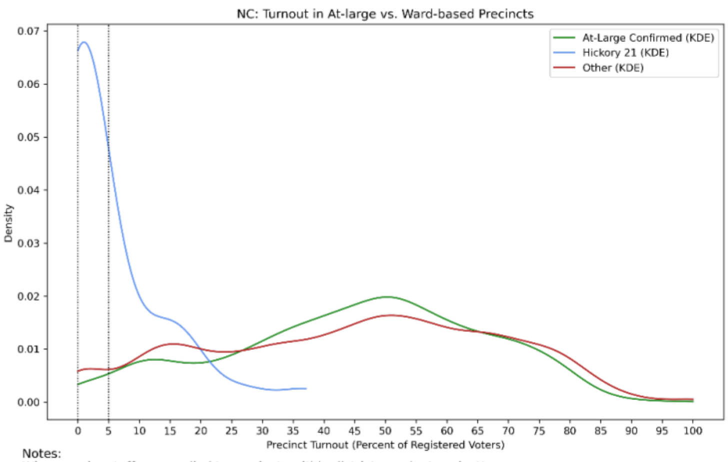
[9] Panel B's Washington graph shows that both a 0% and a 5% cutoff remove all or nearly all voters for some district-by-date observations, but this is expected given the presence of cancelled primaries that we do indeed want to remove from the dataset. Our preferred Washington dataset applies the 0% threshold, and we also drop district-by-election observations with 5 or fewer voters in them. We assume these observations with incredibly few voters are also likely cancelled primaries only "identified" due to small geocoding errors (misplaced voters).

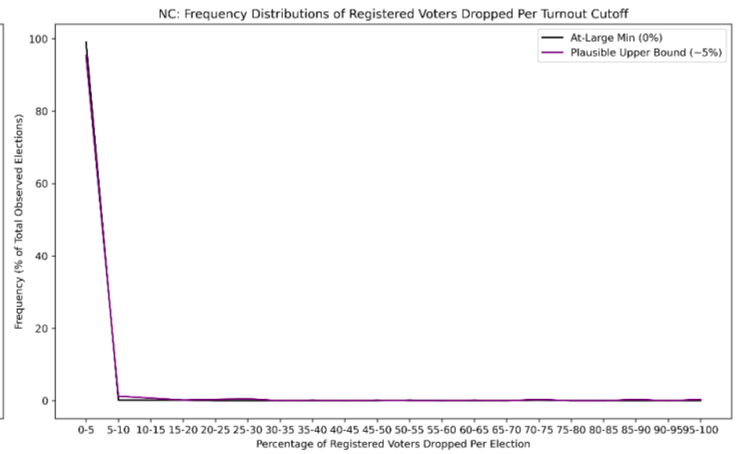
and keeping-in some voters who had no valid election. We prefer our strategy as an evidence-informed path forward, especially using the conservative 0% threshold. Further, we only apply the school district-by-precinct-by-date threshold for voter removal in elections that are off-cycle with Midterm or Presidential contests, as all individuals registered to vote could have voted in Midterms or Presidential elections.

Supplemental Appendix D: Tables and Figures

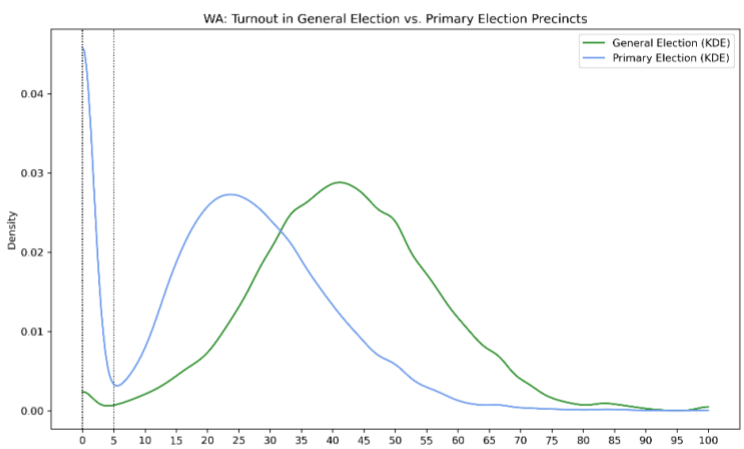
Appendix Figure D1. *Sub-district Turnout Diagnostic Figures for Ward Pruning Strategy*

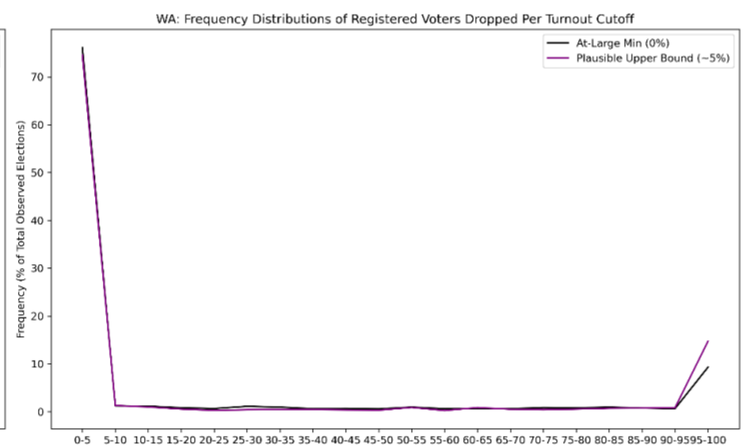
Panel A. Turnout Distributions **Panel B. Amount of Data Pruned Under Cutoff**





Notes:
Where each cutoff was applied to precincts within districts, and a 'precinct'
is the unique combination of leaid, precinct_abbrv, and precinct_desc.
The entire graph drops precincts that are missing both an abbrv and desc value,
and also omit precincts with 5 or fewer registered voters.
Vertical lines show plausible turnout cutoffs: 0% or 5%.





Notes:
Where each cutoff was applied to precincts within districts, and a 'precinct'
is the unique combination of leaid, precinctcode, and precinctpart.
Both graphs drop precincts that are missing both a code and a part value,
and also omit precincts with 5 or fewer registered voters.
Vertical lines show plausible turnout cutoffs: 0% or 5%.

Appendix Table D1. *Mayfield Compilation of Ward-based School Boards*

County	School Board Voting Method	Party Voted for in 2008 Presidential election	Urban/ Rural	Population (2021)
Alamance	At-large	Republican	Urban	173,877
Alleghany	At-large	Republican	Rural	11,049
Ashe	At-large	Republican	Rural	26,711
Avery	At-large	Republican	Rural	17,864
Bertie	At-large	Democrat	Rural	17,505
Brunswick	At-large	Republican	Urban	144,215
Buncombe	At-large	Democrat	Urban	271,534
Burke	At-large	Republican	Urban	87,611
Caldwell	At-large	Republican	Urban	80,463
Camden	At-large	Republican	Rural	10,835
Catawba	At-large	Republican	Urban	161,723
Chatham	At-large	Democrat	Urban	77,889
Cherokee	At-large	Republican	Rural	29,167
Clay	At-large	Republican	Rural	11,309
Cleveland	At-large	Republican	Rural	100,359
Columbus	At-large	Republican	Rural	50,092
Craven	At-large	Republican	Urban	100,674
Davidson	At-large	Republican	Urban	170,637
Davie	At-large	Republican	Urban	43,533
Duplin	At-large	Republican	Rural	48,515
Franklin	At-large	Republican	Rural	71,703
Gaston	At-large	Republican	Rural	230,856
Gates	At-large	Democrat	Rural	10,366
Graham	At-large	Republican	Rural	8,043
Greene	At-large	Republican	Rural	20,417

Haywood	At-large	Republican	Urban	62,476
Henderson	At-large	Republican	Urban	116,829
Hertford	At-large	Democrat	Rural	21,278
Hoke	At-large	Democrat	Urban	53,114
Hyde	At-large	Democrat	Rural	4,508
Jackson	At-large	Democrat	Rural	43,410
Johnston	At-large	Republican	Urban	226,504
Jones	At-large	Republican	Rural	9,255
Lee	At-large	Republican	Rural	64,138
Lincoln	At-large	Republican	Rural	89,670
Macon	At-large	Republican	Rural	37,564
McDowell	At-large	Republican	Rural	44,717
Mitchell	At-large	Republican	Rural	14,963
Moore	At-large	Republican	rural	102,763
New Hanover	At-large	Republican	Urban	229,018
Northampton	At-large	Democrat	Rural	17,129
Onslow	At-large	Republican	Urban	206,160
Orange	At-large	Democrat	Urban	148,884
Pasquotank	At-large	Democrat	Rural	40,821
Pender	At-large	Republican	Rural	62,815
Perquimans	At-large	Republican	Rural	13,130
Person	At-large	Republican	Rural	39,127
Polk	At-large	Republican	Rural	19,656
Randoplph	At-large	Republican	Rural	145,172
Richmond	At-large	Democrat	Rural	42,724
Rowan	At-large	Republican	Rural	148,150
Rutherford	At-large	Republican	Rural	64,586
Sampson	At-large	Republican	Rural	58,990
Scotland	At-large	Democrat	Rural	34,227
Stanly	At-large	Republican	Rural	63,425

Stokes	At-large	Republican	Urban	44,553
Surry	At-large	Republican	Rural	71,152
Swain	At-large	Republican	Rural	14,136
Transylvania	At-large	Republican	Rural	33,165
Tyrrell	At-large	Republican	Rural	3,254
Warren	At-large	Democrat	Rural	18,762
Washington	At-large	Democrat	Rural	10,892
Watauga	At-large	Democrat	Rural	54,234
Wilkes	At-large	Republican	Rural	65,806
Yadkin	At-large	Republican	Rural	37,192
Yancey	At-large	Republican	Rural	18,757
Bladen	Both	Democrat	Rural	29,525
Carteret	Both	Republican	Rural	68,541
Caswell	Both	Democrat	Rural	22,714
Chowan	Both	Republican	Rural	13,722
Cumberland	Both	Democrat	Urban	335,508
Currituck	Both	Republican	Rural	29,653
Durham	Both	Democrat	Urban	326,126
Forsyth	Both	Democrat	Urban	385,523
Guilford	Both	Democrat	Urban	542,410
Halifax	Both	Democrat	Rural	48,272
Lenoir	Both	Republican	Rural	54,706
Madison	Both	Republican	Rural	21,502
Mecklenburg	Both	Democrat	Urban	1,120,000
Pamlico	Both	Republican	Rural	12,344
Pitt	Both	Democrat	Urban	172,169
Robeson	Both	Democrat	Rural	116,328
Rockingham	Both	Republican	Rural	91,266
Union	Both	Republican	Urban	243,648
Wayne	Both	Republican	Urban	116,835

Alexander	District	Republican	Rural	36,644
Anson	District	Democrat	Rural	22,060
Beaufort	District	Republican	Rural	4,569
Caburrus	District	Republican	Rural	231,278
Dare	District	Republican	Rural	37,826
Edgecombe	District	Democrat	Urban	48,359
Granville	District	Democrat	Rural	61,986
Harnett	District	Republican	Rural	135,966
Iredell	District	Republican	Urban	191,968
Martin	District	Democrat	Rural	21,754
Montgomery	District	Republican	Rural	25,798
Nash	District	Republican	Urban	95,176
Vance	District	Democrat	Rural	42,185
Wake	District	Democrat	Urban	1,150,000
Wilson	District	Democrat	Rural	78,369

Note. Sourced via news reporting from Asheville.com (https://www.asheville.com/news/2023/06/new-law-spells-the-end-of-at-large-school-board-voting-in-buncombe-county/), and original spreadsheet from the office of North Carolina State Senator Julie Mayfield (https://docs.google.com/spreadsheets/d/1oAwktlQU1p8861y8dxMoo_MA3zEStFtR/edit?gid=1282000746#gid=1282000746)

Appendix Table D2 . *Manual Checks for Additional Races*

School	Leaid	County	Election Date	Was there an "at-large"/ general election this day?
Hickory City Schools	3702190	catawaba	11/2/2021	no
Newton Conover City Schools	3703360	catawaba	11/3/2015	yes
Hickory City Schools	3702190	catawaba	11/7/2017	yes
Halifax County Schools	3701950	halifax	5/6/2014	yes
Kannapolis City Schools	3702430	cabarrus	5/6/2014	yes
Alexander County Schools	3700090	alexander	11/4/2014	yes
Anson County Schools	3700180	anson	11/4/2014	yes
Asheville City Schools	3700270	buncombe	11/4/2014	yes
Beaufort County Schools	3700330	beaufort	11/4/2014	yes
Bladen County Schools	3700390	bladen	11/4/2014	yes
Cabarrus County Schools	3700530	cabarrus	11/4/2014	yes
ChowanEdenton Schools	3700840	chowan	11/4/2014	yes
Cumberland County Schools	3700011	cumberland	11/4/2014	yes
Currituck County Schools	3701080	currituck	11/4/2014	yes
Elkin City Schools	3701380	surry	11/4/2014	yes
Forsyth Winston Salem County Schools	3701500	forsyth	11/4/2014	yes
Guilford County Schools	3701920	guilford	11/4/2014	yes
Harnett County Schools	3702010	Harnett	11/4/2014	yes
Lenoir County Public Schools	3702610	lenoir	11/4/2014	yes
Lexington City Schools	3702640	davidson	11/4/2014	yes
Martin County Schools	3702880	martin	11/4/2014	yes
Montgomery County Schools	3703060	montgomery	11/4/2014	yes
Mount Airy City Schools	3703210	surry	11/4/2014	yes

Nash-Rocky Mount Schools	3703270	nash	11/4/2014	yes
Pitt County Schools	3700012	Pitt	11/4/2014	yes
Rockingham County Schools	3703990	rockingham	11/4/2014	yes
Union County Public Schools	3704620	union	11/4/2014	yes
Vance County Schools	3704650	Vance	11/4/2014	yes
Wayne County Public Schools	3704880	wayne	11/4/2014	yes
Weldon City Schools	3704890	halifax	11/4/2014	yes
Whiteville City Schools	3704920	columbus	11/4/2014	yes
Wilson County Schools	3705020	wilson	11/4/2014	yes
Asheboro City Schools	3700240	randolph	11/3/2015	yes
Chapel Hill-Carrboro City Schools	3700720	orange	11/3/2015	yes
Roanoke Rapids City Schools	3703900	halifax	11/3/2015	yes
Charlotte-Mecklenburg Schools	3702970	mecklenburg	11/3/2015	yes
Carteret County Public Schools	3700630	carteret	3/15/2016	yes
Caswell County Schools	3700660	caswell	3/15/2016	yes
Clinton City Schools	3700930	sampson	3/15/2016	yes
Dare County Schools	3701110	dare	3/15/2016	yes
Durham Public Schools	3701260	durham	3/15/2016	yes
Edgecombe County Public Schools	3701320	edgecombe	3/15/2016	yes
Granville County Schools	3701800	granville	3/15/2016	yes
Halifax County Schools	3701950	halifax	3/15/2016	yes
Kannapolis City Schools	3702430	cabarrus	3/15/2016	yes
Madison County Schools	3702820	madison	3/15/2016	yes
Pamlico County Schools	3703510	pamlico	3/15/2016	yes
Public Schools of Robeson County	3703930	Robeson	3/15/2016	yes
Alexander County Schools	3700090	alexander	11/8/2016	yes
Anson County Schools	3700180	anson	11/8/2016	yes
Asheville City Schools	3700270	buncombe	11/8/2016	yes
Beaufort County Schools	3700330	beaufort	11/8/2016	yes
Bladen County Schools	3700390	bladen	11/8/2016	yes

Cabarrus County Schools	3700530	cabarrus	11/8/2016	yes
ChowanEdenton Schools	3700840	chowan	11/8/2016	yes
Cumberland County Schools	3700011	cumberland	11/8/2016	yes
Currituck County Schools	3701080	currituck	11/8/2016	yes
Elkin City Schools	3701380	surry	11/8/2016	yes
Guilford County Schools	3701920	guilford	11/8/2016	yes
Harnett County Schools	3702010	Harnett	11/8/2016	yes
Iredell-Statesville Schools	3702310	iredell	11/8/2016	yes
Lenoir County Public Schools	3702610	lenoir	11/8/2016	yes
Lexington City Schools	3702640	davidson	11/8/2016	yes
Martin County Schools	3702880	martin	11/8/2016	yes
Montgomery County Schools	3703060	montgomery	11/8/2016	yes
Mount Airy City Schools	3703210	surry	11/8/2016	yes
Nash-Rocky Mount Schools	3703270	nash	11/8/2016	yes
Pitt County Schools	3700012	Pitt	11/8/2016	yes
Rockingham County Schools	3703990	rockingham	11/8/2016	yes
Union County Public Schools	3704620	union	11/8/2016	yes
Vance County Schools	3704650	Vance	11/8/2016	yes
Wake County Schools	3704720	wake	11/8/2016	yes
Wayne County Public Schools	3704880	wayne	11/8/2016	yes
Weldon City Schools	3704890	halifax	11/8/2016	yes
Whiteville City Schools	3704920	columbus	11/8/2016	yes
Wilson County Schools	3705020	wilson	11/8/2016	yes
Asheboro City Schools	3700240	randolph	11/7/2017	yes
Chapel Hill-Carrboro City Schools	3700720	orange	11/7/2017	yes
Roanoke Rapids City Schools	3703900	halifax	11/7/2017	yes
Charlotte-Mecklenburg Schools	3702970	mecklenburg	11/7/2017	yes
Caswell County Schools	3700660	caswell	5/8/2018	yes
Clinton City Schools	3700930	sampson	5/8/2018	yes
Durham Public Schools	3701260	durham	5/8/2018	yes

Edgecombe County Public Schools	3701320	edgecombe	5/8/2018	yes
Granville County Schools	3701800	granville	5/8/2018	yes
Halifax County Schools	3701950	halifax	5/8/2018	yes
Kannapolis City Schools	3702430	cabarrus	5/8/2018	yes
Pamlico County Schools	3703510	pamlico	5/8/2018	yes
Public Schools of Robeson County	3703930	Robeson	5/8/2018	yes
Granville County Schools	3701800	granville	6/26/2018	yes
Alexander County Schools	3700090	alexander	11/6/2018	yes
Anson County Schools	3700180	anson	11/6/2018	yes
Asheville City Schools	3700270	buncombe	11/6/2018	yes
Beaufort County Schools	3700330	beaufort	11/6/2018	yes
Bladen County Schools	3700390	bladen	11/6/2018	yes
Carteret County Public Schools	3700630	carteret	11/6/2018	yes
ChowanEdenton Schools	3700840	chowan	11/6/2018	yes
Cumberland County Schools	3700011	cumberland	11/6/2018	yes
Currituck County Schools	3701080	currituck	11/6/2018	yes
Dare County Schools	3701110	dare	11/6/2018	yes
Elkin City Schools	3701380	surry	11/6/2018	yes
Forsyth Winston Salem County Schools	3701500	forsyth	11/6/2018	yes
Guilford County Schools	3701920	guilford	11/6/2018	yes
Harnett County Schools	3702010	Harnett	11/6/2018	yes
Iredell-Statesville Schools	3702310	iredell	11/6/2018	yes
Lenoir County Public Schools	3702610	lenoir	11/6/2018	yes
Lexington City Schools	3702640	davidson	11/6/2018	yes
Madison County Schools	3702820	madison	11/6/2018	yes
Martin County Schools	3702880	martin	11/6/2018	yes
Montgomery County Schools	3703060	montgomery	11/6/2018	yes
Mount Airy City Schools	3703210	surry	11/6/2018	yes
Nash-Rocky Mount Schools	3703270	nash	11/6/2018	yes
Pitt County Schools	3700012	Pitt	11/6/2018	yes

Rockingham County Schools	3703990	rockingham	11/6/2018	yes
Union County Public Schools	3704620	union	11/6/2018	yes
Vance County Schools	3704650	Vance	11/6/2018	yes
Wake County Schools	3704720	wake	11/6/2018	yes
Wayne County Public Schools	3704880	wayne	11/6/2018	yes
Whiteville City Schools	3704920	columbus	11/6/2018	yes
Wilson County Schools	3705020	wilson	11/6/2018	yes
Cabarrus County Schools	3700530	cabarrus	11/8/2018	yes
Weldon City Schools	3704890	halifax	11/8/2018	yes
Asheboro City Schools	3700240	randolph	11/5/2019	yes
Chapel Hill-Carrboro City Schools	3700720	orange	11/5/2019	yes
Hickory City Schools	3702190	catawaba	11/5/2019	yes
Newton Conover City Schools	3703360	catawaba	11/5/2019	yes
Roanoke Rapids City Schools	3703900	halifax	11/5/2019	yes
Charlotte-Mecklenburg Schools	3702970	mecklenburg	11/5/2019	yes
Caswell County Schools	3700660	caswell	3/3/2020	yes
Durham Public Schools	3701260	durham	3/3/2020	yes
Edgecombe County Public Schools	3701320	edgecombe	3/3/2020	yes
Granville County Schools	3701800	granville	3/3/2020	yes
Halifax County Schools	3701950	halifax	3/3/2020	yes
Kannapolis City Schools	3702430	cabarrus	3/3/2020	yes
Pamlico County Schools	3703510	pamlico	3/3/2020	yes
Public Schools of Robeson County	3703930	Robeson	3/3/2020	yes
Alexander County Schools	3700090	alexander	11/3/2020	yes
Anson County Schools	3700180	anson	11/3/2020	yes
Asheville City Schools	3700270	buncombe	11/3/2020	yes
Beaufort County Schools	3700330	beaufort	11/3/2020	yes
Bladen County Schools	3700390	bladen	11/3/2020	yes
Cabarrus County Schools	3700530	cabarrus	11/3/2020	yes
Carteret County Public Schools	3700630	carteret	11/3/2020	yes

ChowanEdenton Schools	3700840	chowan	11/3/2020	yes
Clinton City Schools	3700930	sampson	11/3/2020	yes
Cumberland County Schools	3700011	cumberland	11/3/2020	yes
Currituck County Schools	3701080	currituck	11/3/2020	yes
Dare County Schools	3701110	dare	11/3/2020	yes
Elkin City Schools	3701380	surry	11/3/2020	yes
Forsyth Winston Salem County Schools	3701500	forsyth	11/3/2020	yes
Guilford County Schools	3701920	guilford	11/3/2020	yes
Harnett County Schools	3702010	Harnett	11/3/2020	yes
Iredell-Statesville Schools	3702310	iredell	11/3/2020	yes
Lenoir County Public Schools	3702610	lenoir	11/3/2020	yes
Lexington City Schools	3702640	davidson	11/3/2020	yes
Madison County Schools	3702820	madison	11/3/2020	yes
Martin County Schools	3702880	martin	11/3/2020	yes
Montgomery County Schools	3703060	montgomery	11/3/2020	yes
Mount Airy City Schools	3703210	surry	11/3/2020	yes
Nash-Rocky Mount Schools	3703270	nash	11/3/2020	yes
Newton Conover City Schools	3703360	catawaba	11/3/2020	yes
Pitt County Schools	3700012	Pitt	11/3/2020	yes
Rockingham County Schools	3703990	rockingham	11/3/2020	yes
Union County Public Schools	3704620	union	11/3/2020	yes
Vance County Schools	3704650	Vance	11/3/2020	yes
Wake County Schools	3704720	wake	11/3/2020	yes
Wayne County Public Schools	3704880	wayne	11/3/2020	yes
Weldon City Schools	3704890	halifax	11/3/2020	yes
Whiteville City Schools	3704920	columbus	11/3/2020	yes
Wilson County Schools	3705020	wilson	11/3/2020	yes
Asheboro City Schools	3700240	randolph	11/2/2021	yes
Chapel Hill-Carrboro City Schools	3700720	orange	11/2/2021	yes
Newton Conover City Schools	3703360	catawaba	11/2/2021	yes

Roanoke Rapids City Schools	3703900	halifax	11/2/2021	yes
Caswell County Schools	3700660	caswell	5/17/2022	yes
Halifax County Schools	3701950	halifax	5/17/2022	yes
Kannapolis City Schools	3702430	cabarrus	5/17/2022	yes
Public Schools of Robeson County	3703930	Robeson	5/17/2022	yes
Wake County Schools	3704720	wake	10/8/2013	yes
Carteret County Public Schools	3700630	carteret	5/6/2014	yes
Caswell County Schools	3700660	caswell	5/6/2014	yes
Clinton City Schools	3700930	sampson	5/6/2014	yes
Dare County Schools	3701110	dare	5/6/2014	yes
Durham Public Schools	3701260	durham	5/6/2014	yes
Edgecombe County Public Schools	3701320	edgecombe	5/6/2014	yes
Granville County Schools	3701800	granville	5/6/2014	yes
Iredell-Statesville Schools	3702310	iredell	5/6/2014	yes
Madison County Schools	3702820	madison	5/6/2014	yes
Pamlico County Schools	3703510	pamlico	5/6/2014	yes
Public Schools of Robeson County	3703930	Robeson	5/6/2014	yes
Hickory City Schools	3702190	catawaba	11/3/2015	yes
Newton Conover City Schools	3703360	catawaba	11/7/2017	yes

Supplemental Appendix E: North Carolina Election Return Web-scraping

Our web-scraping process obtains official school board election returns from the North Carolina Board of Elections' "Elections Dashboard" landing page, presented earlier as Supplemental Appendix Figure B1. In essence, our strategy involves accessing the HTML source code of this central page via launching a botted Firefox browser controlled in R using functions from the *RSelenium* package (Harrison et al., 2025). We then extract the contents of each dropdown box dictating the date, county, and type of election to be examined in any given query using regular expression patterns, and cycle through combinations of these features to obtain election results. We restrict this process to all dates between the last possible election date in 2012 through the first possible election date in 2023 to align with the school board election date range used for the North Carolina QOR panel (years 2013-2022, inclusive).

Further word searches are employed to filter down raw results in several stages. First, we restrict the initial scraping to probable school board elections by selecting the "ALL LOCAL" option from the "Office" dropdown box, and filtering "Contest" box options to only those containing the words "EDUCATION" or "SCHOOL." We then save an intermediate dataset structured at the election contest-by-date-by-county level (e.g. "CUMBERLAND COUNTY BOARD OF EDUCATION AT-LARGE" on 11/8/2022), where the results table for each contest could be split across different counties if the contest crossed county lines. Each observation of this intermediate dataset contained page metadata (URL, webpage title, county name, election date, and contest name) as well as the actual HTML code of each webpage so that the raw data could be accessed later with or without an internet connection.

Secondly, we screen these stored observations to solely probable at-large elections by excluding election contests whose contest names (webpage titles) implied a sub-district or ward structure via the following key terms: DISTRICT, DIST, AREA, SEAT, WARD, TWP, TOWNSHIP. We also screen out non-school board elections through the REFERENDUM, TAX, and BOND keywords. These terms were selected based on a manual scan of the raw scraping results from the first step.

Third, we create a new candidate-by-school district-by-election date-by-county level at-large elections dataset using the HTML features of the stored election results' webpages, manipulated via the *rvest* package (Wickham, 2024). The election results in our data are stored in HTML elements identified as "#electionResults" on webpages and could be extracted during cleaning as a table in most cases.[10] We frequently needed to employ string replacement techniques based on regular expressions to remove non-numeric features of the results, such as percent signs, parentheses, and commas. We assign the results NCES school district identifiers (NCES IDs) from the list created during our background research, manually reviewing contest names and county names for each contest captured by the web-scraping and entering the appropriate NCES ID for all candidates via our R script. We remove 19 election contests (not candidates) during this process that had irregular names (e.g., "MCDOWELL COUNTY BOARD OF EDUCATION OLD FORT") suggesting a ward structure despite dodging our keyword screening. We also remove 7 other election contests that were for "UNEXPIRED TERM" seats co-occurring with another at-large election in the same district on the same date. We further need to manually change 46 contest names that were presented by the North Carolina Board of Elections using generic language (e.g., "BOARD OF EDUCATION"); we transform these contest names to become unique from other identical contest names by adding relevant metadata— usually just

[10] Some webpages' raw data necessitated deleting blank rows embedded in the HTML tables, and in these cases the county-level returns were instead contained in an element called "#lblBallotsCastPct"

the accurate county or city name as a minimal change— to the generic language (e.g., "PERQUIMANS BOARD OF EDUCATION"). Additionally, we take this opportunity to categorize contests into "County Only", "City", and "County with a City District" using the NCES IDs given that we already knew which IDs fell into which of these three categories.

Fourth, we feed the contests from this dataset back into a *for* loop that identified likely "Vote for 1" (or "VOTE FOR ONE") school board elections and flagged them with a binary 1 indicator based on the presence of a "VOTE FOR 1" or "VOTE FOR ONE" header in the webpage's election results table. For contests in "County Only" districts, we still retained any elections that were not "VOTE FOR 1" given that North Carolina reports separate county level registration, county level unique vote ("ballots cast") totals, and county level turnout rates alongside candidate level election returns. Therefore, our validation analyses where we compare QOR against the official county-wide at-large election returns can still use non-vote-for-one contests. However, receiving a "0" flag for vote-for-one status does exclude a district from being used in the roll-off analyses, which rely on summing votes cast across individual candidates within each contest. Many school board elections allow voters to cast votes for multiple candidates.

Fifth, we aggregate the candidate-by-school district-by-election date-by-county level data back up to the school district-by-election date-by-county level. Recall that the original district-by-election date-by county format previously mentioned only included election *metadata*, as the actual returns per election contest were reported in tables where each row corresponded to one candidate.[11] The final data structure contains: several county level variables (unique ballots cast,

[11] We mention earlier that North Carolina *does* report county level statistics alongside the candidates' vote totals. However, even this information is embedded in the HTML code near the candidate data, so functionally, researchers have to go into the same web pages with the candidate data for each election contest to extract the county level results anyway.

registered voters, and the turnout rate), the cross-candidate vote sum, a binary flag for "vote for 1" status (1= vote-for-one, 0 = not), a binary flag for being a multi-county contest (1 = multi-county, 0 = not), the election date, the election contest name, the county name, the web page URL, and the district type ("County Only", "City", "County with a City District"). As an added quality check, we attempt to collapse these district-by-election date-by-county level observations to the district-by-election date level. Manually reviewing the North Carolina Board of Elections website during our background research yielded that election returns are reported by county, with counties seemingly counting just the votes cast and voters registered within their own county. This is typical in the U.S., where county auditors often certify elections. While school board elections generally do not cross county boundaries in North Carolina (see Supplemental Appendix A), it may be necessary to collapse the web-scraping results across counties in rare edge cases. However, we find that no multi-county contests were left in our validation data after pruning not-at-large contests and (except for any county-wide contests) removing the not-vote-for-one contests. Thus, our data are already at the district-by-election date level after aggregating up from the candidates' election results.

Supplemental Appendix F: Identifying Districts Solely Using Zip Codes

We acknowledge that researchers may be tempted to use coarse district-voter matching strategies which ignore voters' exact addresses and instead rely upon larger geographic units. Zip codes are likely the most accessible (and potentially plausible) alternative to voter geocoding along exact addresses, given the findings in Supplemental Appendix A which suggest that counties are unsuitable for creating school district level turnout datasets. We again turn to NCES Geographic Relationship File (GRF) data, although this time we examine yearly sets of GRF files that record intersections between school districts and all U.S. zip code tabulation areas. The GRF (zip code) matching procedure replaces QOR for both North Carolina and Washington state, assigning voters a single school district based upon the first five digits of their zip code provided during voter registration. We determine these matches by, in most cases, identifying the school district which occupies the largest portion of each zip code's land area.[12] Thus, each zip code is assigned to just one school district, but each school district may pull voters from multiple different zip codes. This makes intuitive sense when considering that zip codes are smaller units than school districts generally speaking, and certainly in North Carolina and Washington. We craft another algorithm with the following major components:

1. Loading the national zip code-district GRF and the Census's national zip code shapefiles for each year.

[12] We use land area here, rather than the sum of land and water area, because the "total area" construction from Supplemental Appendix A was tailored to (1) the need to compare area values across two sources from the NCES and one from the Census, and (2) the fact that county boundaries cast a far wider net and tend to have smoother shapes less aligned to the land than something like zip codes. Many counties intentionally sweep over large bodies of water to indicate political and administrative ownership, whereas zip codes are intentionally created to facilitate mail delivery and receipt at businesses and residences. Further, we only use NCES-generated data for the GRF zip code matching strategy, and do not draw from a different (e.g., Census) source that may calculate areas differently (like with the counties). NCES does use Census data in its own data products but seems to process them in-house.

  a. Filtering these down to only the school districts in each state (North Carolina or Washington) using the list of identifiers within the NCES school district shapefiles temporally closest to the focal matching year.
2. Obtaining the "largest land area" matches that tie each zip code to just one school district.
3. Identifying any school districts that were not matched to any zip codes under this method and giving them "first priority" to obtain at least one zip code.
   a. Functionally this means that the initially unmatched "first priority" districts immediately receive their largest zip code by intersection land area,[13] taking these zip codes out of the matching pool such that larger districts are restricted from consuming them. Then, we obtain the "largest land area" matches for the remaining zip codes left in the pool.
4. Using the yearly datasets of zip code-by-district matches as a replacement for individual voters' address matches in our panel generation code.
   a. We merge the school district names and NCES identifiers into the yearly voter registration data along voters' registration zip codes, where each zip code only corresponds to one school district. Any individual school district, however, could have many zip code matches.
5. Applying the same 0% sub-district turnout threshold (see Supplemental Appendix D) to address likely wards and cancelled primaries.

[13] For North Carolina, this solves the issue and every school district receives at least one zip code match. In Washington, however, it is not possible to match all school districts to at least one zip code in every year. This is because Washington school districts are generally far smaller than in North Carolina, leading to much more sharing of zip codes across districts. We further adjust Washington's algorithm to specify breaking ties in the matching process and also to restrict matches among "first priority" districts to be the largest intersection for the whole *zip code* (rather than for the school district). These changes do not solve the problem, which is apparently intractable through a simple zip code match.

We find considerable justification for skepticism of simple zip code matching as a primary voter-district matching strategy, as detailed in the main manuscript. One additional drawback, though, lies in a fundamental problem where larger school districts exhaust the zip code pool to the detriment of accurately locating voters in small school districts. Our application of simple zip code matching to North Carolina shows that 9 school districts must arbitrarily receive the "first priority" rule in at least one year, or else they would obtain no zip code matches. In Washington state, more concerningly, 13 school districts go without *any* match in at least one year and we could not tweak the simple algorithm such that it matched all districts in all years. This tentatively suggests that zip code matching alone, especially when using GRF files to avoid applying formal Geographic Information Systems techniques, becomes an increasingly poor strategy when switching research foci to states with smaller school districts that have less district-county alignment.

To this point, even though identifying voters' school districts based on zip codes alone is intended as a time saving measure, users would likely need to develop a much more complicated algorithm with arbitrary assignment rules (e.g., which districts should be prioritized over others) if they wanted to merely represent all school districts each election cycle— let alone identify them accurately. The QOR method instead utilizes an intuitive assignment scheme (voters' actual addresses to districts' actual boundaries) to achieve the most precise match under reasonable assumptions. We also recognize the utility of falling back on other geographic units, namely zip codes, when exact address matching is not possible. However, the unique nature of school districts as numerous and broadly autonomous special districts, and their resulting poor alignment with other state and local administrative units, is cause for significant caution in solely using other matching methods besides the fine-grained approach offered by QOR.

Supplemental Appendix G: Regression Results

Appendix Table G1. *Relationships Between Turnout, Representativeness, and District Characteristics*

	Turnout			Representativeness
	(1)	(2)	(3)	(4)
Panel A. Relationships with Economic Characteristics				
% Students Free/Reduced Price Lunch Eligible	-0.078***	-0.051	-0.118***	-0.133***
	(0.022)	(0.061)	(0.020)	(0.032)
Constant	48.615***	49.202***	49.757***	-12.726***
	(1.192)	(4.395)	(1.059)	(2.382)
Panel B. Relationships with Racial Demographics				
% Students of Color	-0.173***	-0.172**	-0.195***	-0.273***
	(0.020)	(0.054)	(0.017)	(0.030)
Constant	51.047***	53.872***	50.726***	-8.744***
	(0.779)	(2.664)	(0.707)	(1.279)
Number of observations	2,034	559	1,475	559
North Carolina	X	X		X
Washington	X		X	

Note. Standard errors clustered at the district level in parentheses. Panels show OLS regression results of separate regressions of the panel measures regressed on the outcome indicated in the column (turnout in columns 1-3 and representativeness of voters of color in column 4).
*** p<0.001.

Supplemental Reference List

Bache S, & Wickham, H. (2025). magrittr: A Forward-Pipe Operator for R. R package version 2.0.4, https://magrittr.tidyverse.org.

Cambon, J., Hernangómez, D., Belanger, C., & Possenriede, D. (2021). tidygeocoder: An R package for geocoding. *Journal of Open Source Software, 6*(65), 3544. https://doi.org/10.21105/joss.03544

Harrison, John, Ju Yeong Kim, Jonathan Volkle, and Indranil Gayen. (2025). RSelenium: R Bindings for 'Selenium WebDriver.' https://docs.ropensci.org/RSelenium/

Izrailev, S. (2024). tictoc: Functions for Timing R Scripts, as Well as Implementations of "Stack" and "StackList" Structures. R package version 1.2.1, https://github.com/jabiru/tictoc.

Lindell, Noah B. (2017). ONE PERSON, NO VOTES. *Wisconsin Law Review 2017* (5): 885–954.

Pebesma, E., & Bivand, R. (2023). *Spatial Data Science: With applications in R*. Chapman and Hall/CRC. doi:10.1201/9780429459016, https://r-spatial.org/book/.

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal, 10*(1), 439–446. doi:10.32614/RJ-2018-009, https://doi.org/10.32614/RJ-2018-009.

Wickham, Hadley. (2024). Rvest: Easily Harvest (Scrape) Web Pages. https://rvest.tidyverse.org/.

Wickham, H., Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. doi:10.21105/joss.01686.