Who Bears the Cost of School Closures?

Evidence from Puerto Rico

Reza Moradi*

September 2026

Abstract

This paper studies the distributional consequences of school consolidation when families differ in access to effective alternatives. Using linked student–teacher records and enrollment before and after school closures in Puerto Rico, I develop and estimate a structural model in which families choose schools, while schools and teachers jointly contribute to achievement. The model distinguishes school contributions from teacher composition and uses families' substitution among schools to measure changes in access. Among closures before Hurricane María, high-poverty rural schools had higher estimated school effects than their surviving local alternatives, on average. Children displaced from these schools attended destinations with estimated school effects averaging 0.16 test-score standard deviations below those of their origins, with little average improvement over the surviving local benchmark. The demand estimates show greater sensitivity to distance among rural households, while poorer households rely more heavily on the public-school system. In the baseline model, rural-poor households account for 46.3 percent of predicted displacement but bear 76.3 percent of model-implied access losses. These findings show why enrollment alone is an incomplete guide to consolidation: evaluating closures requires accounting for both the effectiveness and accessibility of the alternatives available to displaced students.

JEL Classification: I21, I22, I24, I28, H75, R53

Keywords: school closures; school consolidation; school effects; teacher value-added; school choice; distance costs; rural schools; incidence; Puerto Rico

---

* Department of Economics, University of Toronto (email: reza.moradi@mail.utoronto.ca). I am deeply grateful to my supervisors, Gustavo Bobonis and Stephan Heblich, and to my committee member Robert McMillan. I also thank Victor Aguirregabiria, Eduardo Souza-Rodrigues, and participants in the University of Toronto Applied Micro and IO seminars. The administrative data were provided by the Puerto Rico Department of Education under a data use agreement; this research was supported by the University of Toronto Department of Economics, the School of Graduate Studies, and the Ontario Graduate Scholarship. I have no conflicts of interest to disclose, and all errors are my own.

1 Introduction

When enrollment falls, public school systems face pressure to consolidate schools and reduce the fixed costs of maintaining underused facilities (Engberg et al., 2012; OECD, 2018). However, the consequences of closing a school depend on the alternatives available to its students, since children may retain similar educational opportunities if an equally effective school is nearby. At the same time, a closure can be costly when suitable alternatives are far from their homes. Enrollment alone does not tell us which of these situations applies.

This paper asks how the distributional consequences of school closures depend on the effectiveness and accessibility of the schools that remain. I develop and estimate a structural model of school choice and educational production using the closures in Puerto Rico, which displaced approximately 60,000 children. In the model, families choose among schools in different locations, while the school attended and the teachers assigned jointly contribute to achievement. I use linked student-teacher records, enrollment before and after closures, school operations, geographic data, and population data from the Puerto Rico Community Survey.

Studies of school closures show that displaced students' outcomes depend on the schools they leave and enter (Brummet, 2014; Steinberg and MacDonald, 2019). In Puerto Rico, Bobonis et al. (2022) find that achievement effects differ according to the performance of the school a child left. I extend these comparisons by estimating school contributions net of teacher composition, drawing on the joint estimation of school and teacher effects in Mansfield (2015). I also distinguish the comparison of a closed school with its surviving local alternatives from the comparison of a child's actual destination with those alternatives. This allows me to examine the educational environments removed by consolidation separately from where displaced children subsequently enroll.

Epple et al. (2018) evaluate school-closing policies allowing parents to re-sort among the remaining schools. My analysis also draws on research relating school preferences to educational effectiveness (Walters, 2018; Abdulkadiroğlu et al., 2020). I use changes in enrollment after an attended school is removed to learn about substitution among schools, drawing on the information in second choices emphasized by Berry et al. (2004). These substitution patterns and the geographic distribution of schools allow me to estimate how access losses are distributed across households. I combine this analysis with the estimates of school and teacher contributions to examine access losses alongside changes in school-effect exposure, retaining their different units and populations rather than combining them into a single welfare measure.

Among closures before Hurricane María, I compare school effects, net of teacher composition, with a benchmark formed from surviving schools in the same municipality that serve overlapping grades, using preclosure enrollment weights. Closed high-poverty rural schools had higher estimated effects than their benchmark, on average, while closed urban schools had lower effects and the gap for lower-poverty rural schools was close to zero.[1] Children displaced from high-poverty rural origins entered schools with estimated school effects averaging 0.16 test-score standard deviations below those of their origins. Their destinations offered little average improvement over the local benchmark. The destination–benchmark difference is small and statistically indistinguishable from zero in all four origin-school groups, although this comparison does not establish which schools each family could have attended.

The demand estimates show greater sensitivity to distance in rural catchments, while survey and administrative data show that poorer households rely more heavily on the public-school system. I use the choice model to calculate the reduction in expected maximum utility when closing schools are removed, holding surviving-school attributes fixed. In the baseline, rural-poor households account for 46.3 percent of predicted displacement but bear 76.3 percent of model-implied access losses. I weight households by their population weights and predicted preclosure probabilities of attending a closing school; the calculation conditions on the baseline demand specification and the calibrated nesting parameter.[2] The distribution of losses depends jointly on the alternatives that remain, distance sensitivity, outside options, and exposure to closure.

These differences in school-effect exposure and access can coexist with small average achievement estimates. The reduced-form estimates in Section 6.1 are small and statistically indistinguishable from zero. The school-effect comparisons measure a change in one component of educational opportunity, rather than a treatment effect on children's own test scores, while access losses include costs that test scores do not measure.

Identification of educational production uses student and teacher movements, including closures and ordinary transfers, to separate persistent student differences from school and teacher contributions (Abowd et al., 1999). It requires assignment histories to be conditionally unrelated to remaining achievement shocks: forced departure does not make destinations random. I assess the dispersion of the estimated effects using a leave-out correction (Kline et al., 2020) and classroom-split comparisons, and examine a lagged-score ANCOVA specification as a robustness check on the treatment of prior achievement.

For school choice, enrollment before and after closure provides information about substitution when an

[1] Here, high poverty means that the origin school's poverty rate exceeds the unweighted median among closed schools in the full school-comparison dataset. This cutoff is calculated before restricting the sample to pre–Hurricane María closures and excluding schools with insufficient test-score coverage.

[2] The outside option includes private school, non-enrollment, and public enrollment outside the local municipality. Household types use block group rurality and administrative family poverty, while school effect comparisons use origin-school classifications. The measures therefore refer to different populations as well as different units.

attended school is removed. I integrate over possible residential locations, maintaining fixed residence and stable school attributes and utility rankings through displacement. Lagged school-average teacher value-added enters as observed quality; the separate school effect is absorbed by unobserved school demand, so a preference for this component is not separately identified. The distance coefficients describe responses under the information and transportation conditions families face.

The findings imply that enrollment alone is insufficient for assessing the distributional consequences of consolidation, since children leaving a small school may face alternatives that are less effective, more costly to reach, or both. Fiscal savings and the capacity of receiving schools also enter the policy decision. I report savings separately from the school-effect and access comparisons, and evaluating alternative closure programs would require a model of feasible reassignment that also accounts for these constraints.

Section 2 describes the closure policy and the procedures for pupil enrollment and teacher reassignment, and Section 3 introduces the linked records and reporting populations. Sections 4 and 5 present the integrated model and its identification. Section 6 compares closed schools with local alternatives and observed destinations, and reports the distribution of access losses. Section 7 presents the separate fiscal account, and Section 8 concludes.

2 Institutional Background

Puerto Rico consolidated schools during a long decline in enrollment. The program reduced the number of operating schools and moved pupils and teachers to the schools that remained, and it continued through the disruption caused by the 2017 hurricanes. For this study, I distinguish a school's closure from the decisions that followed it, since closing a building did not by itself determine where each child enrolled or which teachers taught at the destination.

2.1 Enrollment and Fiscal Pressure

The Puerto Rico Department of Education (PRDE) administered the public school system across the territory, which the 2016-17 Common Core of Data reports as a single operating district (NCES, nd). Consolidation therefore took place within a common administration. Public enrollment declined from 473,735 in fall 2010 to 365,181 in fall 2016, or about 23 percent before Hurricane María. By fall 2018, enrollment had fallen further to 307,282 (NCES, 2021).

The government's 2014 economic recovery agenda described declining enrollment as a reason for under-used facilities and high maintenance costs, proposing a reduction in the school network along with investments in receiving schools so that pupils would have nearby alternatives with better facilities and academic offerings. Moreover, receiving schools were expected to use part of the anticipated savings to expand specialized support and extracurricular activities (Government of Puerto Rico, 2014). These statements describe what the policy was intended to achieve, rather than showing whether the improvements took place or the savings were realized.

The program had already been operating before the largest rounds of closures, with the government describing consolidation over the earlier three years in its October 2016 fiscal-plan submission and listing academic performance, enrollment, utilization, and distance among its considerations. The plan also proposed moving personnel from administrative positions to school-based positions (Government of Puerto Rico, 2016). Therefore, the fiscal proposal involved the allocation of staff as well as the number of facilities.

2.2 Closure Policy

School size was not the only criterion in the formal procedures. The November 2015 guidelines included enrollment, safety, infrastructure, academic indicators, staffing, operating costs, and location, with possible exceptions where transportation was difficult, schools offered distinctive programs, or special education services would be costly or time-consuming to establish elsewhere. To evaluate a proposed closure, the Department required information on the receiving schools' physical capacity, grade-level enrollment, staffing, transportation, and necessary improvements (PRDE, 2015b).

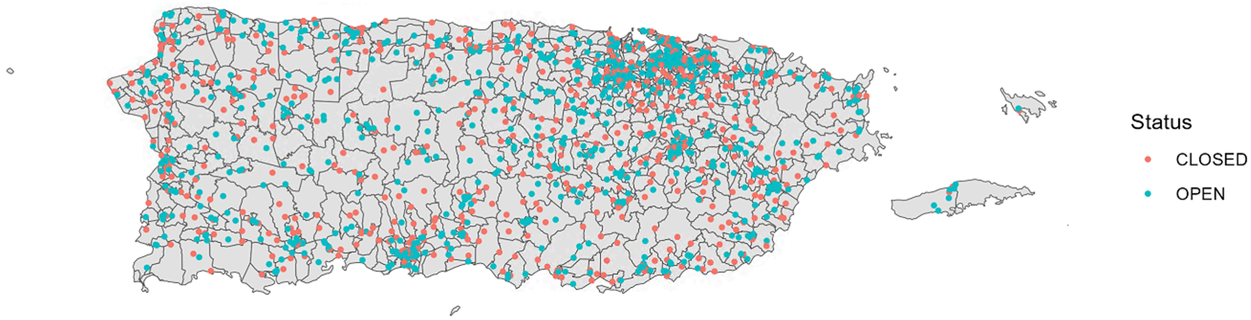
These procedures changed over the study period, as the March 2017 redesign circular replaced the November 2015 guidelines and explicitly considered transport routes and costs, rural or urban location, building utilization, and academic offerings. Families, teachers, schools, and municipalities were to be consulted before the committee recommendations and the Secretary's decision (PRDE, 2017b). The documents show what the formal process was expected to consider, although they do not establish how consistently each criterion affected individual closure decisions.

This distinction matters for the comparison in this paper because low enrollment can indicate unused capacity without measuring a school's contribution to achievement or how costly its replacements are to reach. Similarly, the academic indicators in the closure procedures describe observed performance, which does not separate school effects from pupil characteristics and teacher composition.

2.3 Pupil Enrollment and Teacher Reassignment

Families had a role in selecting schools under the Department's 2017 enrollment procedures, but their requests remained subject to administrative approval. The August instructions and accompanying manual

Figure 1. School Locations and Closure Status



Notes: Each marker shows a school location. Coral markers indicate schools recorded as closed, and teal markers indicate schools recorded as open. The map shows the geographic distribution of schools. Section 3 and Appendix A define the estimation samples and closure windows. Source: author's administrative school map presented at EARIE 2026.

allowed families to request a school and grade, although new enrollments needed the director's approval and directors could transfer a pupil to a different school while recording the reason. Eligibility for school transportation was decided separately (PRDE, 2017c). Under the 2017 procedures, a family's request, the enrollment decision, and access to transportation were related but separate parts of the process.

Teacher placement followed personnel rules. Under the January 2017 circular, regions first assessed whether receiving schools could accommodate teachers from closing schools, and those who could not be placed there became available for reassignment or transfer. Within position categories, priority depended on when a teacher obtained permanent status, followed by experience in the public system and academic qualifications, although teachers already working at receiving schools were not included in this seniority comparison with displaced teachers (PRDE, 2017a).

Since pupil enrollment and teacher placement followed separate procedures that could change different parts of a child's educational environment, I use the linked records to observe both where displaced children enrolled and which teachers taught them. In this comparison, I do not need to assume that teachers followed their former pupils or that reassignment improved estimated effectiveness. Figure 1 shows the schools recorded as closed and those that remained open. Throughout this paper, a school's closure year is its last operating year. Appendix A gives the annual counts and explains how this dating convention differs from others.

2.4 Hurricane Disruption and Closure Selection

Hurricanes Irma and María in September 2017 disrupted the consolidation program while it was already in progress. The education-sector recovery assessment describes damage to school facilities, interruptions to instruction, and reduced schedules after reopening because of electricity and water problems (Nelson et al., 2020). Temporary loss of access to a school is different from permanent consolidation, and outcomes in the later closure waves may reflect hurricane disruption and population movements as well as the removal of schools. I therefore use the earlier closure waves in the principal demand and closure comparisons, although the last destination cohort in the main demand window is observed in 2017–18, as explained in Section 5.2.

Bobonis et al. (2022) examine which observed characteristics predicted closure and find that small classes predicted closure in earlier waves, while low enrollment was especially predictive in 2017-2018 under their dating convention. Conditional on enrollment, they find no evidence of independent associations with achievement, demographic composition, or local political conditions. These linear probability estimates describe selection on observed characteristics; they do not establish a causal rule for the closure decisions.

3 Data

For this study, I use administrative records on pupils, teachers, and schools, together with population data and road distances. The production and school-choice components draw on these sources over different samples, so I describe the records and the construction of the geographic and poverty measures along with school characteristics and sample coverage. Appendix A gives further construction details.

3.1 Administrative data

The longitudinal pupil file records the school, grade, demographic characteristics, and poverty status of each public school child in each academic year. It also contains scores from the *Pruebas Puertorriqueñas de Aprovechamiento Académico*, administered in grades 3 through 8 in mathematics, Spanish and English. I normalize these scores within subject, grade, and year to have mean zero and unit variance, so the resulting scores and component effects are measured in standard deviations relative to a child's subject-grade-year peers. The score panel continues through 2019 and contains 211,879 scored records in the final year.

To identify the teacher assigned to each pupil, I use the student-teacher file, which links them by subject and, for real grades 3 through 8 and academic years ending 2010 through 2019, contains 4.77 million student-subject-year records for 501,291 pupils, 18,439 teachers and 1,294 schools. There is exactly one teacher in each student-year-subject cell, so no allocation rule for team teaching is needed. The school file records operation, grade offerings, enrollment, capacity, the urban/rural designation, coordinates, and closure status, including the final year of operation. I define demand markets using municipality-grade-year cells, where Puerto Rico's 78 municipios are county-equivalent units referred to as municipalities throughout the paper.

The pupil-key match rate between the first two files is 98.1 percent, while the school-code match rate is 87 percent, corresponding to 1,281 of 1,472 codes, with the unmatched codes belonging to high schools and schools outside the frame.

3.2 Population and geographic measures

I use residential population from the Puerto Rico Community Survey at the census block-group level to construct population weights, the density classification underlying rurality, and demand market size. Market size is the resident school-age population in a municipality-grade-year cell, not the number of children enrolled in its schools. Road distances are computed with the OSRM routing engine on the OpenStreetMap network. The school-to-school measure gives the distance between a closed school and a destination, whereas the block-group-to-school measure gives the distance from a possible residence to an alternative and enters household utility in Section 4.2. Neither measure gives the change in a particular child's journey from home.

One limitation of the pupil records is that residence is observed only at the ZIP level, and a Puerto Rican ZIP contains on the order of a hundred block groups, which prevents assigning an observed block-group residence to each child. To account for this limitation, the demand model integrates over block groups within municipality–grade–year markets using population weights, with geographic characteristics in utility referring to a possible residential block group. Using ZIP centroids would introduce measurement error that differs between rural and urban areas, and in the estimates it reverses the rural distance contrast. Appendix A presents the diagnostics for this measurement error.

3.3 Rurality and poverty

I define rurality at the level used in each comparison. For demand and incidence, I use a block-group classification calibrated to the Department's urban/rural school tag, which captures functional rurality on the island. The classification identifies school rurality with 78.8 percent accuracy, covers every agent block group, and assigns 0.441 of agent mass to rural areas. The Census urban share rule is not used because Puerto Rican settlement density exceeds the mainland urban threshold almost everywhere, leaving under a tenth of the island classified as rural. For the reduced-form and achievement comparisons, I use the urban/rural dummy of the child's school in the year before closure, which describes the origin school rather than the child's residence. Roughly 51 percent of displaced pupils attended a school classified as rural before closure.

The poverty measures also depend on the comparison. Demand and incidence use the Department's administrative family-poverty dummy, which is assigned from family income and covers about 81 percent of public-school pupils. Each block group has a poor and a non-poor sub-agent facing the same distances, which holds location fixed in the model, although the residential block groups of students in the destination data remain unobserved. For production and local benchmark comparisons, poverty is measured at the school level, with schools divided at the median poverty rate for group comparisons and child weights used for student-level comparisons. The baseline additive regressions instead use the continuous poverty rate, standardized where indicated, so these measures describe the school a child attended rather than family poverty.

A third measure, Census *income* poverty, which covers about 56 percent of children and is nested inside program poverty at the person level, shifts enrollment in the public system. The Census area poverty rate is not a substitute for the administrative family dummy: their correlation is $+0.02$. Since the dummy is observed only in tested grades and children who leave the system stop testing, I impute the modal poverty status across observed enrollment years for displaced children whose flag is missing. Appendix A reports checks on persistence and agreement for this imputation.

3.4 Sample construction

I date each closure by the school's *last operating year*. A child enrolled there in that year is *treated*, and I classify the child as *displaced* if observed at another public school in the following year, or as having *exited* if absent from the public panel. Exit includes movement to a private school, migration off the island, dropout, and linkage failure, so it measures departure from the public panel and cannot be interpreted as dropout alone. In these records, a closure affects the whole school.

The administrative school file identifies 610 closed schools in the grades 3-8 frame, with a last operating year between 2013 and 2018 recorded for 596 of these schools and no assigned annual wave for the remaining fourteen. Approximately two-thirds of the dated closures occurred in the final two waves. The broader program displaced roughly 60,000 students. Appendix A reports the annual counts.

I use the full scored panel for production. The demand sample covers 2013–2017 origin market years, with destinations observed in 2014–2018. In the closure-selection comparisons, schools must have a last operating year no later than 2017. The primary reduced-form exit estimates use closure cohorts through 2015, and the robustness checks extend the cohorts through 2016.

Table 2 reports the coverage of these samples. The scored production sample has 1,947,300 student-subject-year records for 281,943 students, 11,888 teachers (16,806 teacher-subjects) and 1,235 schools. After restricting to the connected teacher-mover graph and applying leave-out pruning, the sample contains 1,912,673 observations for 279,644 students, 16,417 teacher-subjects and 1,177 schools. The variance components use a separate, wider 2011-2019 frame with 3,452,029 observations, and since these samples differ, I report each statistic with the sample that produced it. The lagged-score ANCOVA robustness specification uses grades 4–8 with the required prior-score data, since grade 3 has no preceding tested grade. Appendix B.3 describes its controls and distinguishes this sample from the pupil-fixed-effect samples above.

The demand frame includes more grades but less information on individual children, covering real grades 1 through 8 in 3,120 municipality-grade-year markets ($78 \times 8 \times 5$), with 1,281 schools, 31,767 school-market records, 285,800 agent records and 26,675 displaced micro records. These displaced records cover grades 1 through 7 because terminal-grade transitions leave the modeled network. Standard errors are clustered by municipality, with 78 clusters. Achievement is observed only in real grades 3 through 8, so I report access losses and changes in school-effect exposure separately because they have different units and cover different populations: grades 1–8 for access and grades 3–8 for school-effect exposure.

The closure comparisons have narrower samples, including 317 pre-María closures in 76 municipalities, each compared with surviving schools in the same municipality that serve overlapping tested grades, and 34,118 children from these origins in the student-level analysis. The incidence calculation covers 992 closure markets, 122,350 agent records, and 334 closed schools.

3.5 Descriptive statistics and sample coverage

Enrollment in the grades 3-8 administrative frame decreased from 253,087 in 2013 to 176,607 in 2019, a decline of 30.2 percent. This population and period differ from those used to describe the population decline in Section 2.

Table 1 compares observed school characteristics by closure status. The median enrollment was 139 in schools that closed and 324 in schools that never closed, with closed schools being smaller in both rural and urban areas, and mean capacity utilization was 0.689 and 0.758, respectively. These comparisons describe the schools affected, although they are unconditional and do not identify the separate role of each characteristic in closure decisions. Table 2 then distinguishes the pupils, teachers, schools, and repeated observations used in the production and demand samples. I use the model to connect pupils' school choices with the school and teacher inputs in educational production.

4 Model

I use a model in which families choose schools and schools and teachers jointly contribute to children's achievement, with the production equation separating pupil, school, and teacher contributions and the choice equation allowing families to respond to distance, teacher quality, other observed and unobserved school

Table 1. School Characteristics

	Closed	Never closed
Median enrollment in the last operating year	139	324
rural schools	125	284
urban schools	170	342
Utilization of building capacity, mean	0.689	0.758

Notes: I use administrative enrollment and capacity records to describe closed and surviving schools. The unit of observation is a school. Enrollment is measured in the closed school's last operating year and in the same year for surviving schools. Utilization is enrollment relative to the building's rated capacity. The two groups differ in utilization by seven percentage points. These comparisons are unconditional and concern observed school characteristics.

attributes, and idiosyncratic tastes. When a school closes, families choose among the remaining alternatives, and the resulting changes in attendance affect the educational inputs children receive as well as their access to public education.

The link between the two equations is lagged school-average teacher value-added, which I obtain by estimating production first and then use as the quality measure in school choice. The school effect also enters the unobserved component of school demand, but its contribution cannot be separated from other unobserved determinants of demand, so the choice equation does not identify a preference for this component. The model takes observed teacher assignments as given rather than explaining schools' staffing decisions or teachers' choices of employer. I hold the attributes of surviving schools fixed in the access counterfactual, while the analysis of educational exposure uses observed student and teacher transitions.

Table 2. Estimation Samples and Coverage

Sample	Window	Unit of observation	N
Panel A. School and teacher effects: the grades-3–8 scored frame			
Linked teacher–pupil frame	2010–2019	pupil–subject–year	4,771,061
Scored value-added sample	2010–2019	pupil–subject–year	1,947,300
Connected-sample estimation	2010–2019	pupil–subject–year	1,912,673
Match-variance sample	closure movers	teacher	1,626
Panel B. School and teacher effects: the variance-component frame			
Variance-component estimation	2011–2019	pupil–subject–year	3,452,029
Panel C. What the closure rule removed			
Closure-selection comparison	2013–2017	closed school	317
Realized change in the school effect	pre-María	displaced child's move	34,772
Displaced teachers, exit	pre-María	displaced teacher	879
Displaced teachers, destination	pre-María	destination assignment	541
Panel D. Demand and incidence			
Demand estimation	2013–2017	school–market row	31,767
Welfare and incidence	pre-María	agent–market row	122,350
Panel E. Fiscal accounting			
Staff roster window	2016–2018	exposed staff-year	10,819
Extrapolation universe	2013–2018	closed school	610

Notes: Academic years are labeled by their ending year. *Panel A.* The linked frame covers real grades 3 to 8 and includes 501,291 pupils, 18,439 teachers, and 1,294 schools. The scored sample contains 281,943 pupils, 11,888 teachers (16,806 teacher–subjects), and 1,235 schools. Of these schools, 1,179 (95 percent) belong to a single connected teacher-mover graph. The connected-sample estimation uses 279,644 pupils, 16,417 teacher–subjects, and 1,177 schools. It provides the individual school and teacher effects used in Section 6.3. Section 5.2 describes the saved teacher-quality input used in demand. The match-variance sample also includes 469 teachers with at least two scored years on each side of the move. *Panel B.* I obtain the variance components in Table 4 from a separate estimation on a wider 2011–2019 frame, containing 395,030 pupil effects and 1,274 schools. The two estimations are not interchangeable. I report each statistic with the sample from the run that produced it. *Panels C–E.* I cluster the closure-selection comparisons and demand estimation by municipality, with 76 and 78 clusters, respectively. The child-level regressions cluster by origin school: 325 schools for the cell means and 317 for the decomposition. The welfare bootstrap resamples 76 municipalities. The demand frame contains 3,120 municipality–grade–year markets, 1,281 schools, 26,675 displaced micro rows, and 285,800 agent rows. The welfare frame contains 992 closure markets and 334 closed schools. Closures displaced 2,145 teachers of tested subjects and about 60,000 pupils. The sample window depends on the analysis. I use the full scored window for production, but restrict every causal reduced-form comparison and closure-selection comparison to waves that closed before Hurricane María on 20 September 2017. The 2018 wave follows the hurricane. More specific analyses use smaller frames: roughly 688 schools appear in the system-wide 2010–2019 counts, 610 in the grades-3–8 panel, 334 in the pre-María access frame, and 317 in the frame with both access and achievement channels.

4.1 Educational Production

Let i index students, $s \in \{\text{mathematics}, \text{English}, \text{Spanish}\}$ subjects, t academic years, k teachers, and j schools. Write $k(i,s,t)$ for student i's teacher in subject s and year t, and $j(i,t)$ for the school attended that year. Achievement is generated by

$$Y_{ist} = a_i + \mu_{k(i,s,t),s} + \phi_{j(i,t)} + e_{ist}, \tag{1}$$

where Y_{ist} is a test score standardized within subject, grade, and year; a_i is a pupil effect; μ_{ks} is teacher k's value-added in subject s; ϕ_j is the school effect; and e_{ist} collects the remaining production disturbances, including classroom, school-year, and pupil shocks and any residual teacher–school match component. School and teacher effects are measured in student test-score standard deviations.

The school effect captures persistent differences in school organization and the peer environment, while pupil effects control for persistent individual differences without separating the contribution of peers from other school inputs. Although including group averages of observed student characteristics could help make this distinction (Altonji and Mansfield, 2018), I do not estimate that extension here. Changes in school-effect exposure therefore describe a broad component of educational production rather than the causal effect of closure on a student's own score.

I allow teacher effects to differ across subjects while pooling student and school effects across subjects, with any separate subject intercepts absorbed by the teacher-subject fixed effects. Standardizing scores makes the outcome units comparable, but it does not impose zero conditional subject intercepts. School effects are also held constant over time, so school-year deviations enter e_{ist}; Appendix B examines the consequences of this restriction.

To measure the teacher input at each school, let $n_{jt,ks}$ be the number of scored student-subject observations taught by teacher k in subject s at school j in year t, and let $n_{jt} = \sum_{k,s} n_{jt,ks}$. School-year teacher quality is $q_{jt} = \sum_{k,s}(n_{jt,ks}/n_{jt})\mu_{ks}$. Pooling years gives $\bar{q}_j = \sum_t (n_{jt}/n_j) q_{jt}$, where $n_j = \sum_t n_{jt}$. I define the composite school effect as $\Psi_j = \phi_j + \bar{q}_j$ and its school-year counterpart as $\Psi_{jt} = \phi_j + q_{jt}$. These averages summarize teacher inputs at the school level, whereas the input received by an individual student in (1) depends on the teacher assigned to that student. I use the lagged school-year measure $q_{j,t-1}$ in school choice, keeping in mind that it concerns tested skills and that effects on behavior and later attainment need not coincide with effects on test scores (Jackson, 2018; Beuermann et al., 2023).

4.2 School Choice

I model families' choices as a response to teacher quality $q_{j,t-1}$, distance, and other school attributes. Persistent school effects ϕ_j can also influence those choices through the unobserved component of school demand, but since their coefficient cannot be identified separately from other unobserved determinants of demand, the choice estimates describe substitution and access without identifying willingness to travel for the school-effect component. Destinations outside the local public-school menu are combined in one alternative. In the demand analysis, $q_{j,t-1}$ denotes the estimated counterpart of the production measure defined above, and I suppress hats except when distinguishing measured quality from its latent counterpart.

I define a market m by municipality of residence, grade, and year, with the inside menu $\mathscr{J}_m$ containing public schools that serve that grade in the municipality. Alternative 0 includes private school, non-enrollment, and public enrollment across a municipal boundary. For a family i in market m, let t denote the market's academic year and allow school–market mean utility δ_{mj} to vary by school and market, including grade. This component enters utility identically for households within the market. Family i obtains

$$\begin{aligned} u_{imj} \;=\; \delta_{mj} \;+\; \big(\lambda_0 + \lambda_{\text{rural}} R_i + \lambda_{\text{pov}} P_i\big)\, d_{ij}^{0.8} \;+\; \big(\beta_{\text{rural}} R_i + \beta_{\text{pov}} P_i\big)\, q_{j,t-1} \\ +\; \pi_{\text{pov}} I_i \;+\; \varepsilon_{imj}, \qquad u_{im0} = \varepsilon_{im0}, \qquad \delta_{mj} = \beta_q\, q_{j,t-1} + x_{jt}'\gamma + \alpha_{ct} + \xi_{mj}, \end{aligned} \tag{2}$$

where the outside option's deterministic utility is normalized to zero. Road distance in kilometers is denoted by d_{ij}; $q_{j,t-1}$ is estimated lagged teacher quality, rather than an average of realized student scores; x_{jt} contains lagged capacity and an urban indicator; and α_{ct} are region-year fixed effects, with c denoting the school's region. The unobserved school–market demand component ξ_{mj} absorbs the contribution of ϕ_j to school utility. Section 3 defines the three household indicators: block group rurality R_i, administrative family poverty P_i, and census income poverty I_i. The last is nested inside P_i and shifts the utility of the inside option. I split the integration agents in each block group by administrative family poverty, so that poor and non-poor agents in that block group face identical distances. Throughout the paper, I write $\lambda_{\tau(i)} = \lambda_0 + \lambda_{\text{rural}} R_i + \lambda_{\text{pov}} P_i$ for compactness. The distance exponent is calibrated at $p = 0.8$, as described in Section 5.2. Kilometer equivalents therefore require solving the nonlinear distance tradeoff.

The errors follow a nested-logit distribution with $0 \le \rho < 1$, in which public schools in the municipality form one nest and the outside alternative forms a singleton nest. I use the variance-components structure of Cardell (1997) to obtain the corresponding choice probabilities and logsum (Train, 2009). Let $D_i = \sum_{j \in \mathscr{J}_m} \exp(v_{ij}/(1-\rho))$, where v_{ij} is the deterministic part of (2), with the market index suppressed. Using the same convention, the choice probabilities are

$$p_{ij} = \frac{\exp\big(v_{ij}/(1-\rho)\big)}{D_i} \cdot \frac{D_i^{1-\rho}}{1 + D_i^{1-\rho}}, \qquad p_{i0} = \frac{1}{1 + D_i^{1-\rho}}. \tag{3}$$

4.3 Educational Exposure after Closure

Let o index a closed school, and let $cy(o)$ be its closure year. The administrative file records the last operating year, so displaced pupils first appear at other schools in $cy+1$. The *surviving local menu* of o is

$$M(o) \;=\; \big\{\, j : m(j) = m(o),\;\; j \text{ operating in } cy(o)+1,\;\; \text{tested-grade spans of } j \text{ and } o \text{ overlap},\;\; j \neq o \,\big\}, \tag{4}$$

where $m(\cdot)$ denotes municipality. I use this set as a local benchmark because its schools have overlapping grade spans, although that overlap does not imply that each school serves every displaced child's grade. The *menu gap* compares the closed school's effect with a weighted average of the effects at these surviving schools:

$$\begin{aligned} \operatorname{gap}_{\phi}(o) &= \hat{\phi}_o - \sum_{j \in M(o)} w_{oj} \hat{\phi}_j, \\ \operatorname{gap}_{q}(o) &= \hat{q}_{o,cy(o)} - \sum_{j \in M(o)} w_{oj} \hat{q}_{j,cy(o)}. \end{aligned} \tag{5}$$

The primary weights are $w_{oj}^{\text{enr}} = E_{j,cy(o)} / \sum_{\ell \in M(o)} E_{\ell,cy(o)}$, where E is preclosure enrollment. The robustness weights are proportional to $\exp\{-d_{oj}/(5\,\text{km})\}$ and are also normalized to sum to one within $M(o)$. Both school-average teacher-quality terms use the preclosure year $cy(o)$. A positive gap, $\operatorname{gap}_{\phi}(o) > 0$, means that the closed school had a higher estimated school effect than its surviving local benchmark.

I separate the realized change in school-effect exposure into the contribution of the local menu and the contribution of the destination chosen. Let $d(i)$ and $o(i)$ denote student i's destination and origin, and write $\bar{\phi}_{M(o)} = \sum_{j \in M(o)} w_{oj} \hat{\phi}_j$. Then:

$$\Delta\hat{\phi}_i = \hat{\phi}_{d(i)} - \hat{\phi}_{o(i)} = -\operatorname{gap}_{\phi}(o(i)) + \left[\hat{\phi}_{d(i)} - \bar{\phi}_{M(o(i))}\right]. \tag{6}$$

The first term gives the change in exposure if the origin were replaced by its local benchmark, while the second shows how the actual destination differs from that benchmark. These terms describe the estimated educational environments a student leaves and enters; neither identifies closure's treatment effect on the student's score or establishes which destinations were feasible for that student.

4.4 Access Losses

I use the choice model to measure how removing schools changes expected maximum utility. Let v_{ij} be the deterministic utility in (2), with the market index suppressed. Given one inside nest of same-municipality public schools, a singleton outside option, and nesting parameter ρ, expected maximum utility over the open set $\mathscr{O}$ is

$$W_i(\mathscr{O}) = \log\left[1 + \exp\left\{(1-\rho) IV_i(\mathscr{O})\right\}\right], \qquad IV_i(\mathscr{O}) = \log \sum_{j \in \mathscr{O}} \exp\left(\frac{v_{ij}}{1-\rho}\right), \tag{7}$$

and the nonnegative access loss of a closure set C is $L_i(C) = W_i(\mathscr{O}^{\text{pre}}) - W_i(\mathscr{O}^{\text{pre}} \setminus C)$, holding the mean utilities of the survivors, all distances, all agent weights and all parameters fixed.

The change $\Delta W_i(C) = -L_i(C)$ follows the discrete-choice logsum measure of Small and Rosen (1981), which I evaluate using families' effective demand responses under the information and transportation conditions they face. I interpret it as a change in access. It also measures a welfare change when the utility governing choices represents the welfare being evaluated, although imperfect information need not satisfy this condition (Kapor et al., 2020). The baseline calculation retains all six demand shifters, and I report the resulting access losses separately from school-effect exposure because the measures have different units and populations. Combining them would require a utility weight for ϕ_j that the model does not identify and could count the same attributes twice if those attributes already enter the unobserved component of school demand.

To aggregate these losses, let w_i be an integration weight normalized to sum to one within market m, and let M_m be resident school-age population. In the baseline, I use displaced-equivalent weights $h_i = w_i M_m s_{i,C}^{\text{pre}}$, where $s_{i,C}^{\text{pre}} = \sum_{j \in C} p_{ij}^{\text{pre}}$ is the preclosure probability of attending a closing school. Group g's population share is $\sum_{i \in g} h_i / \sum_i h_i$, and its share of access losses is $\sum_{i \in g} h_i L_i(C) / \sum_i h_i L_i(C)$. These are ex ante shares weighted by attendance probabilities rather than counts of observed displaced students. I also consider an alternative population in which all residents of closure markets receive weights $w_i M_m$, with the calculation and inference described in Appendix D.

5 Identification and Estimation

I use changes in student and teacher assignments to separate pupil, school, and teacher contributions to achievement, and school choices before and after closure to study how families respond when their school is removed. Student and teacher movements link schools in the production equation, while displaced students' destinations show how families substitute among the remaining alternatives. I estimate the two components sequentially, using lagged school-average teacher value-added from the production equation as the quality measure in school choice.

5.1 Educational Production

The decomposition in (1) uses changes in teachers and schools within students' histories: observing a student with different teachers helps separate teacher contributions from persistent student heterogeneity, while student moves provide comparisons across schools. Teachers who teach the same subject at more than one school also link teacher-subject and school effects across institutions. Together, these assignment changes provide the variation needed to separate a_i, μ_{ks}, and ϕ_j, extending the mobility-based fixed-effect framework of Abowd et al. (1999) and Card et al. (2013) to educational production. I estimate the effects by least squares using the linked assignment panel.

After the fixed-effect normalizations, identification requires sufficient rank of the joint student, teacher-subject, and school design. For the closure comparisons, I use the individual effects from the connected, leave-out-pruned grades-3–8 sample in Table 2 and Appendix B, where both ordinary transfers and closure-induced moves provide the mobility links.[3]

I impose strict exogeneity on the assignment histories:

Assumption 1 (Strict exogeneity of assignment)**.** *Let $\mathscr{A}$ collect the full student, teacher-subject, and school assignment histories and the observation pattern in the estimation sample, and let a, μ, and ϕ collect the student, teacher-subject, and school effects, respectively. The remaining production disturbance satisfies $\mathbb{E}[e_{ist} \mid \mathscr{A}, a, \mu, \phi] = 0$ for every observed student–subject–year record.*

Assumption 1 permits students and teachers to sort across schools on the persistent characteristics captured by the fixed effects, but requires the combined remaining disturbance, including any residual teacher–school match component, to satisfy the mean-independence restriction. Reassignment in response to earlier achievement shocks can violate this assumption (Rothstein, 2010). Closures provide useful variation since students and teachers must leave their origin school, reducing selection through voluntary departure as studied by Kramarz et al. (2015), although the receiving school and classroom assignments are still subject to the restriction. The same restriction must hold for ordinary transfers.

The additive equation separates the contribution attached to the teacher from the contribution attached to the school, assuming that μ_{ks} is portable across schools and that ϕ_j is stable over the estimation period. This restriction is consistent with evidence that closure-induced relocation does not reduce teacher value-added on average (Hill and Jones, 2021), and I examine how well it fits these data through the teacher–school match exercise in Section 6.2.3. Appendix B compares specifications for student heterogeneity and the stability of school effects, while the forecast exercises in Section 6.2.1 ask whether the estimated components predict achievement changes among displaced students.

As a robustness check on the treatment of prior achievement, I also estimate a lagged-score ANCOVA that replaces pupil fixed effects with flexible controls for previous scores and individual demographics, retaining teacher–subject and school effects (Chetty et al., 2014). This specification permits assignment to depend on observed prior achievement, but interpreting its effects still requires a conditionally mean-zero

[3] Connectivity concerns the teacher-subject and school assignments in the estimating equation, together with student mobility. A teacher who changes subject when moving does not link the same teacher-subject effect across schools. Appendix B describes the graph restrictions and specification diagnostics.

disturbance after these controls. Appendix B.3 sets out the specification and the scope of the comparison; agreement across specifications would not by itself establish Assumption 1.

For the pupil-fixed-effect specification, estimation error contributes to the dispersion of estimated fixed effects alongside differences in educational contributions, so I correct the variance components by adapting the leave-out estimator of Kline et al. (2020), using calibrated variance components to allow for classroom-year and school-year shocks. I also estimate covariances between effects obtained from separate classroom samples. If estimation errors are uncorrelated across those samples, the covariance recovers the variance of the common component. Appendix B describes the correction and sample definitions and reports sensitivity to the common-shock components.

5.2 School Choice

School closures provide information beyond aggregate enrollment shares, since they reveal where students enroll when the school they initially chose is no longer available. I combine shares before and after closure with displaced students' destinations, the income composition of public enrollment, and the extensive-margin response to closure, using the alternative chosen when an initial choice becomes unavailable to help distinguish demand responses across households (Berry et al., 2004). The coefficients describe responses to proximity and measured teacher quality under the information and transportation conditions faced by each group, which may combine preferences, information about quality, and the costs of reaching school. The data do not separately identify these mechanisms (Bayer et al., 2007; Hastings and Weinstein, 2008).

Let θ collect the six coefficients estimated from the demand moments:

$$\theta = (\lambda_0, \lambda_{\text{rural}}, \lambda_{\text{pov}}, \beta_{\text{rural}}, \beta_{\text{pov}}, \pi_{\text{pov}})'.$$

For each candidate θ, I recover the school–market mean utilities $\delta_{mj}(\theta)$ that reproduce observed enrollment shares (Berry, 1994; Berry et al., 1995). Let $\delta_m(\theta)$ be the vector of these mean utilities in market m; recovering mean utilities allows unobserved determinants of school demand common to families to be correlated with teacher quality, and the destination moments inform the distance and quality interactions conditional on the recovered mean utilities. Mean utilities that are common to households cannot absorb arbitrary school attributes valued differently across groups, so the choice specification must represent remaining systematic group differences through its observed interactions and conditional error distribution.

Because residential census block groups are unobserved, I integrate choice probabilities over household characteristics $z_i = (b_i, P_i, I_i)$, where b_i denotes the residential block group and P_i and I_i are the administrative family-poverty and Census income-poverty indicators defined in Section 4.2. Residential rurality R_i is determined by b_i. Let F_m denote the joint distribution of these characteristics in the resident school-age population of market m. The identification argument maintains the population weights in Appendix A and the conditional nested-logit specification in Section 4.2.

For an inside choice set $\mathscr{S} \subseteq \mathscr{J}_m$, let $p_{mj}(z;\mathscr{S},\theta)$ denote the probability that a household with characteristics z chooses alternative $j \in \mathscr{S} \cup \{0\}$. This is the probability p_{ij} in (3), written as a function of household characteristics, available schools, and coefficients. I evaluate it at $\delta_m(\theta)$, recovered from the original enrollment shares, keeping the distance exponent and nesting parameter at their calibrated values. If an origin is removed, the mean utilities of the remaining schools stay at these values.

Let O_i denote the school initially attended. Since a household chooses origin o with probability $p_{mo}(z;\mathscr{J}_m,\theta)$, the composition of students attending a school need not match that of the market population. The model-implied joint distribution of residence and poverty characteristics among students attending o is proportional to $p_{mo}(z;\mathscr{J}_m,\theta)\,dF_m(z)$, accounting for both the population composition and selection into the origin school.

Let J_i^{obs} denote the observed destination, with destinations outside the modeled public-school menu coded as 0, and let C_o be an indicator equal to one if origin o closes. For movers observed at another school in the modeled public-school menu, $o(i)$ and $d(i)$ in Section 4.3 are the observed realizations of O_i and J_i^{obs}, respectively. Let $\mathscr{H}_m$ collect the school attributes, school–market mean utilities, and population composition entering market m's choice model. Write u_{ij} for the origin-year utility u_{imj}, suppressing the market index, and let ε_i collect the individual utility disturbances across alternatives. The notation $\mathscr{L}(\cdot \mid \cdot)$ denotes a conditional distribution.

Assumption 2 (Closure-induced substitution)**.** *For each closing origin with positive enrollment, selection into closure satisfies*

$$\mathscr{L}(z_i,\varepsilon_i \mid O_i = o, C_o = 1, \mathscr{H}_m) = \mathscr{L}(z_i,\varepsilon_i \mid O_i = o, \mathscr{H}_m). \tag{8}$$

Conditional on $O_i = o$ and $C_o = 1$, the destination satisfies

$$J_i^{\text{obs}} = \underset{j\in(\mathscr{J}_m\setminus\{o\})\cup\{0\}}{\arg\max}\; u_{ij}, \tag{9}$$

with the same utility realization used to represent initial attendance.

The first restriction allows closures to depend on school attributes, mean utilities, and enrollment implied by the model, while excluding additional selection on household composition or latent school rankings, including anticipatory departures that the model does not capture. Enrollment-related selection is consistent with the closure predictors reported by Bobonis et al. (2022, p. 9), although their school-level evidence does not establish the restriction on families' substitution patterns. Under the second restriction, the observed destination is the next-best alternative after the origin is removed, which requires changes in rankings or school availability to leave the modeled destination choice unchanged.[4] The comparison deletes one origin from its origin-year menu. Other contemporaneous closures must therefore leave the modeled destination choice unchanged.[5]

Under Assumption 2, the joint probability of choosing o initially and $j \neq o$ after its removal, conditional on z, is

$$T_{moj}(z;\theta) = p_{mj}(z;\mathscr{J}_m \setminus \{o\},\theta) - p_{mj}(z;\mathscr{J}_m,\theta). \tag{10}$$

Deleting o changes the choice only of families who initially attended it, so the increase in j's probability is the mass that moves from o to j (Conlon and Gortmaker, 2025). Integrating this joint probability over F_m accounts for selection into the origin. For the observed and predicted destination moments, I then apply the same conditioning on inside re-enrollment, poverty, rural exposure, and data coverage, with the normalization given in Appendix C.1.

The distance moments contain a pooled mean relocation distance and sixteen rank shares across four poverty-by-rural-exposure cells. Utility depends on distance from each possible residence, whereas the observed ranks and mean distance are measured from the closed school, so changes in λ_0, λ_{rural}, and λ_{pov} affect predicted relocation patterns through the relative attractiveness of schools at different residential distances. For quality, I use fifteen contrasts: three independent comparisons across cells for four retained quality-rank bins and a mean quality gap, which inform β_{rural} and β_{pov} through quality-related substitution jointly with distance. Although the contrasts remove an additive discrepancy common to the compared quality moments, they do not eliminate the dependence of nonlinear choice probabilities on school mean utilities. The appendix states this weaker quality-moment restriction separately from Assumption 2.

Administrative poverty is observed at the student level, but residential rurality is not, so the rural and urban cells weight each origin by its catchment's rural exposure and the complement of that exposure.

[4] The May 2015 notice to parents at Los Caños permitted another grade-appropriate school, subject to capacity, and provided for automatic transfer to a recommended receiver when parents did not respond (PRDE, 2015a). These rules support a role for family choice but do not establish unrestricted choice. Capacity, administrative placement, transportation, and receiving-school changes must be adequately represented for (9) to hold.

[5] Origin years are 2013–2017 and destination years are 2014–2018, labeled by the end of the academic year. The last destination cohort is observed in 2017–18: its closure decision precedes María, but its destination observation need not. Model probabilities use origin-year attributes and mean utilities. The transition restriction covers this entire interval, rather than only the date of the closure decision.

The same origin can contribute to both cells. The rural interactions therefore rely on the specified residential distribution and variation across catchments rather than comparisons of observed rural and urban students within an origin. The within-municipality comparisons and residence-measurement exercises in Appendix C.8 examine this source of variation, without establishing the absence of other systematic differences across catchments.

I estimate the six coefficients jointly by generalized method of moments, using the thirty-two destination restrictions and the composition and exit targets. Census income poverty informs π_{pov}, which shifts all inside utilities relative to the outside option; this poverty measure differs from administrative family poverty used in the destination moments. The estimation conditions on the distance exponent $p = 0.8$ and nesting parameter $\rho = 0.953524$, with travel-time evidence motivating the exponent and the extensive margin informing the calibration of ρ. The exit target remains in the criterion and restricts the free coefficients conditional on this calibration; all moment blocks operate jointly. A sufficient condition for local identification is that the six coefficients produce linearly independent changes in fitted moments after share inversion, as stated in the appendix. Standard errors condition on the calibrated inputs.

The common quality coefficient β_q requires a separate argument from the six coefficients in θ, since its contribution is absorbed by δ_{mj} and identifying it requires a restriction on the relation between teacher quality and the unobserved component of school demand. Let b_q denote the population coefficient from projecting mean utility on lagged quality, the school controls, and region-year effects in (2). With correctly measured quality and correctly recovered mean utility, this projection satisfies

$$b_q = \beta_q + \frac{\mathrm{Cov}(q^{\perp}, \xi^{\perp})}{\mathrm{Var}(q^{\perp})}, \qquad \mathrm{Cov}(q^{\perp}, \xi^{\perp}) \geq 0 \implies \beta_q \leq b_q. \tag{11}$$

Here q denotes lagged quality $q_{j,t-1}$ and ξ denotes ξ_{mj}, with school and market indices suppressed. Superscript $\perp$ denotes residualization on the projection's controls, with its sample and weights held fixed. The sign restriction is motivated by the possibility that schools with higher quality also offer omitted amenities valued by families (Abdulkadiroğlu et al., 2020, p. 1527), but its applicability to the teacher component used here remains an assumption. Measurement error in teacher quality requires additional restrictions, including on its effect on recovered mean utilities. I therefore report the estimated projection as a conditional association, with the limits of an upper-bound interpretation discussed in Section 6.2.5 and Appendix C.10.

The access counterfactual in Section 4.4 also requires these effective responses to remain applicable when schools are removed; that section distinguishes the access index from fully informed welfare.

5.3 Reduced-Form Comparisons

I compare displaced students with students who were not displaced within cells defined by municipality of residence, grade, year, poverty bin, and population-density quartile, following the comparison logic of coarsened exact matching (Iacus et al., 2012). The regressions include cell fixed effects, weight observations by the number of students, and cluster standard errors by municipality. I restrict these comparisons to the pre-María closure waves and estimate the outcome dynamics with the interaction-weighted event-study estimator of Sun and Abraham (2021), which allows the effects to differ across closure cohorts.

Identification of the dynamic effects requires that outcomes would have followed parallel trends within the comparison groups in the absence of closure, with no anticipation and no effect of the closures under study on comparison outcomes. Preclosure coefficients provide evidence on the parallel-trends restriction. The distance comparisons describe relocation within the school network and are presented in Section 6.1 together with the enrollment and achievement dynamics.

5.4 Closed-School Comparisons

I use the production estimates to compare the educational contributions of closed schools with those of surviving schools in the local menu. For $x \in \{\phi, q\}$, I regress the menu gap defined in Section 4.3 on the closed school's rurality and poverty:

$$\mathrm{gap}_x(o) \;=\; \alpha \;+\; b_{\mathrm{rural}}\,\mathbf{1}[\mathrm{rural}_o] \;+\; b_{\mathrm{pov}}\,\frac{\mathrm{pov}_o}{\mathrm{sd}(\mathrm{pov})} \;+\; \gamma_{cy(o)} \;+\; \varepsilon_o, \tag{12}$$

where pov_o is the school's continuous share of program-poor students, scaled by its cross-school standard deviation, and $\gamma_{cy(o)}$ denotes closure-year fixed effects. Regressions are weighted by displaced enrollment and standard errors are clustered by municipality. The four descriptive groups split origin schools at the median poverty rate; this school-level classification differs from the family-level poverty indicators in the choice model.

The coefficients describe which educational environments were removed relative to the surviving local menu. Their interpretation as differences in educational contributions rests on the production assumptions above, but the regressions concern selection among closed schools rather than causal effects of rurality, poverty, or closure. Since the benchmark uses surviving schools' preclosure enrollment, it is independent of where displaced students subsequently enroll; conditional mean-zero estimation error in school effects enters the dependent variable and affects precision. The primary sample contains 317 closed schools in

76 municipalities, with last operating years no later than 2017, and Section 6.3.3 examines sensitivity to including the 2018 closure wave.

6 Results

This section examines how the consequences of closure depend on the schools that remain. I begin with observed displacement outcomes and then present the production and school-choice estimates. These estimates distinguish school contributions from teacher composition and allow me to compare changes in educational exposure and access across groups. I report educational exposure and access in their respective units and samples.

6.1 Displacement Outcomes

I first examine how pupils responded to closure using the comparisons described in Section 5.3. Reassignment distances and destination shares show where displaced children enrolled, while the exit and achievement estimates describe their subsequent participation in the public system and performance.

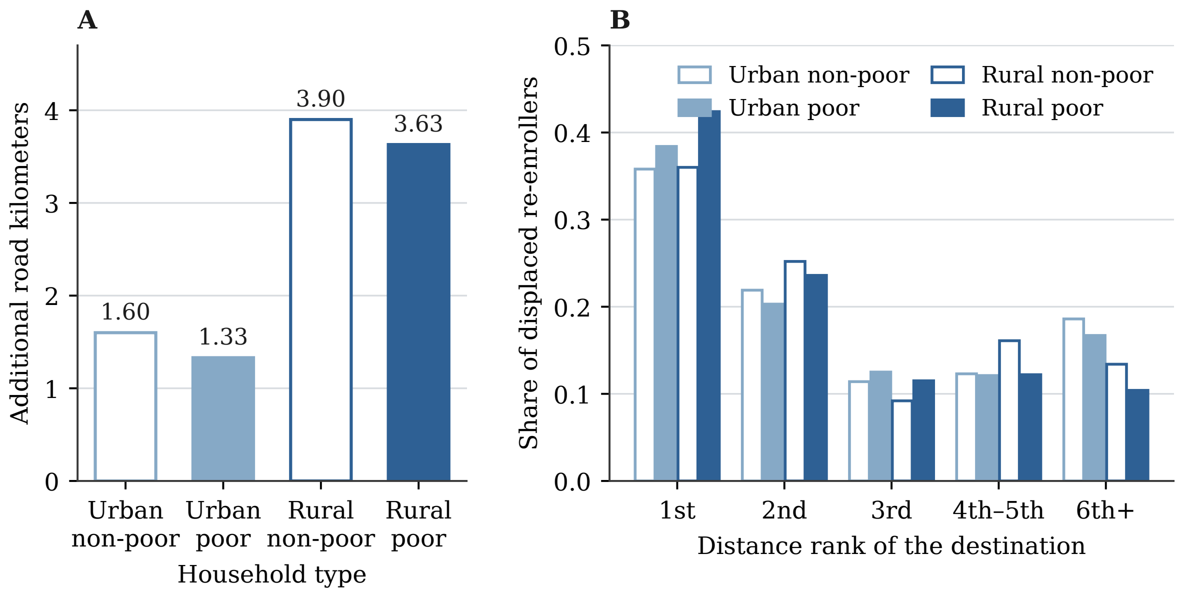
6.1.1 Reassignment Distances

Children from rural schools moved farther after closure: their destination school was on average 4.46 road kilometers from the closed school, with a median distance of 4.04, compared with 1.96 and 1.31 kilometers for children from urban schools. I measure these distances between schools for the 52,386 of 60,174 displaced children whose origin–destination pair could be routed, or 87.1 percent of the total. Figure 2 shows the distributions for each group.

Comparing displaced and non-displaced children gives a similar pattern, with displacement adding 1.60 kilometers to reassignment distance on average ($t = 14.2$) and a further 2.303 kilometers for children from rural schools ($t = 15.3$). Changing the comparison group puts this rural premium between 2.3 and 2.5 kilometers. The poverty difference is smaller and has the opposite sign: poor displaced children moved 0.269 kilometers *shorter* distances than non-poor displaced children ($t = -4.95$). Poor children are disproportionately urban, where schools are closer together. Across these comparisons, the implied additional travel is 3.63 to 3.90 kilometers for rural children and 1.33 to 1.60 for urban children.

The rural difference in ordinary school moves is much smaller. Among children who were not displaced, the rural main effect is 0.266 kilometers ($t = 5.56$), about one ninth of the 2.3-kilometer rural premium following closure. Longer moves may matter for attendance, as displacement distance predicts absences after

Figure 2. Displacement Distances and Destination Ranks



Notes: The figure uses displaced pupils from the pre-María period. Distances are OSRM road kilometers between the closed school and the destination. They measure school-to-school relocation rather than the change in the journey from home. *Panel A.* The bars show additional reassignment distance relative to comparable non-displaced children. I use coarsened exact matching within cells, with cell fixed effects, child weights, and standard errors clustered by municipality (about 78 clusters). The estimated distances are 1.60, 1.33, 3.90 and 3.63 kilometers. This is a difference in realized move distance, not a difference-in-differences estimate, because the outcome is a mechanical function of treatment when every treated child moves; the underlying mean distances are 4.46 kilometers for rural and 1.96 for urban displaced children, using the 52,386 of 60,174 displaced children whose origin–destination pair could be routed. *Panel B.* The bars show the shares of displaced re-enrollers whose destination is the nearest, second-nearest, third-nearest, fourth- or fifth-nearest, and sixth-or-more-distant surviving school in the local set. I calculate these shares directly from observed choices. Rural-poor children enroll at the nearest surviving school 6.7 percentage points more often than urban non-poor children (0.424 against 0.357), and the share falls monotonically across the more distant ranks. The nearest-school shares are similar among non-poor children (0.357 compared with 0.360). Among poor children, the shares differ (0.384 urban compared with 0.424 rural), consistent with differences in the proximity of surviving schools. No standard errors are reported for these shares, and they sum to one within rounding in each cell. Panel A uses the origin school's urban/rural flag and the administrative family-poverty flag. Panel B uses the kernel-weighted rural exposure of the closed school's catchment. The panels therefore use different definitions of rurality.

closure in Philadelphia (Steinberg and MacDonald, 2019), although measuring access losses also requires accounting for families' preferences and the alternatives available to them.

6.1.2 Destination Schools

Among displaced children who remained in public schools, the share enrolling at the *nearest* surviving school is 0.357 for urban non-poor children, 0.384 for urban poor, 0.360 for rural non-poor, and 0.424 for rural poor. Table 3 shows the full distribution of destination ranks. Rural-poor children attend the nearest school 6.7 percentage points more often than urban non-poor children. These are observed shares, which I report without standard errors. The comparison within poverty groups shows almost the same nearest-school attendance in rural and urban markets among non-poor children, at 0.360 and 0.357, whereas among

Table 3. Displacement, Distance, and Public-School Exit

	Urban non-poor	Urban poor	Rural non-poor	Rural poor
Panel A. Group comparisons				
Share landing at the nearest surviving school	0.357	0.384	0.360	0.424
Additional travel relative to comparable non-displaced pupils (road km)	1.60	1.33	3.90	3.63
	Rural	Urban	Poor	Non-poor
Panel B. Contrasts available only on the margins				
Realized reassignment distance, mean (road km)	4.46	1.96	3.91	4.30
median	4.04	1.31	—	—
N with a routed origin–destination pair	52,386		21,932	3,841
Raw closure-induced exit (pp)	3.9	4.6	2.9	5.2
Exit gradient within residential cells (pp)	n.s.		-1.9 ($t = -2.39$)	

Notes: Panel A compares groups by rural exposure and poverty. Panel B reports the available marginal contrasts. Each observation is a displaced pupil. The program displaced about 60,000 pupils; I obtain an origin-to-destination route for 52,386 of 60,174 pupils (87.1 percent). *Panel A.* I calculate the rank-one share directly from pre-María choices. It is the share of displaced re-enrollers who attend the nearest surviving school in the local set. I report this share without a standard error. Additional travel compares the distance moved by displaced pupils with that of comparable non-displaced pupils. I use coarsened exact matching within cells, with cell fixed effects, child weights, and standard errors clustered by municipality (about 78 clusters). This is a difference in realized move distance rather than a difference-in-differences estimate. *Panel B.* Distances are OSRM road kilometers between the closed school and the destination, rather than changes in the journey from home. The rural and urban columns show raw means for displaced pupils with a routed pair. The poor and non-poor columns show mean realized distances for displaced re-enrollers in the demand frame. Raw closure-induced exit measures the increase in exit from the public system for each group in the pre-María period. The exit-gradient row reports the coefficient on displacement interacted with the group indicator, holding the residential cell fixed. Poor pupils have about 2 percentage points *less* exit (-0.0187, $t = -2.39$). The rural interaction is not statistically distinguishable from zero (no point estimate is reported). Across designs and horizons the causal exit effect lies between 2.2 and 4.4 percentage points, with about 3 points as the central figure. Rurality uses kernel-weighted catchment exposure in the rank-one row and the prior school's urban/rural flag in the other rows. Poverty uses the administrative family-poverty flag. These rurality measures differ from the demand model's block-group density classification.

poor children the rural share is higher by 4.0 percentage points. Thus, the rural difference in nearest-school attendance is concentrated among poor children.

6.1.3 Public-School Exit

Closure increased the probability of leaving the public system by roughly 3 percentage points, compared with annual baseline attrition of about 7 percent. Across comparison groups and follow-up horizons, the estimates range from 2.2 to 4.4 points, with a larger increase at longer horizons. Poor displaced children were about 2 percentage points *less* likely to leave than non-poor displaced children. Once residential cells are held fixed, rurality has no association with exit.

Some children subsequently returned: 26.6 percent of displaced leavers re-entered, compared with 11.5 percent of comparable non-displaced leavers, and the adjusted difference within cells is 9.0 percentage points ($t = 3.9$). Once I account for this re-entry, the permanent extensive-margin loss is at most about 2.4 percentage points.

The broader participation pattern also differs by income. The share outside the public system is 8.7 percent among income-poor children and 42.6 percent among the non-poor. Non-enrollment is the same in both groups, at 1.8 percent, so private-school enrollment accounts for the difference. These shares show poorer families' greater reliance on public schools, although they cannot establish whether a particular family could afford a private school. The event-time estimates and pre-closure placebo appear in Appendix A.

6.1.4 Achievement Outcomes

The average achievement estimates are small and statistically indistinguishable from zero. For mathematics, the interaction-weighted estimate averages $+0.039$ over the four post-closure years (standard error, SE, 0.031; $p = 0.202$). The joint test of the pre-period coefficients does not reject. The matched design used here gives -0.063 on normalized scores ($t = -1.41$), with a pre-closure placebo of $+0.002$, also indistinguishable from zero. Using the same closures and administrative universe, the companion study finds that enrollment initially declines and recovers within two years, with little evidence of a persistent average achievement effect but gains for children from underperforming schools and losses for those from overperforming schools (Bobonis et al., 2022). These comparisons measure changes in students' own scores, whereas Section 6.3 examines changes in the school component of educational production; averaging those school-effect contrasts would therefore not give the average treatment effect on achievement.

Linked labor-market records show reductions in high school completion, college attendance, and earnings in the mid-twenties, with larger reductions among secondary-school students and economically disadvantaged children (Kim, 2026). Since my panel ends in 2019, I cannot examine those longer-run outcomes.

The achievement changes also differ among displaced children. For children from rural schools, higher poverty at the origin school predicts a larger decline in the child's own score. The association grows over time and is absent in the two years before closure, while among urban displaced children and voluntary movers it is statistically indistinguishable from zero. Appendix A reports the estimates and the imprecisely estimated rural–urban difference. Since this comparison concerns variation within the displaced population, it does not measure the average effect of closure; poverty here refers to the origin school, and the outcome is the change in the child's own score.

To examine these outcomes further, I use the production estimates to distinguish school and teacher contributions and the choice estimates to measure access costs. Section 6.2 presents both sets of estimates.

6.2 Achievement and School-Choice Estimates

6.2.1 Achievement Forecasts

I first examine whether the estimated effects predict score changes after forced moves. Forecast tests usually use variation outside the estimation sample (Angrist et al., 2017), but some observations here enter both estimation and the forecast exercise, limiting its value as an independent test. The sample contains 24,263 displaced pupils whose score change and both schools lie in the connected set. Regressing their realized score change on the change in the composite school effect gives an attenuation-corrected coefficient of 0.940, with a bootstrap standard error of 0.011, compared with an uncorrected coefficient of 0.916. The composite therefore predicts score changes approximately one-for-one in this sample. Correcting attenuation does not resolve the overlap between the estimation and forecast samples, and the forecast concerns the composite effect. Since moving an offset from the school component to the teacher component leaves their sum unchanged, coefficients near one cannot validate the decomposition on their own, even in a regression with both components (Appendix B). The appendix describes the samples and quality measures used in each forecast exercise.

I also assess reliability by splitting classrooms, without using the calibrated variance inputs. The implied reliability of the full-sample school effects is approximately 0.63, with similar results across two independent splits and two weighting schemes. Reliability is 0.80 for the change in school effects across a forced move, the quantity used in the child-level comparisons in Section 6.3. Individual teacher–subject effects have reliability of approximately 0.35. I therefore use posterior means when individual teacher effects enter a regression.

For small rural schools, the estimation error in $\hat{\phi}_j$ is about as large as the main closure contrast, so I do not rank individual schools by their estimated effects. Averaging across schools can reduce uncertainty in cell means, and using the estimated gap as an outcome avoids errors-in-variables attenuation when measurement error is mean zero, although these averages still have sampling uncertainty, which I report.

6.2.2 School and Teacher Effects

Table 4 shows the variance components using the two-way decomposition format of Card et al. (2013), with the classroom-split estimates in a separate column. Correcting for estimation noise substantially reduces the dispersion of each component. Before correction, the standard deviations are 0.44 for teacher–subject effects, 0.44 for school effects, and 0.25 for the school-average teacher component. After correction, the standard deviation of teacher-by-subject value-added is 0.225 and ranges over $[0.14, 0.28]$ across the cali-

Table 4. Teacher and School Variance Components

	Uncorrected plug-in	Leave-out estimate	Range across calibrations	Classroom split
Panel A. Standard deviations, in test-score standard deviations				
Teacher-by-subject value-added, σ_μ	0.437	0.225	[0.14, 0.28]	0.253 / 0.240
School effect, σ_ϕ	0.445	0.361	[0.34, 0.38]	0.339 / 0.323
School-average teacher component, $\mathrm{sd}(q)$	0.252	0.089	[0, 0.15]	0.141 / 0.112
Composite school effect, $\mathrm{sd}(\Psi)$	0.409	0.393	$\pm 2\%$	—
Teacher–school match, σ_ψ	—	0, bound [0, 0.10]	—	—
Panel B. Derived quantities				
$\mathrm{Var}(\phi)/\mathrm{Var}(\mu)$		$0.1305/0.0504 = 2.6$		
Reliability of $\hat{\phi}$		0.63	[0.569, 0.697], [0.563, 0.681]	
Reliability of $\Delta\hat{\phi}$ across a forced move		0.80		

Notes: The unit of observation is a pupil–subject–year record. I estimate (1) using scores normalized within subject, grade, and year. Panel A reports standard deviations in these test-score units. The variance ratio and reliabilities in Panel B are dimensionless. The estimation uses 3,452,029 observations from academic years 2011–2019, with 395,030 pupil effects and 1,274 schools. Teacher and school effects are jointly identified on the connected mover graph. Column 1 reports the naive plug-in dispersion of estimated effects, and column 2 applies the leave-out correction. Column 3 gives the range of point estimates as the calibrated classroom-year and school-year variance components vary over the calibration surface. Column 4 reports cross-half covariances from two random classroom partitions, each with two disjoint halves, for four half-sample estimations in total. Columns 1–3 are unweighted across schools or teacher–subjects. In column 4, school-level estimates are observation-weighted and teacher–subject estimates are unweighted. I report the corresponding unweighted school estimates in Appendix B. Brackets in Panel B contain 500-resample school bootstrap intervals. The variance ratio in Panel B uses the leave-out estimates, rather than the split-sample estimates with different weights. I report bounds for the school-average teacher component and the match component. For the teacher component, the leave-out point and split estimate cannot be separated once variation in split assignments is accounted for. The shared estimation-error component explains at least 85 percent of the variance of estimated school-average teacher value-added across the four half-sample estimates. For the match component, three independent approaches using 1,626 closure movers, including 469 with two scored years on both sides, give upper limits of 0.105, 0, and 0.095. The $[0, 0.15]$ range in column 3 for school-average teacher quality concerns the leave-out point as the two calibrated inputs vary. It differs from the bound on this component in Section 6.2.2, roughly 0.09 to 0.15. The covariance between the school effect and school-average teacher quality cannot be distinguished from zero. Its implied correlation also changes sign across statistically identical half-sample estimates, so I report no correlation between the two components. Appendix B presents this diagnostic together with the calibration surface, the reliability of individual teacher–subject effects, and the comparison with the value-added literature.

bration surface, while the two split-sample estimates are 0.253 and 0.240. For school effects, the corrected standard deviation is 0.361, with a calibration range of $[0.34, 0.38]$; splitting the sample gives estimates of 0.339 and 0.323 when weighted by scored observations, and 0.375 and 0.373 without weights. Together, these approaches place school-effect dispersion between 0.32 and 0.38. For the school-average teacher component, the standard deviation is roughly 0.09 to 0.15, well below the uncorrected estimate. The leave-out estimate of 0.089 lies at the lower end of this range. The composite school effect has a standard deviation of 0.393 and is the component least sensitive to the calibrated inputs.

School effects account for more variation than teacher–subject effects: their variance is 2.6 times as large, at 0.131 compared with 0.050. Every correction in the table gives the same ordering. Since the two calibration surfaces do not overlap, changing the calibrated classroom-shock input does not reverse this comparison at any point on the surface. These results motivate examining how closures change children's exposure to school effects, a quantitatively important component of achievement. Although the model separates this component from teacher inputs, the variance comparison alone cannot establish whether teacher reassignment could offset a loss in school exposure.

The school-level teacher component is smaller because it averages over the teacher roster, and teacher quality varies less *across* schools than across classrooms. At the main leave-out estimates, school effects account for 94 percent of $\mathrm{Var}(\phi) + \mathrm{Var}(\bar{q})$ and the teacher component for 6 percent. This calculation excludes the covariance between the components; the reported shares are therefore not shares of $\mathrm{Var}(\Psi)$. The comparison of school averages still allows substantial variation in the effectiveness of individual teachers. The covariance between the school-level components is statistically indistinguishable from zero, and its implied correlation changes sign across equivalent split-sample estimations. For this reason, I do not report a signed correlation between the components. Appendix B presents this diagnostic and the calibration surface, and compares the estimated dispersion with the teacher value-added literature.

The ANCOVA comparison in Appendix B.3 examines sensitivity to replacing pupil fixed effects with lagged-score and demographic controls. Across teacher–subject pairs estimated in both specifications, the correlation is 0.849, with each specification using its available estimation sample. This is a comparison of estimated effects, without a correction for estimation error, and does not establish robustness of the corrected variance components in Table 4.

6.2.3 Teacher–School Matches

If a teacher is more productive at some schools than at others, the effect of reassignment depends on the match at both the origin and destination. I denote the match component by ψ_{kj}, which adds a term specific to the teacher–school pairing to (1) and has mean zero across pairings. To estimate its dispersion, I use teachers observed at two schools, whose residual performance changes by the match difference plus estimation noise, for which the leave-out estimator corrects. The sample contains 1,626 closure movers with a scored year on each side of the move, and 469 with at least two scored years on both sides. Since the panel ends in 2019, no teacher displaced in the 2018 wave enters the stricter sample.

The point estimate of match dispersion is zero, and the reported bound is $\sigma_\psi \in [0, 0.10]$. Three approaches to correcting estimation noise give upper limits of 0.105, 0, and 0.095. This is a bound on the dispersion of matches, rather than an estimate of the share of teacher-effect variance that matches explain. The placebo in Appendix B uses teachers who stayed at the same school to isolate mild decay in teacher value-added over time, although the approaches used for the bound do not depend on the placebo or its decay estimate.

Jackson (2013) finds higher teacher effectiveness after a move and attributes roughly a quarter of its variation to match quality. The settings, types of moves, and estimators differ, which limits a direct comparison. In my sample, closure forces teachers to move, and although the leave-out estimator corrects noise in the mover's contrast directly, the mover's own pupils also contribute to the estimated school effects, which can attenuate the contrast. I therefore report a dispersion bound and do not interpret the zero point estimate as evidence that teacher–school matches have no effect.

Having estimated school and teacher contributions to achievement, I next examine how families choose among the schools that provide those inputs. Section 6.3 uses the production estimates to measure the changes in educational exposure associated with closure.

6.2.4 School-Choice Parameters

Table 5 presents the school-choice estimates: I estimate the three distance coefficients and the inside-option shifter, report the rural-quality coefficient as a sensitivity range and the common quality coefficient as a projection, and calibrate the nesting parameter and distance exponent. The base distance coefficient is $\lambda_0 = -0.0429$ (0.0030), with $t = -14.28$ and inference based on 78 municipality clusters.

To compare the distance coefficients, Panel B translates them into the marginal utility cost of a kilometer at the median displaced move of 3.4 kilometers, where the cost increases from urban non-poor to urban poor, rural non-poor, and rural-poor families. Rural families face 2.4 times the cost for urban non-poor families; for rural-poor families, the ratio is 2.7.

To estimate the poverty increment, I use the observed administrative family-poverty indicator and integrate over unobserved residence. Since poor and non-poor model households within a block group face the same distances and school mean utilities, their destination moments compare choices conditional on school–market mean utility, which remains in the choice probabilities. The estimate of λ_{pov} is one-sixth of λ_{rural}, increasing the marginal cost of distance by about a quarter in urban markets and a tenth in rural markets. Observed reassignment distances show that poor displaced children moved 3.914 kilometers on average, compared with 4.298 for non-poor children. The inside-option shifter, π_{pov}, is $+21.3$ in within-nest units, or three times the median distance disutility. Thus, the larger poverty difference concerns participation in public schools, while rurality accounts for the larger difference in estimated distance costs.

Table 5. School Demand Parameters

Parameter	Symbol	Estimate	Status
Panel A. Utility parameters			
Distance, base	λ_0	-0.0429 (0.0030)	estimated
Distance, rural increment	λ_{rural}	-0.0612 (0.0071)	estimated
Distance, family-poverty increment	λ_{pov}	-0.0102 (0.0020)	estimated
Income-poverty inside-option shifter	π_{pov}	$+0.9898$ (0.0741)	estimated
Teacher quality, family-poverty increment	β_{pov}	-0.0393 (0.0084)	estimated
Teacher quality, rural increment	β_{rural}	$[-0.14, +0.20]$	sensitivity range
Teacher quality, common projection	b_q	$+0.1393$ (0.0716)	projection
Lagged capacity		-1.2×10^{-5} (9.3×10^{-5})	estimated
Urban-school indicator		-0.0191 (0.0494)	estimated
Nest parameter	ρ	0.953524	**calibrated**
Exponent on distance	p	0.8	**calibrated**
Panel B. What the distance parameters imply			
		Utils per km at 3.4 km	Km per SD of δ
Urban non-poor		-0.0269	3.46
Urban poor		-0.0332	2.65
Rural non-poor		-0.0652	1.14
Rural poor		-0.0716	1.02
Rural to urban ratio		2.4×	
Rural-poor to urban-non-poor ratio		2.7×	
Auxiliary travel per SD of teacher quality		sweep endpoint ≈ 150 meters	

Notes: Each observation is a school–market row. Origin years are 2013–2017 and destination years are 2014–2018, labeled by the end of the academic year. The sample contains 3,120 municipality–grade–year markets, 1,281 schools, 31,767 product–market rows, 26,675 displaced micro rows, 285,800 agent rows, and 78 municipality clusters. I use the utility specification in (2), with one nest containing same-municipality public schools, a singleton outside nest, region-by-year fixed effects, and road distance in kilometers. I estimate the nested-logit model by GMM using 34 moments: 32 destination moments, one composition moment, and one exit moment (Appendix Table C1). The exit moment remains in the criterion with ρ fixed; its fit is not an independent validation. Parentheses report standard errors clustered by the 78 municipalities. These standard errors condition on the calibrated nesting parameter. The interval for rural quality summarizes sensitivity across specifications and does not define an identified set. The coefficient cannot be separated from the smaller choice sets in rural markets; controlling for menu size gives -0.03, SE 0.11. I obtain the common quality projection by regressing δ on lagged quality, without an excluded instrument. I use the nesting parameter from an earlier joint GMM estimation and hold it fixed in the baseline run to match the calibration target for leaving the local public-school choice set. I therefore do not report a standard error on ρ from that run. The exponent on distance is measured from road geometry (Appendix Figure C2). *Panel B.* I evaluate the marginal utility of distance at $d = 3.4$ km, the median displaced move, using $\partial v/\partial d = 0.8\lambda_\tau d^{-0.2} < 0$. Marginal cost is the negative of this expression. The kilometer column shows the tradeoff between distance and one within-market standard deviation (SD) of mean utility (0.116 utils), measured relative to a zero-distance alternative: $\kappa_\tau = [0.1158/(-\lambda_\tau)]^{1/0.8}$. It is not additional travel evaluated at 3.4 km. For the travel calculation, I use a homogeneous share index, the upper 95 percent endpoint of the projection coefficient, the smallest absolute distance coefficient on an auxiliary grid, and reliability 0.125. The point figures range from 15 to 45 meters. The grid omits the baseline model's base-distance coefficient, whose absolute value is smaller. Its endpoint therefore does not provide an upper bound across baseline household types. Appendix C.10 states the additional sign, measurement-error and normalization assumptions. The rural increment uses variation in kernel-weighted rural exposure across closure catchments. It therefore compares catchments rather than rural and urban families within the same catchment. The within-municipality version of the underlying contrast is about 39 percent of the pooled one.

6.2.5 Demand for Teacher Quality

Since the destination moments do not identify the common quality coefficient, Table 5 reports the projection coefficient $b_q = +0.139$ (0.072), which I obtain by regressing recovered mean utility on lagged quality and school controls, including region-by-year fixed effects. Its interpretation requires assumptions about measurement error and the unobserved school–market demand component.

For a separate travel calculation, I construct a homogeneous nested-logit index from aggregate shares at the calibrated ρ and regress it on q with market fixed effects, obtaining a slope of $+0.0021$ (0.0025). This index differs from the mean utilities recovered by the heterogeneous model; after adjusting for the reliability of $\hat{q}$, the implied distance tradeoff is 15 to 45 meters across the recorded sweep, and about 150 meters at its upper endpoint. These auxiliary values depend on the grid: the upper endpoint combines its smallest absolute distance coefficient, the upper end of the 95 percent interval for the slope, and reliability of 0.125, while the smaller absolute base-distance coefficient of the baseline model is excluded. An upper-bound interpretation would further require restrictions on quality measurement and mean-utility recovery, nonnegative conditional covariance between true quality and omitted determinants of school utility in the auxiliary projection, and a common utility normalization (Appendix C). I therefore do not report these values as an identified bound on households' willingness to travel.

The micro moments identify the poverty contrast $\beta_{\text{pov}} = -0.0393$ (0.0084), $t = -4.71$. Poor families place a lower estimated utility weight on measured teacher quality conditional on school–market mean utility, a contrast that I estimate jointly with distance responses using group differences in destination moments and conditioning on recovered mean utilities. This describes quality-related substitution under the information and access conditions that families face, but it cannot separate these mechanisms or identify peer sorting: within choice sets, the correlation between q and school poverty is 0.011. Here, q measures the school-average teacher value-added. As Section 6.3 shows, the mean teacher-quality gap between closed schools and surviving menus is statistically indistinguishable from zero.

For the rural-quality contrast, the comparison is between zones with different school menus, and the design cannot separate menu size from rurality. When I control for menu size, the coefficient is -0.03 (0.11); I therefore report a sensitivity range across specifications rather than a formal identified set.

6.2.6 Demand Sensitivity and Interpretation

One concern for the estimated utility levels is the measure of market size. With resident school-age population as the denominator, the population-weighted inside share is 0.5481, compared with public-school participation of 0.7645 in the same census extract. The difference is 21.6 percentage points. Of this, 7.8 points reflect children attending schools across municipal boundaries, whose schools are excluded by the same-municipality restriction; the remaining 13.8 points reflect population counts exceeding realized enrollment.

I recorded this discrepancy before estimation. Adding a rural parameter to absorb it could attribute a geographically uneven population error to preferences, so the model includes no such adjustment. Since the discrepancy affects the extensive margin, I calibrate ρ to departure from the local public-school choice set using a target that combines estimated public-system exit with an approximate adjustment for cross-municipality switching. The reported access losses are conditional on this calibration.

Although the factor $1/(1-\rho) = 21.5$ scales the utility index within the public-school nest, it does not multiply access losses or kilometer equivalents by a common amount. Aggregate inside shares also cannot establish how the market-size discrepancy affects comparisons across types, since in a heterogeneous model it need not shift all mean utilities equally and access-loss shares can vary with the nest parameter. Appendix D examines this sensitivity for the earlier distance-only calculation, and Appendix C reports the evidence on the outside option.

I examine three related concerns about the rural distance increment, with the diagnostics reported in Appendix C. First, the observed moment compares rural *catchments*, since the records do not identify each student's residential block group; I weight moment cells by the catchment's rural share, whereas rurality in utility refers to the model household's block group. Although the observed targets and model predictions apply the same conditioning rule, the data therefore do not provide a moment based on observed household rurality.

Second, most of the identifying variation is *between* municipalities. The rural-minus-urban contrast in rank-one attendance is $+0.0262$ when calculated within municipality, compared with $+0.0669$ in the pooled sample. The within-municipality contrast is about two-fifths as large, showing sensitivity to the source of variation without bounding the structural coefficient.

Third, the magnitude changes with the distance specification. The rural-to-urban ratio is 2.4 under $d^{0.8}$ and 1.8 under $\log(1+d)$, although its sign is stable across all structural runs. For families living near a boundary, the same-municipality restriction may also exclude nearby schools and affect both $|\lambda|$ and the estimated access cost, although this does not establish a general lower bound for either. Section 6.4 examines the distribution of access losses when all households receive the urban non-poor distance coefficient, using the distance-only calculation.

The production estimates measure school and teacher contributions to achievement, while the choice estimates describe substitution and access costs. I next use these estimates to examine the educational exposure and access consequences of closure, without imposing a separately identified preference for school effects.

6.3 School Closures and Educational Exposure

6.3.1 Closure Selection

In Table 6, I compare closed schools with surviving local alternatives using estimated school effects and model-based access costs to evaluate the observed closure set. Among the 317 schools with the required estimates, the mean school-effect rank within the market is 0.499, compared with a random-selection benchmark of 0.502. The mean ranks for total and per-child access costs are 0.404 and 0.479. The ranks describe the schools selected for closure without identifying the decision makers' criteria, and the access ranks use the travel cost of school removal, which differs from the logsum measure of access loss.

Using Equation (12), I compare the estimated effects of closed schools with those of surviving schools in the same municipality to describe the composition of closures, rather than to estimate a treatment effect on pupils. Section 2 explains why these comparisons cannot recover the decision makers' selection criteria. The average within-market school-effect rank is near one half (Table 6). Conditional on municipality and scale, regressing school effects on an indicator for ever closing gives -0.017, with a standard error of 0.031. School effects are also weakly associated with urban status, poverty, and grade span, with observables together explaining about 9 percent of true school-effect variance, almost all through school size.

The overall rank conceals differences across local comparisons, and the location of alternatives offers one possible explanation: although the median rural market contains more schools than the median urban market

Table 6. Closure Selection and Estimated Costs

	Mean percentile	p-value
School effect, against the surviving menu	0.499	0.90
random-selection null	0.502	
Access cost of the catchment, total	0.404	< 0.001
Access cost of the catchment, per child	0.479	0.003

Notes: The unit of observation is a closed school. I rank each closed school against its surviving local alternatives: schools in the same municipality, operating in the year after closure, with an overlapping tested-grade span. I use the midrank $(r-0.5)/n$ and average these within-market percentiles across the 317 closed schools with both an estimated school effect and an access-cost measure. Random draws on the same data give a random-selection benchmark of 0.502. School effects are the estimated production component $\hat{\phi}_j$. To measure access costs, I use the estimated school-demand model to evaluate school removal in total and per child. Both are estimated quantities, rather than directly observed outcomes. An average rank close to the random-selection benchmark does not establish random selection. A rank below that benchmark also does not establish that access was ignored when schools were selected for closure.

(Appendix Figure A1), rural displaced children travel more than twice as far to their destinations. School counts therefore give an incomplete picture of accessibility, and the municipal school-effect benchmark used below does not determine which alternatives a particular family could reach.

6.3.2 Closed Schools and Local Alternatives

Table 7 and Figure 3 show the local comparisons. After controlling for school poverty and closure-year effects, I find a menu gap for rural closed schools that is 0.148 standard deviations larger than for urban closed schools, with a standard error of 0.046 and a wild cluster bootstrap p-value of 0.011. This is the adjusted rural–urban difference, rather than the average gap for rural closures. Only 36 of the 76 municipalities contain both rural and urban closures, so I report the bootstrap p-value (Cameron et al., 2008; MacKinnon and Webb, 2017). Weighting surviving schools by road proximity instead of enrollment gives a coefficient of 0.110, with a standard error of 0.047.

Sixty-one percent of rural-poor closures removed a school with an estimated effect above the local benchmark, compared with 54 percent of rural non-poor closures, 34 percent of urban non-poor closures, and 25 percent of urban-poor closures. Because these percentages concern estimated gaps, mean-zero measurement error alone does not make them lower bounds on the shares of schools with positive true gaps. The mean gaps show a related but non-monotonic pattern: in urban markets they are -0.059 for non-poor closures and -0.097 for poor closures, while in rural markets they are -0.004 and $+0.137$, respectively. On average, urban closures therefore removed schools with *lower* estimated effects than the surviving alternatives, as in the accountability closures studied by Bifulco and Schwegman (2020) and Carlson and Lavertu (2016). The comparison is reversed for rural-poor schools. Inference must account for multiple comparisons because I examine all four rural-by-poverty groups; I report every cell estimate, but not a stepwise-adjusted test of the largest contrast (Romano and Wolf, 2005).

When I allow the rural–urban difference to vary with origin-school poverty, the rural-by-poverty interaction is $+0.201$, with a standard error of 0.100 and a bootstrap p-value of 0.053; this is not a comparison of rural-poor schools with all other schools combined. Since the total interaction falls short of conventional statistical significance, while the residual component discussed below meets the 5 percent threshold, I report the total alongside all four cell means.

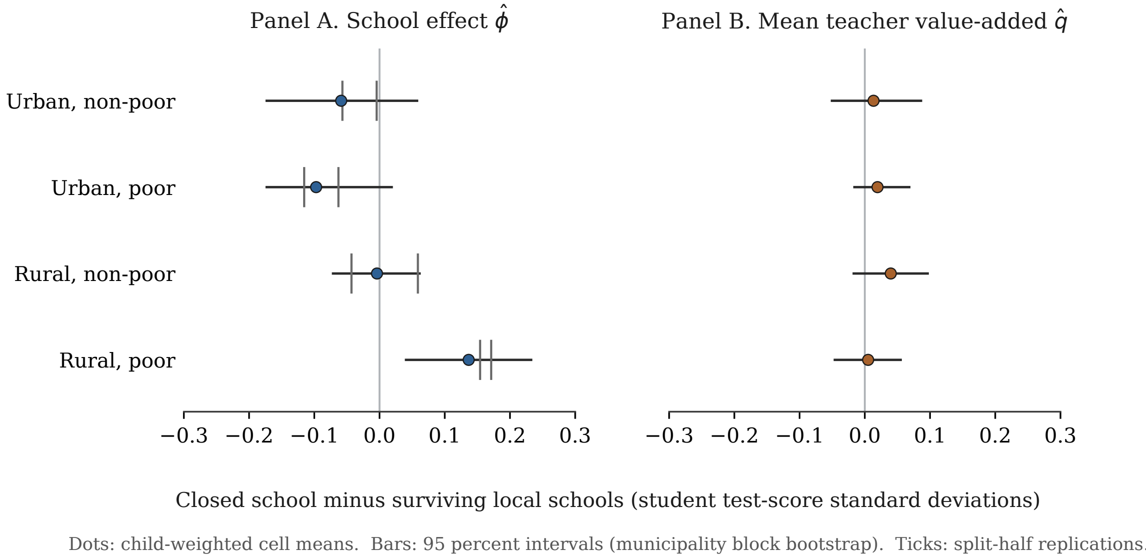
I next examine whether school size, tested-grade span, and catchment breadth, defined as the number of distinct residence markets from which a school draws, account for this interaction. Of the $+0.201$ interaction, these characteristics predict -0.021, with a standard error of 0.021, under one set of gradients, and -0.012, with a standard error of 0.023, under the other. Neither estimate is statistically distinguishable

Table 7. Local School-Effect Gaps

	Urban non-poor	Urban poor	Rural non-poor	Rural poor
Panel A. The gap to the surviving local menu, by cell				
Mean gap in the school effect, gap_ϕ	-0.059	-0.097	-0.004	$+0.137$
Share of closures above their local menu	0.338	0.246	0.539	0.610
Closed schools (n)	39	66	115	97
Panel B. Regression of gap_ϕ *on school characteristics*				
	Coefficient		Inference	
Rural, baseline specification	$+0.148$ (0.046)		wild $p = 0.011$	
Rural $\times$ poor	$+0.201$ (0.100)		$p = 0.053$	
Menu weighted by road proximity, not enrollment	$+0.110$ (0.047)		$t = 2.36$	
Both measured biases subtracted	$+0.135$ (0.048)		$t = 2.83$	
Adding log school size	$+0.099$ (0.047)		$t = 2.08$	
Full window, including the 2018 wave	$+0.070$ (0.043)		$t = 1.63$	
School-average teacher value-added gap, gap_q	-0.003 (0.027)		$t = -0.12$	

Notes: Each observation is a closed school. The sample contains 317 schools that closed before Hurricane María in 76 municipalities. Only 36 municipalities have both rural and urban closures. For a school o that closes in year cy, I define the surviving menu as schools in the same municipality that operate in $cy+1$ and whose tested-grade span overlaps o's. The gap is $\text{gap}_\phi(o) = \hat{\phi}_o - \sum_{j \in M(o)} w_{oj} \hat{\phi}_j$ with w_{oj} proportional to preclosure enrollment and normalized to sum to one within the origin's menu. I use the same weights for the school-average teacher value-added gap and measure teacher value-added at both the origin and the surviving schools in the preclosure year. A positive gap means the closed school had a higher estimated effect than the enrollment-weighted mean of its surviving local alternatives. Units are student test-score standard deviations normalized within subject, grade and year. I estimate the regressions by child-weighted least squares with closure-year fixed effects. Parentheses report standard errors clustered by municipality. "Wild p" is the p-value from a wild cluster bootstrap-t with the null imposed, Rademacher weights, and municipality clusters. This procedure addresses the small number and unequal size of the clusters (Cameron et al., 2008; MacKinnon and Webb, 2017), following Roodman et al. (2019). Measurement error enters the outcome. If it has conditional mean zero, it does not cause errors-in-variables attenuation. In Panel A, I weight cell means by children and give each school equal weight when calculating shares. The shares describe positive estimated gaps and do not provide lower bounds on the shares of positive true gaps. The additive regression uses the continuous school poverty share divided by its standard deviation; the four cells use a median split. The interaction is marginal and is not described as statistically significant. I quantify two possible biases: differences in testing participation and the reduction in a receiving school's measured effectiveness when it receives displaced pupils. Together they imply biases of 0.013, compared with a coefficient of 0.148. The row "both measured biases subtracted" removes both. Only the last robustness row extends beyond the pre-María sample. Adding the post-María 2018 wave reduces the estimate because those schools were well above their local alternatives on average, with no rural gradient. This pattern differs from the pre-hurricane waves and does not identify the criteria used to select schools in that round. Table B3 reports the full set of robustness specifications, including municipality fixed effects and the restriction to schools with low absorption. The teacher-value-added row reports the adjusted rural–urban contrast in preclosure school-average teacher value-added gaps. This comparison excludes later teacher inflows.

Figure 3. Closed Schools and Local Alternatives



Notes: The figure uses 317 schools that closed before Hurricane María (closure years 2013–2017) in 76 municipalities. Each school has an estimable school effect and at least one surviving school in its local menu. The unit of observation is a closed school. Menus and gaps are defined in (4) and (5). I weight surviving schools by pre-closure enrollment and normalize the weights to sum to one within each origin school's menu. A positive gap means the closed school had a higher estimated effect than the enrollment-weighted mean of its surviving local alternatives. Panel A uses the estimated school effect $\hat{\phi}_j$; Panel B applies the same construction to the school-average teacher value-added $\hat{q}_{j,t}$, the observation-weighted mean of teacher-by-subject value-added, measured in the origin school's pre-closure year for both the origin and surviving schools. Both panels share one horizontal scale. Units are student test-score standard deviations, normalized within subject, grade and year. Markers are child-weighted cell means: -0.059, -0.097, -0.004 and $+0.137$ in Panel A, and no cell distinguishable from zero in Panel B. Horizontal bars are 95 percent intervals from a nonparametric bootstrap that resamples the 76 municipalities. The two short vertical ticks at each estimate show the cell mean calculated separately on two disjoint random halves of the classroom sample. Agreement across halves provides a replication check. Using the products of these estimates to correct for noise also requires uncorrelated estimation errors. In child-weighted regressions of gap_ϕ on origin-school characteristics with closure-year fixed effects and standard errors clustered on municipality, the rural coefficient is $+0.148$ (SE 0.046; wild cluster bootstrap $p = 0.011$) and the rural-by-poverty interaction is $+0.201$ (SE 0.100; $p = 0.053$), reported in Table 7.

from zero. For the additive rural coefficient, the same exercise attributes 31 to 42 percent to observables, but the residual is statistically insignificant under conservative inference that propagates sampling error in the gradients, so the significance claims apply to the totals. Appendix Table B5 reports all four cells and both inference approaches. The exercise leaves the concentration of positive menu gaps among rural-poor origins unexplained. To see how these local gaps relate to the school effects experienced by displaced children, I subsequently compare their destinations.

6.3.3 Robustness

Table 7 presents the main robustness checks, and Appendix Table B3 reports the full sequence. Changing corrections and controls puts the coefficient between $+0.135$ and $+0.154$, still statistically distinguishable from zero, while dropping each closure wave in turn gives positive estimates with t-statistics of at least 2.00, ranging from +0.105 to +0.176. Excluding the most influential wave, 2015, reduces the coefficient by 29 percent.

Differential testing and absorption of displaced pupils are two other possible sources of bias: closed schools tested *more* of their pupils than the surviving schools, while rural and urban schools absorbed displaced children at similar rates. Together, the implied biases sum to +0.013, less than 9 percent of the coefficient. School performance may also have changed just before closure, but it was above the schools' own averages in the preceding years, by +0.029 in the year immediately before closure, so the time-invariant estimate *understates* their observed preclosure performance.

Controlling for log school size reduces the coefficient to +0.099, with a standard error of 0.047, as discussed in Section 6.3.4. Adding the post-María 2018 wave reduces it to +0.070; schools closed in that wave were well above their local alternatives on average, without a statistically detectable geographic gradient, although this difference cannot identify how storm damage affected closure selection. Finally, using the composite school effect avoids dependence on its allocation between school and teacher components and gives a rural coefficient of +0.145 to +0.149, with a rural-poor versus urban-poor difference of 0.248 standard deviations, or roughly seven tenths of school-effect dispersion.

6.3.4 School Size and the Rural Contrast

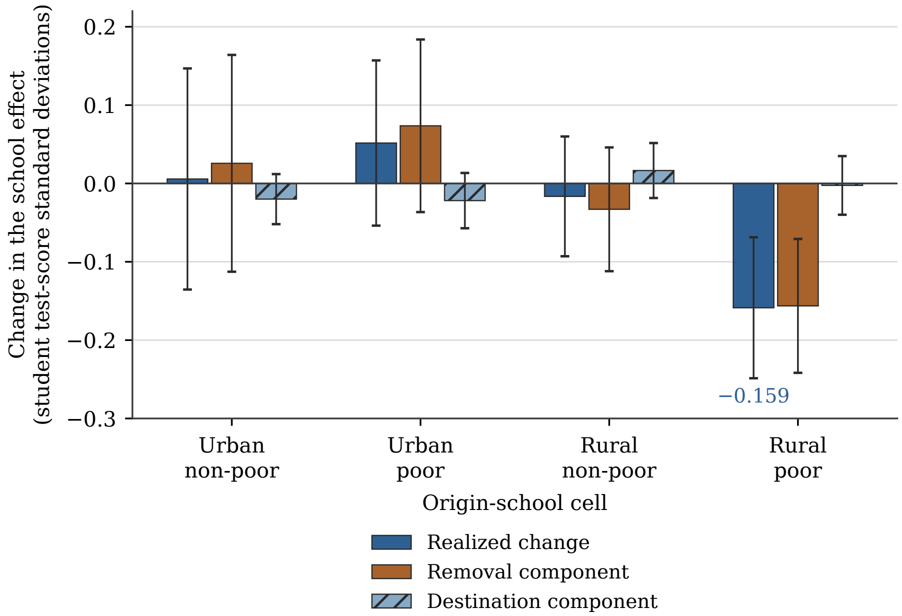
School size may connect the enrollment rule to the rural school-effect gap. In Appendix Table B4, enrollment is negatively associated with school effects when size, tested-grade span, and catchment breadth enter jointly. Controlling directly for school size reduces the rural coefficient from +0.148 to +0.099. The gradient-based attribution in Section 6.3.2 assigns 31 to 42 percent of the contrast to these characteristics. Neither approach explains the rural-poor interaction. School size thus accounts for part of the rural difference: where smaller schools are more effective, closing them can remove school effects above the local benchmark, although the estimated size gradients leave the concentration among rural-poor origins unexplained.

6.3.5 Destination School Effects

The local menu comparison describes the alternatives that remain. I now examine whether displaced children attended schools with estimated effects above that benchmark.

For children forcibly displaced before Hurricane María, the sample contains 34,772 moves from 325 origin schools. The estimated change in school effects, measured as destination minus origin, is +0.006 with a standard error of 0.072 for urban non-poor children, +0.052 with a standard error of 0.054 for urban

Figure 4. School-Effect Changes after Displacement



Notes: The figure uses 34,772 moves by children displaced from 325 origin schools in the pre-María period. Each observation is a move from a closed school to a destination school. The outcome is $\Delta\hat{\phi}$, the estimated school effect of the destination minus that of the closed school, in student test-score standard deviations normalized within subject, grade and year. I define the cells using the origin school. Rurality follows the school's urban/rural tag, and poverty divides schools at the median poverty rate. These groups therefore differ from the household groups in Table 9. Dark bars show the realized change. The middle bars show the removal component, which evaluates the change at the enrollment-weighted mean of the surviving local menu. Hatched bars show the destination component, which depends on the surviving school the family chose. The two components sum to the realized change by construction. Realized changes are +0.006, +0.052, −0.017 and −0.159; the destination component never exceeds 0.022 in absolute value in any cell. Whiskers show 95 percent intervals using standard errors clustered by origin school. Appendix Table B6 tabulates the cell means and standard errors plotted here.

poor children, −0.017 with a standard error of 0.039 for rural non-poor children, and −0.159 with a standard error of 0.046 for rural-poor children. Inference is clustered by origin school. Children from high-poverty rural schools therefore moved to destinations with estimated effects 0.16 standard deviations below those of their origins, while the mean changes for the other three groups are statistically indistinguishable from zero. Figure 4 shows the estimates, which are also reported in Appendix Table B6. The change in effects across a forced move has reliability of 0.80.

For the decomposition, I use 34,118 children from 317 origin schools and include close-year fixed effects, obtaining an additive rural coefficient of −0.006 with a standard error of 0.078 and an additive poverty coefficient of +0.051 with a standard error of 0.084. The rural-by-poverty interaction is −0.204 with a standard error of 0.103; school observables account for 7 percent of this interaction and the residual for 93 percent. The largest decline is therefore among rural-poor origins, whereas the additive rural coefficient of +0.148 in the menu comparison averages over this heterogeneity, combining a near-zero rural non-poor contrast with a larger rural-poor contrast.

The interaction is imprecisely estimated: its coefficient-to-standard-error ratio is 1.98. Without assumptions about the true effect and its sampling distribution, this observed ratio cannot by itself identify power, the expected exaggeration conditional on statistical significance, or a bias-corrected true effect (Gelman and Carlin, 2014). I therefore report the decomposition and the rural-poor mean of -0.159 with their sampling uncertainty, without adjusting them using such an illustrative calculation.

Equation (6) separates the change from the origin to its local benchmark from the destination component. The latter is no larger than 0.022 in absolute value in any group. For rural-poor children, the removal component is -0.156 with a standard error of 0.044, and the destination component is -0.003 with a standard error of 0.019. Most of the rural-poor average thus reflects replacing the origin with its local benchmark, although the decomposition describes the change in exposure without identifying what determined the destination choice. The child-level removal component of -0.156 and the school-level rural-poor mean of $+0.137$ have opposite sign conventions but point in the same direction. They use different samples and weights, so their magnitudes need not match.

The urban destination components are also small and statistically indistinguishable from zero, but small mean differences still allow sorting and cannot establish whether families could have reached better schools. Section 6.3.8 uses the choice estimates and the geography of surviving schools to examine substitution and access costs. These comparisons do not identify preferences for the separately estimated school-effect component.

6.3.6 Effects on Receiving Schools

To examine whether closure also affects pupils already attending the receiving schools, I use pupil–subject–year residuals from the production model, which retain within-school time variation. Regressing these residuals on absorption intensity, with school and year fixed effects and school-clustered inference, gives a decline of 0.063 standard deviations per unit, with a standard error of 0.019. Restricting the sample to incumbent pupils gives a larger coefficient of -0.074. This estimate suggests that adjustment extends to pupils who did not move. Appendix B reports the specification.

Since receiving schools differ in location and in nearby closures, absorption is not randomly assigned, although the regression compares a school with itself over time and accounts for persistent school differences and common year effects. The larger incumbent coefficient cannot arise solely from arriving pupils' transition-related score changes, but the design cannot separate crowding, peer changes, and other time-varying conditions. The association is consistent with receiving-school effects in the closure literature

(Brummet, 2014; Steinberg and MacDonald, 2019), and with evidence that teacher reshuffling can disrupt achievement beyond changes in measured teacher quality (Ronfeldt et al., 2013).

6.3.7 Teacher Reallocation

The preceding comparisons concern ϕ; comparing the school-average teacher component at closed schools with that of their surviving menus gives a gap close to zero in each group. The rural coefficient is -0.003 with a standard error of 0.027 under enrollment weights and $+0.004$ under distance weights. The poverty coefficient is -0.030 with a standard error of 0.015. The estimates provide little evidence that closed schools had higher school-average teacher value-added than their local alternatives, consistent with limited sorting of teacher quality across schools by pupil disadvantage (Mansfield, 2015). Since I compare preclosure school-average teacher value-added, these estimates exclude subsequent reassignment, whereas using the destination's postclosure roster would also include teachers who arrived from a closed school.

Appendix Table B7 shows teachers' subsequent assignments. Of the 2,145 teachers of tested subjects displaced by closure, 999, or 47 percent, left the tested panel, although leaving the panel includes reassignment to an untested grade or an administrative position. Teachers with higher value-added were less likely to leave, with an increase of one standard deviation in true value-added associated with a 0.049 lower exit probability and a standard error of 0.023. Exit was concentrated among senior teachers. Appendix B decomposes these patterns.

For teachers who remained and were reassigned, I compare colleagues displaced from the same school using origin-school fixed effects. The coefficient relating the destination school's preclosure mean teacher value-added to the teacher's own value-added is $+0.036$, with a standard error of 0.029, and is statistically indistinguishable from zero along with the other reported destination coefficients. The bound on the match standard deviation, $[0, 0.10]$, measures teacher–school match dispersion without determining whether another allocation could offset the change in school exposure. That question requires a counterfactual with feasible teacher assignments. Section 6.3.8 instead relates observed destinations to the school-choice estimates.

6.3.8 School Choice and Educational Exposure

Section 6.3.5 shows little average improvement in destination school effects over the local benchmark, including among rural-poor origins. The demand estimates help describe the cost of reaching other schools, although they cannot establish which destinations were feasible for an individual family.

One within-market standard deviation of mean utility is 0.116 utils. Relative to a zero-distance alternative, its distance equivalent is $\kappa_\tau = [0.1158/(-\lambda_\tau)]^{1/0.8}$. This gives 3.46 kilometers for urban non-poor families and 1.02 kilometers for rural-poor families, with 2.65 and 1.14 for the urban-poor and rural-non-poor groups. These values express the estimated utility–distance tradeoff, without determining how far a family could travel or which schools it could attend. Using the same reference, a move of 3.4 kilometers, the median displaced distance, costs urban non-poor families 0.99 standard deviations of mean utility and rural-poor families 2.65. The same distance therefore represents a larger utility cost for rural-poor families under the estimated preferences.

Measured teacher quality, which enters demand, explains about 4.3 percent of the within-market variation in mean utility. Other observed attributes and the unobserved demand component account for roughly 96 percent. Although the auxiliary quality projection is also small (Section 6.2.5), these findings alone do not explain the small average destination component in Section 6.3. The model estimates greater distance costs for rural households without identifying their willingness to travel for the school-effect component, and determining whether they could have reached better destinations requires additional information about each family's available alternatives.

These comparisons also do not establish that rural-poor families value a good school less, since a given utility difference corresponds to a shorter distance when travel costs are higher. Demand includes q, the school-average teacher value-added, while preferences for the school-effect component remain unidentified.

6.4 Distributional Effects

I now compare how changes in educational exposure and access are distributed across groups. I measure changes in estimated school-effect exposure, net of teacher composition, by combining observed destinations with the production estimates, and losses of access to the school network by comparing the preclosure and postclosure menus using the choice model in Section 4.4. These measures refer to different attributes, units, and populations, so I report them separately.

For the access comparisons I classify households by residential block-group rurality and administrative family poverty, while the school-effect comparisons classify origin schools by rurality and school poverty. The baseline access sample contains 992 closure markets, 122,350 agent records, and 334 schools, with inference clustered on 76 municipalities. The exhibits state which classification and sample each comparison uses.

6.4.1 Educational Exposure and Access

Table 8 shows educational exposure and access by rurality and poverty. Panel A first reports observed travel and destination choices. Compared with similar non-displaced pupils, children from rural schools moved about 3.6 to 3.9 additional road kilometers, while children from urban schools moved 1.3 to 1.6 additional kilometers. Among rural-poor displaced children, 42.4 percent enrolled at the nearest surviving school, compared with 35.7 percent among urban non-poor children. Nearest-school attendance in the rural-poor group is 3.7 percentage points higher than an additive specification of rurality and poverty would predict. The final rows show each group's population share and share of access losses. Rural-poor families account for 46.3 percent of the displaced-equivalent population and 76.3 percent of the model-implied access loss. Urban non-poor families account for 6.9 percent of this population and 1.5 percent of the loss.

Panel B compares school effects: 24.6 percent of urban-poor closures removed a school whose estimated effect exceeded the enrollment-weighted mean of the surviving alternatives, compared with 61.0 percent of rural-poor closures. The other two shares are 33.8 and 53.9 percent. The mean menu gap, which is positive when the origin's estimated effect exceeds its local benchmark, changes sign between urban-poor and rural-poor markets, from -0.097 in the former to $+0.137$ in the latter. The change in school effects from origin to destination is statistically indistinguishable from zero in three groups. Among rural-poor origins, it is -0.159 (0.046). The destination's estimated effect differs little from the local benchmark: the absolute difference never exceeds 0.022 in any group, and is -0.003 (0.019) for the rural poor. Taken together, the rural-poor comparisons show longer reassignment distances, more frequent nearest-school attendance, and lower school-effect exposure, with the small average destination component indicating little improvement over the local benchmark even though sorting across schools can still occur.

6.4.2 Distribution of Access Losses

To calculate the access loss, I remove the closed schools and hold fixed surviving schools' mean utilities, distances, household weights, the nest parameter, and all six preference parameters. The fall in expected maximum utility is the discrete-choice welfare measure of Small and Rosen (1981), applied to the nested logit in Section 4.2 and defined in Section 4.4. Table 9 reports the distribution across household types in the displaced-equivalent population. I express losses in utility units because a kilometer equivalent uses a different numeraire for each type, mechanically reducing the reported loss for households with greater distance aversion.

Rural-poor families account for 46.3 percent of displaced-equivalent mass but bear 76.3 percent of the

Table 8. School-Effect Changes and Access Losses

	Urban non-poor	Urban poor	Rural non-poor	Rural poor
Panel A. School access				
Additional travel over comparable non-displaced pupils (km)	1.60	1.33	3.90	3.63
Share enrolling at the nearest surviving school	0.357	0.384	0.360	0.424
Share of the displaced-equivalent population (%)	6.9	38.2	8.6	46.3
Share of the model-implied access loss (%)	1.5	12.9	9.3	**76.3**
Panel B. School-effect changes				
Share of closures removing a school above its local benchmark	0.338	0.246	0.539	0.610
Menu gap at the closed school, gap_{ϕ}	−0.059	−0.097	−0.004	+0.137
Realized change for displaced children, $\Delta\phi$	+0.006	+0.052	−0.017	−**0.159**
	(0.072)	(0.054)	(0.039)	(0.046)
of which destination component	−0.020	−0.022	+0.017	−0.003
	(0.016)	(0.018)	(0.018)	(0.019)

Notes: Both school-effect quantities are measured in student test-score standard deviations, normalized within subject, grade, and year. A positive menu gap means the estimated effect at the closed school exceeds the enrollment-weighted mean of its surviving local alternatives. A negative realized change means the destination has a lower estimated effect than the origin. The group definitions differ across rows, and I do not difference the rows against one another. Mass and loss shares use the demand model's household types: rurality is the block-group density classification, and poverty is the administrative family-poverty flag. The travel and rank-one rows use the same poverty flag. For rurality, the travel row uses the prior school's urban/rural tag, while the rank-one row uses kernel-weighted catchment exposure. In Panel B, rurality is the origin school's urban/rural tag and poverty is defined by a median split of origin schools. *Panel A.* Additional travel is the implied reassignment distance above baseline, measured in road kilometers between schools relative to the movement of comparable non-displaced pupils. I tabulate the rank-one share directly from displaced re-enrollers' observed destinations and report it without a standard error. Mass and loss shares use the displaced-equivalent population in Table 9: 992 closure markets, 122,350 agent records, and 334 pre-María closures. The loss shares are model-implied and conditional on the calibrations. If I set every household's distance cost to the urban non-poor value, rural families still lose 2.1 times as much as urban non-poor families because their next-best alternatives are farther away. This ratio uses the distance-only welfare evaluation: the three distance parameters in Table 5, comprising the base coefficient and its rural and poverty increments. It excludes the type-specific outside-option and quality shifters. *Panel B.* The first two rows describe 317 closed schools relative to their surviving menus (Table 7). The shares concern positive estimated gaps; they are not lower bounds on the shares of positive true gaps. The last two rows describe 34,772 forced moves across 325 origin schools, with standard errors clustered by origin school. The destination-component row measures the part of the realized change attributable to the surviving school chosen by the family. Its absolute value does not exceed 0.022 in any cell.

model-implied access welfare loss. The standard error of the loss share is 2.8 percentage points, and the 95 percent interval is 70.7 to 81.8. Together, rural families in both poverty groups bear 85.6 percent of the loss (SE 3.0). The mean loss per rural-poor family is 7.54 times that of an urban non-poor family (SE 1.58), with losses of 0.46 and 0.06 standard deviations when measured relative to the within-market dispersion of mean utility. The ordering of the four groups is unchanged in all 2,000 municipality block-bootstrap resamples.

The reporting population matters. In Population B above, each household is weighted by its model-implied probability of attending a closing school, whereas Population A uses population weights for all households in closure markets. Under the latter weights, rural-poor families' loss share is 59.5 percent (SE 4.6) and their mean-loss ratio is 4.35 (SE 0.89). Since the estimates describe different populations, they are not endpoints of a range.

Table 9. Access Losses by Household Type

	Urban non-poor	Urban poor	Rural non-poor	Rural poor
Panel A. The displaced-equivalent population				
Share of the population (%)	6.9	38.2	8.6	46.3
Mean welfare change, utils	−0.00714	−0.01104	−0.03540	−0.05378
in within-market SD of mean utility	−0.06	−0.10	−0.31	−0.46
Share of the aggregate loss (%)	1.5	12.9	9.3	**76.3**
Panel B. Aggregate statistics with standard errors				
	Point	Bootstrap	Parameter	Combined
Rural-poor loss share	0.763	0.0194	0.0208	0.0284
Rural loss share	0.856	0.0227	0.0199	0.0302
Mean-loss ratio, rural-poor to urban non-poor	7.536	1.049	1.176	1.576

Notes: Each observation is an agent–market cell. The sample contains 992 closure markets out of 3,120, 122,350 agent rows out of 285,800, 334 schools that closed before Hurricane María, and 76 municipality clusters. For the displaced-equivalent population, I weight each agent by the resident school-age population of its market and by its model-implied probability of attending a closing school. This is the main reporting population. It describes an ex ante calculation weighted by attendance probabilities, rather than a count of observed displaced pupils. I define household types using the demand model's block-group density classification and administrative family-poverty flag. These types differ from the origin-school groups in Table B6. I evaluate welfare using the two-level nested logsum $W_i = \ln[1+\exp((1-\rho)\,\mathrm{IV}_i)]$ before and after removing the closure set. The mean utilities of surviving schools, distances, agent weights, and parameters remain fixed. These are model-implied quantities conditional on the calibrated nesting parameter. The reported welfare change is $\Delta W = -L$, where $L \geq 0$ is defined in Appendix D; all six preference shifters enter that calculation. I report standard errors for the aggregate statistics in Panel B, which summarize the distribution of access losses. Panel A contains point estimates for each cell. For the bootstrap column, I resample the 76 municipalities while holding the parameters fixed. For the parameter column, I use a numerical delta method over the six preference-shifter parameters and the nesting parameter while holding the sample fixed. This component uses a worst-case covariance bound across the parameter contributions. The combined column adds the bootstrap and parameter components in quadrature. This combination omits covariance between the two components, so it is not a worst-case bound on total uncertainty. The reported approximate ninety-five percent intervals are $[0.707, 0.818]$ and $[0.797, 0.915]$ for the two loss shares and $[4.45, 10.62]$ for the mean-loss ratio. The ordering of the four cells never reverses across 2,000 resamples. Parameter uncertainty is concentrated in the rural distance increment. Applying the same calculation to all families living in closure markets gives a rural-poor loss share of 59.5 percent (SE 4.6, interval $[0.506, 0.685]$) and a mean-loss ratio of 4.353 (SE 0.89, interval $[2.61, 6.10]$). These estimates change only the population weights. They therefore describe a different population and should not be combined with the displaced-equivalent estimates to form a range.

The estimates also depend on the nesting parameter, which I calibrate to departure from the local public-school choice set by combining estimated public-system exit with an approximate cross-municipality switching adjustment. Welfare levels are conditional on this parameter, and loss shares can depend on it as well. Appendix D reports a sensitivity exercise for the earlier distance-only specification, which does not re-estimate the six-parameter baseline. The welfare levels also depend on the outside-option normalization discussed in Section 6.2.6. Of the demand parameters, the rural distance increment, estimated from differences between catchments, contributes most to the reported standard errors through its uncertainty.

6.4.3 Geography and Preference Heterogeneity

The rural difference in access losses may arise from both geography and estimated preference differences. To examine geography under common preferences, I recompute welfare for Population A using the distance-only specification in Table 5, giving every household the urban non-poor distance coefficient and omitting the type-specific inside-option and quality shifters. Rural families still lose 2.1 times as much as urban non-poor families because their alternatives are farther away. The ratio in kilometers is 2.04. These ratios are close in this particular calculation; nonlinear distance costs do not generally make utility and kilometer ratios equal.

The rural gradient therefore remains when I remove estimated preference heterogeneity from the calculation, conditional on the specified choice sets and common distance-cost function. Travel speed provides little support for a separate topographic explanation. Across 279,759 school-to-school routes, average minutes per kilometer vary little by geography.

6.4.4 Exposure and Outside Alternatives

The poverty increment in the distance coefficient is -0.0102 (SE 0.0020), about a sixth of the rural increment, -0.0612 (SE 0.0071). The larger poverty difference concerns exposure to the public-school network. Rural-poor families make up 32.6 percent of the population in closure markets, compared with 8.2 percent for rural non-poor families, or about four to one. Poor families therefore form the larger rural group exposed to longer reassignment distances, and they rely more heavily on public schools: 8.7 percent of income-poor children are outside the public system, compared with 42.6 percent of the non-poor, while non-enrollment is the same in both groups at 1.8 percent. Private-school enrollment accounts for the whole difference. The inside-option shifter π_{pov} captures this participation difference and is large relative to the poverty increment in distance costs. Although these patterns show greater reliance on the shrinking public network, they do not reveal whether private alternatives are available or affordable for a particular family.

The school-effect comparisons also place the largest losses among rural-poor origins. In Section 6.3, the additive rural and poverty coefficients are small and statistically insignificant, and school observables explain little of the interaction contrast.

How income, geography, and public-school participation are related therefore matters for the distribution of losses: poor families are more numerous in the rural population exposed to closure and are less likely to attend private schools, so a large poverty difference in distance aversion is not necessary for them to bear more of the loss.

6.4.5 Scope and Limitations

I do not convert the achievement changes into dollars, since monetizing the school-effect loss in Table 8 would require a valuation of test-score standard deviations that I do not estimate and adding this value to the access loss would count twice the part of school quality that families already consider when choosing a school. Appendix D values additional travel under transport-appraisal assumptions, pricing marginal kilometers for a family that already drives rather than the full logsum loss. It also reports the change in the share of rural-poor displaced-equivalent children choosing a school within walking range, since families without motorized transport may have difficulty reaching more distant schools. I keep the household and fiscal accounts separate for the reasons discussed in the next section. Since the calculation holds residence fixed, it excludes the longer-run residential responses documented in other settings (Di Cataldo and Romani, 2024; Gibbons et al., 2024).

Determining how much of the loss could have been avoided would require comparing alternative closure rules using information available to the Department before its decisions, a counterfactual I do not estimate. I turn next to the potential fiscal savings from the observed consolidation.

7 Fiscal Implications

The facilities and staffing savings from closing schools depend on what happens to the closed buildings and the positions they release. Using the enrollment and utilization patterns in Section 3, I calculate estimates or bounds for these components and report the fiscal account separately from the household losses in Section 6.4, since the two cannot be combined directly for the reasons in Section 7.3.

7.1 Fiscal Accounting

Puerto Rico does not publish spending by school, although Law 85-2018 required annual reports on direct-to-student spending, a requirement that went unimplemented for more than five years. I therefore calculate potential savings component by component. For a set of closures C in year t, the annual budget account relative to no further closures is

$$S_t(C) = \underbrace{F_t(C)}_{\text{plant, net of mothballing}} + \underbrace{P_t(C)}_{\text{positions freed}} - \underbrace{K_t(C)}_{\text{transition costs}} - \underbrace{\Delta T_t(C)}_{\text{transport}} + \underbrace{G_t(C)}_{\text{enrollment leakage}}, \tag{13}$$

where the terms are dollar flows in year t. One-time expenditures enter $K_t(C)$ in the year they occur. I quantify only the first two components.

Table 10. School Size and Staffing

	Pupil–teacher ratio	Enrollment
Panel A. Observed enrollment and staffing		
Smallest enrollment quintile	13.4	75
Middle enrollment quintile	19.6	195
Largest enrollment quintile	26.8	411
Median closed school (7 teachers)	15.8	108
System norm	19.8	
Panel B. Potential staffing reductions		
Positions freed at the median closure	1.32	
Share of closures freeing a positive number of positions	0.748	

Notes: Panel A. I summarize enrollment and staffing using administrative records. Closed schools must have a teacher roster to enter the staffing sample, so the sample differs from Table 1. I form quintiles by school enrollment and report mean enrollment within each quintile. The system norm is the territory-wide pupil–teacher ratio. *Panel B.* I calculate positions freed by subtracting the teachers needed to accommodate a closed school's pupils at the system norm from that school's teacher headcount. This is an accounting calculation, not a measure of observed layoffs or an estimate from the school-choice model. It gives an upper bound on released positions if receiving schools absorb the new pupils at the norm without additional hiring. Under this benchmark, approximately three quarters of closures release positions. Payroll savings also require those positions to remain unfilled, and the data do not identify that rate.

Much of the spending follows displaced pupils to receiving schools, and most teachers are reassigned. Of the 10,819 exposed staff-years, 88.0 percent remained employed by the Department in the following year. The available evidence does not establish the sign of the change in public transport spending, so I leave that component unestimated. I also report enrollment leakage separately: counting expenditure reductions from public-school exit as a benefit would assign a fiscal gain to children's departure from the system.

For facilities, I calculate avoided costs from a spending relationship estimated on 127 surviving schools, which predicts expenditure of $35,498 at the median closing school and leaves $29,038 per closed school-year after subtracting the cost of maintaining a mothballed site. If only the fitted intercept is avoidable and the per-student component follows the pupils, the saving falls to $12,197, giving a range of $12,197 to $29,038 per closed school-year. Appendix D describes the calculations. The official operating-cost figures in Table 11 are higher because they measure gross expenditure at open schools, whereas my range measures the expenditure avoided after accounting for mothballing.

Table 10 presents observed staffing patterns and the benchmark used to calculate positions that could be released. Larger schools have higher pupil–teacher ratios, and I calculate the potential reduction by subtracting the number of teachers needed to serve displaced pupils at the system ratio of 19.8 from the closed school's teacher headcount. The median reduction is 1.32 positions, with a positive reduction in 74.8 percent of closures. These are potential reductions under the receiving-school benchmark, rather than layoffs, and payroll savings occur only if the released positions remain unfilled.

The separate staffing-cost calculation gives a bound of $0 to $19.2 million a year using the staff-flow and wage data described in Appendix D, with the lower endpoint reflecting the unidentified rate at which positions are refilled. I do not obtain this bound by multiplying the median position reduction in Table 10 by a wage.

When a teacher retains her pay after reassignment, moving her to another school leaves the wage bill unchanged. In the payroll data, teacher fixed effects explain 83.5 percent of pay variation, compared with 49.5 percent for school fixed effects, while pay and measured effectiveness have a correlation of 0.035. Appendix D discusses the limitations of the payroll file.

7.2 Potential Savings

Applying these calculations to the 610 schools identified as closed gives annual facilities savings of $7.4 to $17.7 million and staffing savings of $0 to $19.2 million. The total is $7.4 to $36.9 million a year, compared with a Department budget of $2,628,835,000, or between 0.3 and 1.4 percent. The facilities cost basis and the rate of refilling positions determine this partial-identification bound, whose endpoints and assumptions appear in Table 11. It is not a confidence interval, and I do not treat its midpoint as a preferred estimate.

Two other quantities help interpret the fiscal evidence but are excluded from this account. The roughly $100,000 of stranded movable property per closed school in audited regional inventories is a registered asset value, rather than a payment made when the school closes. The government's three announced savings figures—a point estimate, a floor, and an approximate target—cover different cost categories without stating their horizons. Appendix Table D1 reproduces their wording and sources and shows that a utilities-saving calculation almost exactly reconciles the 2017 figure.

7.3 Comparability and Limitations

An illustrative transport calculation puts the potential savings in context. At national average expenditure of $1,197 per transported student-year, transporting the 24,451 displaced-equivalent students would cost about $29.3 million annually. For the same 334 closures, potential annual savings are $8.3 to $20.6 million, putting the implied transport expenditure between 1.4 and 3.5 times that range. Although this comparison uses no value of time, discount rate, or welfare weight, average expenditure per transported pupil may differ from the marginal cost of transporting newly displaced children.

The household and fiscal calculations also cover different samples and horizons: the monetary household calculation is a present value for one displaced cohort over its remaining exposure, covering 334 closures, whereas the fiscal calculation is an annual flow for a roster of 467, with only 197 schools in both samples.

Table 11. Fiscal Savings from School Closures

Component	Per closed school-year	Program, per year
Plant, net of mothballing	\$12,197–\$29,038	\$7.4–17.7 m
Teaching positions freed	\$0–\$31,600	\$0–19.2 m
Total		**\$7.4–36.9 m**
as a share of the education budget		0.3–1.4%
Operating cost of an open school, for comparison		
Secretary's 2017 statement, utilities only	\$43,017	
Oversight board, stated	\$47,000	
Oversight board, \$15–20 m over 255 schools	\$68,627	

Notes: The table reports annual plant and staffing savings alongside gross operating-cost benchmarks. The annual budget account is given by (13), with each component measured in year-t dollars and one-time costs recorded in the year incurred. I quantify only the plant and staffing components. Audited regional inventories record stranded movable property worth about \$100,000 per closed school. This is a registered asset value rather than a payment at closure, so I document it without entering it as a one-time cost. I leave additional public transport costs unestimated because the evidence does not establish their sign. I also exclude the leakage term: counting the per-pupil cost of children who leave the public system as a saving would treat their departure as a fiscal benefit. The reported range is a partial-identification interval rather than a confidence interval. The data do not identify a preferred point within it. Its width reflects two assumptions about plant costs: the fitted facilities-spending relationship evaluated at the median closing school and net of mothballing, or the fitted intercept alone. It also reflects uncertainty about the rate at which freed positions are refilled, which these data do not identify. I extrapolate the program figures to the 610 schools recorded as ever closed. To calculate the budget share, I use the Department's consolidated budget of \$2.63 billion in the fiscal year 2017–18 strategic plan. The positions term is identified only on the 2016–2018 cohorts, 467 of the 610 closures, and about 88 percent of exposed staff were still employed by the Department the following year. The three comparison figures measure gross operating costs at schools still in use. They do not subtract mothballing costs from avoided spending, so all three exceed this paper's plant estimate. Appendix D gives the construction of each component. I do not subtract the household costs in Table D3 from this fiscal account. The two calculations cover different closures, with 197 schools in common. The household account is a present value for one displaced cohort, whereas the fiscal account is an annual flow. Netting the two would also assume that the savings were returned to Puerto Rican households.

Since the household calculation also values additional travel rather than all consequences of closure, these estimates do not provide a directly comparable per-closure break-even calculation or a marginal value of public funds.

Combining the accounts would further require identifying who receives the fiscal savings and deciding how to weight those gains against local household losses, and I cannot establish that the savings return to the Puerto Rican households affected by closure.[6]

The fiscal effects of district mergers provide a different comparison. Duncombe and Yinger (2007) find that doubling enrollment reduces operating cost per pupil by 61.7 percent for a 300-pupil district, or 31.5 percent after adjustment costs. Those mergers combine whole rural districts and their administrations, whereas Puerto Rico already operates as a single district and closing an individual school does not combine district administrations. Studies of closures within districts are closer to this setting. Nguyen-Hoang (2019) finds

[6] Capps et al. (2010) find a related distinction for five urban hospital closures. Aggregate cost savings exceeded the patient-welfare loss, but the savings were distributed nationally while the access losses fell on the local community.

that Ohio closures reduced operating expenditure by roughly $280 per pupil, mostly through classroom instruction. In a report for Getting Down to Facts III, Pearman (2026) finds that California closures reduced expenditure and revenue by about $440 per pupil each, leaving district deficits unchanged.

Declining enrollment and fiscal pressure motivated consolidation, and these calculations describe the consequences of the closures that occurred. To determine whether consolidation was necessary or another set of closures would have been preferable, I would need counterfactual policies with feasible school and teacher assignments.

8 Conclusion

This paper studies how the consequences of school consolidation depend on the effectiveness and accessibility of the alternatives that remain. The structural model distinguishes school contributions from teacher composition and describes families' substitution among schools. I use its estimates to compare changes in educational exposure and access across groups. Both this paper and the previous study by Bobonis et al. (2022) report small average achievement estimates that are statistically indistinguishable from zero, but these estimates do not summarize the changes in school-effect exposure and access examined here. The school-effect comparisons concern one component of educational opportunity, rather than a treatment effect on children's own test scores, while the access measure also includes costs that test scores do not capture.

Enrollment and class size were the main observed predictors of closure. Although closed schools overall did not have systematically lower estimated school effects than surviving schools, closed high-poverty rural schools had higher effects than their enrollment-weighted surviving benchmark, on average, while closed urban schools had lower effects. The corresponding mean teacher-quality differences were small and statistically indistinguishable from zero, and children from high-poverty rural origins entered schools with lower estimated effects than their origins and little average improvement over the local benchmark. At receiving schools, achievement residuals covary with the intensity of displaced-pupil absorption, although this association does not isolate a causal crowding effect.

The choice estimates show greater distance sensitivity among rural households and greater reliance on public schools among poorer households. The outside option includes private enrollment, non-enrollment, and public enrollment outside the municipality. Rural-poor families bear most of the model-implied loss while accounting for less than half of the displaced-equivalent population, and this ordering holds in every municipality resample. Since the access comparisons classify households and the school-effect comparisons classify origin schools, the results do not measure two losses for the same group of children.

The access calculation measures the fall in expected maximum utility across all available alternatives, while the school-effect comparison uses an enrollment-weighted local benchmark, so neither takes the single best surviving school as its reference. Enrollment alone does not measure these differences: the distributional consequences of closure depend on where schools are and what alternatives remain nearby; the local comparisons matter where small schools are rural, alternatives are distant, and school contributions to achievement are substantial. The differences between urban and high-poverty rural origins also suggest that average results from dense districts may not describe settings with more distant alternatives.

The quantified fiscal components amount to \$7.4 to \$36.9 million a year, compared with a departmental budget of \$2.6 billion. In Section 7, I compare these amounts with an illustrative transport cost, although that comparison does not estimate the cost of compensating families or the net fiscal benefit of closure.

One possible counterfactual would hold the number of closures fixed and change which schools close, using only information available before the decision: school effects estimated from earlier classrooms and years, distances within each catchment, and the Department's enrollment counts. I do not estimate this counterfactual, and closing the same number of schools would not produce equal savings if their costs or the requirements at receiving schools differ. To hold a fiscal target fixed instead, I would need an explicit budget constraint and a rule for aggregating household gains and losses. Epple et al. (2018) study the relevant decision problem, allowing parents to re-sort after closure. Abdulkadiroğlu et al. (2026) provide a useful benchmark: reallocation to improve matches produces only modest gains when places at good schools are scarce. Chrzan and Pearman (2026) provide a simulation precedent with 5,040 closure scenarios. Applying such a counterfactual here would require pre-decision school-effect estimates, propagation of their sampling error into the ranking, feasible placements for displaced pupils, and a treatment of hurricane-driven closures, elements that are not available in the present analysis.

Another counterfactual would reassign teachers across surviving schools, changing the teacher contribution in the achievement equation. The small mean teacher-quality gaps between closed schools and their local alternatives do not establish that reassignment would have no effect, although whether it could offset a change in school exposure depends on feasible assignments, class sizes, and teacher availability. I do not estimate this counterfactual either. Bobba et al. (2021) examine related questions of teacher compensation and allocation in Peru. For the United States, Bates et al. (2025) show that a policy acting on only one side of the match can be ineffective or harmful when both teachers' and principals' preferences depart from assignments that maximize disadvantaged students' achievement.

Several limitations remain. Since the baseline fixes the nesting parameter at its estimate from an earlier joint demand estimation, welfare levels and incidence shares are conditional on this specification, a dependence that taking ratios does not generally eliminate. Rurality in the utility equation refers to a household's residential block group, while in the observed substitution moments it refers to rural exposure within a closure catchment. The corresponding within-municipality contrast is about two-fifths of the pooled one, indicating sensitivity rather than a formal bound on the structural coefficient. I measure the accessibility of substitutes by their distance, rather than the number within a fixed radius. Rurality and poverty also apply to different units in the access and school-effect analyses: groups with the same names contain different members.

The estimated effect of a closing school uses all its operating years, including years after the closure decision, and the event-time profile in Section 6.3 shows that its observed performance immediately before closure was above its estimated average. Although pupil fixed effects account for the child's persistent achievement, they do not remove classmates' contributions, so the school effect can include peer effects; the estimates also do not separate first-year disruption at the destination from the destination school's own effect. I do not separately estimate displaced pupils' score changes for the two-by-two rurality and poverty groups, for which the reduced-form gradient in Section 6.1 provides the closest comparison.

Finally, the largest wave, 262 schools in 2018, is outside the principal demand and access window, and the reported incidence covers 334 of the panel's 610 closures.

Declining enrollment created pressure to close schools, but the consequences for families depended on which schools closed and on the effectiveness and accessibility of the schools that remained.

References

Abdulkadiroğlu, A., Pathak, P. A., Schellenberg, J., and Walters, C. R. (2020). Do parents value school effectiveness? *American Economic Review*, 110(5):1502–1539.

Abdulkadiroğlu, A., Pathak, P. A., and Walters, C. R. (2026). Who gets what in education: Can school matching improve student achievement? NBER Working Paper 34936, National Bureau of Economic Research.

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Altonji, J. G. and Mansfield, R. K. (2018). Estimating group effects using averages of observables to control for sorting on unobservables: School and neighborhood effects. *American Economic Review*, 108(10):2902–2946.

Andrews, M. J., Gill, L., Schank, T., and Upward, R. (2008). High wage workers and low wage firms: Negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society, Series A*, 171(3):673–697.

Angrist, J. D., Hull, P. D., Pathak, P. A., and Walters, C. R. (2017). Leveraging lotteries for school value-added: Testing and estimation. *Quarterly Journal of Economics*, 132(2):871–919.

Bates, M. D., Dinerstein, M., Johnston, A. C., and Sorkin, I. (2025). Teacher labor market policy and the theory of the second best. *Quarterly Journal of Economics*, 140(2):1417–1469.

Bayer, P., Ferreira, F., and McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, 115(4):588–638.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Berry, S., Levinsohn, J., and Pakes, A. (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy*, 112(1):68–105.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, 25(2):242–262.

Beuermann, D. W., Jackson, C. K., Navarro-Sola, L., and Pardo, F. (2023). What is a good school, and can parents tell? Evidence on the multidimensionality of school output. *Review of Economic Studies*, 90(1):65–101.

Bifulco, R. and Schwegman, D. (2020). Who benefits from accountability-driven school closure? Evidence from New York City. *Journal of Policy Analysis and Management*, 39(1):96–130.

Bobba, M., Ederer, T., León-Ciliotta, G., Neilson, C., and Nieddu, M. G. (2021). Teacher compensation and structural inequality: Evidence from centralized teacher school choice in Perú. NBER Working Paper 29068, National Bureau of Economic Research. Revised most recently September 2026.

Bobonis, G. J., Sotomayor, O. J., and Wagner, J. (2022). The effects of school consolidation on students and teachers: Evidence from an underperforming system. Working paper, University of Toronto, December 2, 2022.

Brummet, Q. (2014). The effect of school closings on student achievement. *Journal of Public Economics*, 119:108–124.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3):414–427.

Capps, C., Dranove, D., and Lindrooth, R. C. (2010). Hospital closure and economic efficiency. *Journal of Health Economics*, 29(1):87–109.

Card, D., Heining, J., and Kline, P. (2013). Workplace heterogeneity and the rise of West German wage inequality. *Quarterly Journal of Economics*, 128(3):967–1015.

Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory*, 13(2):185–213.

Carlson, D. and Lavertu, S. (2016). Charter school closure and student achievement: Evidence from Ohio. *Journal of Urban Economics*, 95:31–48.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.

Chrzan, M. L. and Pearman, F. A. (2026). A sandbox for hard choices: Using simulation to explore school closure scenarios and their consequences. EdWorkingPaper 26-1440, Annenberg Institute at Brown University.

Conlon, C. and Gortmaker, J. (2020). Best practices for differentiated products demand estimation with PyBLP. *RAND Journal of Economics*, 51(4):1108–1161.

Conlon, C. and Gortmaker, J. (2025). Incorporating micro data into differentiated products demand estimation with PyBLP. *Journal of Econometrics*, page 105926. Article 105926; volume not confirmed at verification. NBER Working Paper 31605 (August 2023) is the working-paper version.

Di Cataldo, M. and Romani, G. (2024). Rational cuts? The local impact of closing undersized schools. *Regional Science and Urban Economics*, 109:104057.

Duncombe, W. D. and Yinger, J. (2007). Does school district consolidation cut costs? *Education Finance and Policy*, 2(4):341–375.

Engberg, J., Gill, B., Zamarro, G., and Zimmer, R. (2012). Closing schools in a shrinking district: Do student outcomes depend on which schools are closed? *Journal of Urban Economics*, 71(2):189–203.

Epple, D., Jha, A., and Sieg, H. (2018). The superintendent's dilemma: Managing school district capacity as parents vote with their feet. *Quantitative Economics*, 9(1):483–520.

Gelman, A. and Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, 9(6):641–651.

Gibbons, S., Heblich, S., and Pinchbeck, E. W. (2024). The spatial impacts of a massive rail disinvestment program: The Beeching Axe. *Journal of Urban Economics*, 143:103691.

Government of Puerto Rico (2014). Agenda para la recuperación económica 2014–2018. Economic recovery agenda, Estado Libre Asociado de Puerto Rico.

Government of Puerto Rico (2016). Commonwealth of Puerto Rico fiscal plan. Fiscal-plan submission, Commonwealth of Puerto Rico. Submitted October 14, 2016.

Hastings, J. S. and Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *Quarterly Journal of Economics*, 123(4):1373–1414.

Hendren, N. and Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. *Quarterly Journal of Economics*, 135(3):1209–1318.

Hill, A. J. and Jones, D. B. (2021). The effect of school closings on teacher labor market outcomes and teacher effectiveness. *Education Finance and Policy*, 16(4):584–608.

Iacus, S. M., King, G., and Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1):1–24.

Jackson, C. K. (2013). Match quality, worker productivity, and worker mobility: Direct evidence from teachers. *Review of Economics and Statistics*, 95(4):1096–1116.

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

Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2020). Heterogeneous beliefs and school choice mechanisms. *American Economic Review*, 110(5):1274–1315.

Kim, J. (2026). The long shadow of school closures: Impacts on students' educational and labor market outcomes. *American Economic Journal: Economic Policy*. Forthcoming.

Kline, P., Saggio, R., and Sølvsten, M. (2020). Leave-out estimation of variance components. *Econometrica*, 88(5):1859–1898.

Kramarz, F., Machin, S., and Ouazad, A. (2015). Using compulsory mobility to identify school quality and peer effects. *Oxford Bulletin of Economics and Statistics*, 77(4):566–587.

MacKinnon, J. G. and Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32(2):233–254.

Mansfield, R. K. (2015). Teacher quality and student inequality. *Journal of Labor Economics*, 33(3):751–788.

NCES (2021). Enrollment in public elementary and secondary schools, by region, state, and jurisdiction: Selected years, fall 1990 through fall 2030. National Center for Education Statistics, Digest of Education Statistics 2021, Table 203.20. Table prepared March 2022; historical enrollment from the Common Core of Data.

NCES (n.d.). Number of operating public schools and districts, student membership, teachers, and pupil/teacher ratio, by state or jurisdiction: School year 2016–17. National Center for Education Statistics, Common Core of Data, Table 2. 2016–17 provisional release, version 1a; accessed September 12, 2026.

Nelson, C. D., Prado Tuma, A., Marsh, T., Andrew, M., Anderson, D. M., Whitaker, A. A., Karoly, L. A., Murphy, R. F., Nanda, N., Ryan, J. L., Smith, T. D., and Chandra, A. (2020). The education sector in Puerto Rico after Hurricane Maria: Predisaster conditions, hurricane damage, and themes for recovery. Research report RR-2858-DHS, Homeland Security Operational Analysis Center operated by the RAND Corporation.

Nguyen-Hoang, P. (2019). Does the cost-cutting strategy of closing public schools provide financial benefits? Evidence from Ohio. *Public Budgeting and Finance*, 39(3):3–21.

OECD (2018). *Responsive School Systems: Connecting Facilities, Sectors and Programmes for Student Success*. OECD Reviews of School Resources. OECD Publishing, Paris.

Pathak, P. A. and Shi, P. (2021). How well do structural demand models work? Counterfactual predictions in school choice. *Journal of Econometrics*, 222(1):161–195.

Pearman, F. A. (2026). The fiscal consequences of school closures in California: Evidence from a statewide synthetic difference-in-differences design. Report, Getting Down to Facts III, Stanford Graduate School of Education. Report, not peer reviewed.

PRDE (2015a). Notificación a los padres, madres o encargados de los estudiantes cuyas escuelas fueron seleccionadas para consolidación. Parent notification, Puerto Rico Department of Education. May 21, 2015. Los Caños school, code 10108.

PRDE (2015b). Política pública sobre las normas y los procedimientos para el cierre, la consolidación o la reestructuración de escuelas. Carta Circular 21-2015-2016, Puerto Rico Department of Education. November 17, 2015.

PRDE (2017a). Política pública para la reubicación, reasignación, traslado y reclutamiento del personal docente de las escuelas, de los institutos tecnológicos, la escuela de troquelería y herramentaje y la PR Aviation Maintenance Institute en el departamento de educación. Carta Circular 31-2016-2017, Puerto Rico Department of Education. January 27, 2017.

PRDE (2017b). Procedimiento general para el rediseño de escuelas del sistema público de Puerto Rico. Carta Circular 33-2016-2017, Puerto Rico Department of Education. March 10, 2017.

PRDE (2017c). Proceso de matrícula en línea: Entrada y actualización de datos al nuevo sistema. Memorandum and attached director manual, Puerto Rico Department of Education. August 3, 2017; manual for school year 2017–18.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458.

Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.

Ronfeldt, M., Loeb, S., and Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, 50(1):4–36.

Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., and Webb, M. D. (2019). Fast and wild: Bootstrap inference in Stata using boottest. *Stata Journal*, 19(1):4–60.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, 125(1):175–214.

Small, K. A. and Rosen, H. S. (1981). Applied welfare economics with discrete choice models. *Econometrica*, 49(1):105–130.

Steinberg, M. P. and MacDonald, J. M. (2019). The effects of closing urban schools on students' academic and behavioral outcomes: Evidence from Philadelphia. *Economics of Education Review*, 69:25–60.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, 2nd edition. ISBN 9780521766555 (hardback); 9780521747387 (paperback).

Walters, C. R. (2018). The demand for effective charter schools. *Journal of Political Economy*, 126(6):2179–2223.

A Data Construction and Samples

I describe the construction of the data used in Sections 3 and 6.1, starting with the academic-year and grade codes and closure dates, and then the record linkage, residence and geographic measures, and estimation samples. Table 2 shows how much of the data each sample covers. The appendix also reports the dynamic exit estimates behind the range in Section 6.1, together with the comparison of origin-school poverty and score changes among displaced pupils.

A.1 Calendar and the closure flag

I label academic years by the calendar year in which they end, and record a school's closure in its final operating year, so children enrolled there in that year are first observed elsewhere in the following year. For a closure in year c, the surviving menu consists of schools operating in $c+1$.[7] Closures affect the entire school in these records: in 100 percent of school-years, a school either remains open in all its grades or closes in all of them.

Using each school's last operating year, I count 8, 70, 59, 41, 156 and 262 schools for 2013 through 2018 respectively, a total of 596, two thirds of them in the last two waves. Official and press sources use different conventions for the 2017 and 2018 waves. For the last wave, they report 262, 281, 283 or 259 schools, which is why I state the counting convention when reporting the size of a closure wave. The administrative file records 610 distinct schools as ever closed within the grades-3–8 frame, while the wave-by-wave counts sum to 596; the file does not assign the remaining fourteen schools to a wave. The closure counts differ across samples: roughly 688 schools in system-wide counts for 2010 through 2019, 610 inside the grades-3–8 panel, 334 in the pre-María access frame, and 317 with both the access and the achievement channel measured. An administrative school code does not always identify a physical building, so before including a school in a closure comparison I distinguish permanent closure from recoding, merger, grade reconfiguration, and temporary interruption. The menu construction in Section 4.3 covers all 596 schools assigned to closure waves; every menu is nonempty, and none is widened. I exclude 10 schools whose operating-status records suggest recoding and 24 schools with fewer than two scored years or fewer than 30 scored records. The resulting pre-María primary sample contains 317 schools.

A.2 Grade coding

The recorded grade codes require an adjustment. Coverage of the normalized mathematics score is exactly 0.000 at grade codes 3 and 4 and approximately 0.92 at codes 5 through 10, while testing covers real grades 3 through 8, so code 5 corresponds to real grade 3 and I subtract two from the recorded codes to recover

[7] Bobonis et al. (2022, p. 7, n. 14) date closure by the calendar year in which the first zero-enrollment academic year ends. A school last operating in 2016–17 is therefore dated 2017 here and 2018 in their convention. Their primary-school sample includes closures through academic year 2017–18 and excludes the 2018–19 round (p. 8). Their counts are not interchangeable with the counts in this paper.

actual grades. The analysis range “real grades 3 through 8” is therefore codes 5 through 10, and the demand frame’s real grades 1 through 8 are codes 3 through 10. Applying a real-grade restriction to an uncorrected code retains real grades 1 through 6 and drops 7 and 8, which is the reverse of the intended sample.

The enrollment grade in the pupil file and the test grade in the teacher-link file disagree for about 0.3 percent of children, who are tested off their enrollment grade. I use the enrollment grade for sequencing and market definition and the test grade for the within-subject–grade–year normalization of the outcome, and I merge scores to teacher links on pupil and year rather than on grade, so that off-grade testers are not dropped.

A.3 Merge keys and rates

I begin with the raw teacher-link file, which contains 3.08 million pupil-year records for 2010 through 2023 across 1,472 school codes, and drop the years 2020 through 2023. Since the row identifier identifies a record, I use the pupil identifier when counting pupils. Using the pupil identifier, I link 98.1 percent of the teacher-link file to the scored pupil file. Using the school code, I link 87 percent to the school and closure file, or 1,281 of 1,472 codes. The unmatched records are high schools and schools outside the study frame, and coverage is nearly complete within grades 3 through 8.

Before estimating teacher effects, I reshape the subject columns in the teacher-link file to one record per pupil, subject, and year. Each pupil–year–subject cell maps to exactly one teacher, without requiring a rule for allocating pupils across teachers. One teacher teaches all three subjects in about 1.5 percent of assignments. This share reaches its maximum of 5.4 percent in grade 3, and the median class size is 26.

A.4 Connectivity and pruning of the production sample

I first restrict the scored sample of 1,947,300 records to the largest connected component of the teacher-mover graph, which contains 1,914,465 records across 1,179 schools. I then drop 587 pupils with a single observation and 89 high-leverage observations, leaving 1,912,673. Fifty-one schools fall outside the connected component; they are disproportionately urban, with a rural share of 0.31 against 0.56 in the estimation sample. Connectivity of the teacher-by-subject mover graph is 78 percent in mathematics, 79 percent in English and 62 percent in Spanish, the Spanish figure being the weakest. For about fifteen schools, the school effects are not separately identified from the pupil effects and a small ridge penalty shrinks them to approximately zero. I exclude these effects from all density plots and distributional figures; none belongs to a school that closed. Across the full enrollment frame, without the score requirement, 34 percent of teachers are observed at more than one school; the score requirement therefore restricts the production sample more than teacher mobility.

A.5 Residence, population, and the block-group allocation

The data record residence at the ZIP level. A Puerto Rican ZIP contains on the order of a hundred block groups, so I cannot observe any child's residential block group. The demand model therefore treats a market's resident school-age population, aggregated from Puerto Rico Community Survey block groups, as its market size, and integrates household choice over the block groups in the market using population weights. The agent weights sum to one within a market. When aggregating quantities across markets, I multiply these weights by market size to account for differences in the number of resident children, since an unweighted average of markets would give each market the same weight. The agent frame covers 2,527 block groups, of which 1,084 contain a school and 1,443 do not, and only 0.3 percent of the school-less block groups fall outside the density range over which the rural classifier was calibrated.

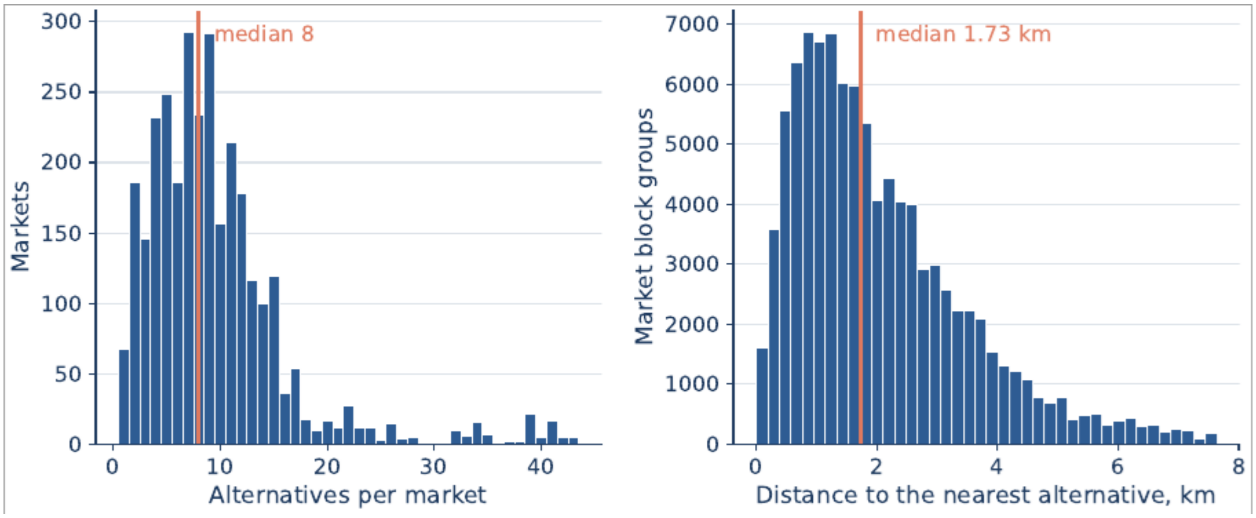
I also constructed an alternative measure by assigning each child to the population-weighted centroid of the residential ZIP and measuring travel from that point. However, rural ZIPs are large and their centroids lie closer to the settled areas, giving the distance measure more measurement error for rural children, which differencing does not remove because distance enters utility nonlinearly. The estimated rural distance contrast has the wrong sign. I therefore do not use this allocation. I retain a student-level rural share from the same ZIP mapping for descriptive purposes only. This continuous measure has a mean of 0.325 and is constant within 76 percent of markets. Constructing a binary indicator requires a threshold: truncating the share to an integer would classify almost every child as urban.

A.6 Rural and poor

To classify block groups as rural, I calibrate the indicator to the Department's school urban/rural tag. I use each school's earliest observed tag to avoid changes in composition caused by closure. Log density separates schools tagged as rural from those tagged as urban with an area under the receiver operating characteristic curve of 0.849. The cutoff that minimizes misclassification is a log density of -8.681, and classification accuracy at that cutoff is 0.788. Coverage is 100 percent of agent block groups and the mean is 0.441 of agent mass. To construct the rural cells for the demand model's rank moments, I use the continuous, kernel-weighted rural share of the closed school's catchment. Its mean is 0.498 over 1,467 closures. Every displaced child enters both the rural and the urban cell with complementary weights, so these cells measure exposure rather than divide the children into separate groups.

The family-poverty flag comes from the Department's family-income classification and is recorded for tested grades. For a displaced child whose flag is missing, I use the most frequent status across all raw enrollment years, which recovers 6,052 of 6,954 missing records and leaves 902. This requires poverty status to be persistent. Within-child consistency is 0.937, and 80.7 percent of children never change status. For the 377,537 children with both an observed flag and a modal flag, the two agree in 96.8 percent of cases. The imputed flag is a modal status—exact for the four fifths of children who never change, and majority

Figure A1. Local School Availability and Distance



Notes: The left panel shows the number of inside alternatives in each of the 3,120 municipality-by-grade-by-year markets in the pre-María demand frame. The mean is 10.18, the median is 8, the tenth percentile is 3, the ninetieth percentile is 17, and the maximum is 81. Urban-majority markets ($n = 821$) have a mean of 13.09 and a median of 6; rural-majority markets ($n = 2{,}299$) have a mean of 9.14 and a median of 9, so the median urban market has *fewer* alternatives than the median rural one and the correlation between a market's mean urbanness and its number of alternatives is $+0.22$. The right panel shows road distance from each market block group to its nearest inside alternative. The mean is 2.22 km, the median is 1.74 km, and the ninetieth percentile is 4.29 km. The unit of observation is a market in the left panel and a market-by-block-group pair in the right panel. The figure is descriptive, and distances are OSRM road kilometers.

status for the rest. Alternative income and eligibility fields in the raw files use different and undocumented definitions and are not used.

A.7 Geographic measures

I calculate road distances and driving durations from the OpenStreetMap network using the OSRM routing engine. This provides a route between the origin and destination for 52,386 of 60,174 displaced children, 87.1 percent. When constructing the menus in Section 4.3, I impute distance for 112 of 10,851 origin–survivor pairs. Figure A1 shows the number of inside alternatives per market and the road distance from each block group to its nearest alternative. The median urban market offers fewer alternatives than the median rural market. The access comparisons concern distance to surviving schools rather than the number of alternatives.

School size also requires care because the enrollment file repeats each school's total on every grade row, with the same total across grade rows in 100 percent of school-years. In addition, 60.5 percent of school–year–grade cells appear on more than one row, with as many as 13 rows. Summing enrollment across rows inflates school size: the median becomes 1,776 compared with the true 256. The sum also combines school size, tested-grade span, and catchment breadth, so using it in a regression imposes the same coefficient on all three relationships. I therefore use the enrollment total on each row to measure school size.

A.8 Estimation windows, by block

Estimation windows differ for the reasons given in Section 3. The production estimates use within-teacher and within-pupil variation in scores standardized within subject, grade, and year. This normalization removes common shifts in score means and scales; remaining shocks are subject to the identifying assumption in Section 5.1. Comparisons based on closure timing face additional confounding from Hurricane María, whose damage was correlated with rural, mountainous, and poor municipalities. Migration after the hurricane is recorded in these data as exit from the public system. I therefore use the full scored window for the production estimates and estimate the variance components over 2011 through 2019. The demand estimation and the incidence calculation use closure years through 2017, the closure-selection comparison uses closure years through 2017 with the 2018 wave excluded, and the reduced-form exit estimates use cohorts through 2015 in the primary specification and through 2016 as a robustness check. The 2018 wave cannot contribute to the stricter match-variance sample requiring at least two scored years on each side of the move, because the panel ends in 2019.

A.9 Exit: the event-time path and the pre-closure placebo

Section 6.1 reports the closure-induced increase in the probability of leaving the public system as a band of 2.2 to 4.4 percentage points with about 3 points central, against a baseline attrition of 6.8 to 7.7 percent a year. This range reflects differences in both comparison groups and follow-up horizons. Across comparison groups, the lower estimate comes from the interaction-weighted estimator of Sun and Abraham (2021) with comparison schools less than 5 percent displaced, the central figure from the matched design with never-treated comparison children on pre-María cohorts, and the higher estimate from the same matched design with all non-displaced children as the comparison group. Across horizons, the estimated response grows with time since closure, from roughly two percentage points at impact toward roughly four by the third year. I keep this range in the paper, stating the comparison group and horizon for each estimate because both enter the comparison.

To assess preclosure differences, I follow a single pre-María cohort of schools whose last operating year is 2015, starting two years before closure. The lead coefficients are close to zero, while the impact coefficient is large: -0.006 ($t = -1.14$) two years before, $+0.002$ ($t = 0.36$) one year before, and $+0.054$ ($t = 3.46$) in the year of displacement. The 2014 cohort corroborates the pattern, with a lead of -0.003 that is not distinguishable from zero and an impact coefficient of $+0.017$. For this exercise, the leads are school-based: I compare the cohort's school with comparison schools in the same cell in each pre-period year. Tracking the cohort of children instead generates a pre-trend mechanically, because children who left the closing school early are removed from the pre-period by construction. Flat leads do not establish that counterfactual exit trends would have been parallel. The interpretation of the exit estimates does not depend on a pre-test p-value.

A.10 The distribution of score changes within the displaced

Section 6.1 reports variation in score changes among displaced pupils. Among rural children displaced by closure, a one-standard-deviation increase in the poverty of the school that closed is associated with a further 0.109 standard deviation decline in the child's own test score (SE 0.041, $t = -2.69$) after correcting for the pre-closure gradient; the association grows rather than fades, reaching 0.144 (SE 0.057) by the second year, which is inconsistent with a purely transitory adjustment cost, and the same gradient estimated on the two years *before* closure is $+0.018$ (SE 0.025), with a 95 percent interval that excludes a pre-trend as large as the result. I also estimate a specification using the within-child change in mean normalized score across all displaced children. The association differs by geography: it is -0.068 (SE 0.025, $t = -2.74$) among rural displaced children, $+0.011$ (SE 0.050) among urban ones, and null among children who moved schools voluntarily; I report the two subsamples separately because their difference is imprecisely estimated. The rural–urban interaction has a t-statistic of 1.53.

The comparison is within the displaced population, so it describes an association rather than the average effect of closure. Since the outcome is the child's own score change and poverty is measured at the origin school rather than at the family level, these estimates do not show that poor children were assigned to worse schools, nor do they separately identify the contributions of origin and destination quality to the score change.

B School and Teacher Effects: Estimation Details and Robustness

I describe here the estimation of the production components in Section 5.1, along with the alternative specifications and decompositions used in Section 6.3. Throughout these comparisons, effects and their standard deviations are measured in test-score standard deviations relative to subject–grade–year peers; variances and covariances are in squared test-score units, and correlations, reliabilities and variance shares are dimensionless.

B.1 The leave-out variance-component estimator

I write (1) in matrix form as $y = X\beta + e$ with $X = [D_a \mid D_\mu \mid D_\phi]$ the pupil, teacher–subject and school design matrices and $\beta = (a', \mu', \phi')'$, and let $\hat{\beta} = (X'X)^{\dagger}X'y$ with hat matrix $P = X(X'X)^{\dagger}X'$. The symbol $\dagger$ denotes the Moore–Penrose generalized inverse of the normal-equations matrix, accommodating the fixed-effect normalizations. Here r indexes pupil–subject–year records, $i(r)$ denotes the pupil in record r, and P_{rr} is that record's leverage. Let x_r' be row r of X, $\hat{e}_r = y_r - x_r'\hat{\beta}$ the fitted residual, and $\sigma_r^2 = \mathrm{Var}(e_r \mid X)$ the conditional error variance. The variances of the teacher, school, school-average teacher and composite school components are quadratic forms $\theta = \beta' A \beta$ with A positive semi-definite. Let $M_m = I_m - \mathbf{1}_m\mathbf{1}_m'/m$. Then $A_\mu = S_\mu'(M_{N_\mu}/N_\mu)S_\mu$, where S_μ selects the μ block; A_ϕ is defined analogously; and $A_q = S_\mu' W'(M_{N_\phi}/N_\phi)WS_\mu$, where W is an $N_\phi \times N_\mu$ matrix of pooled pupil–subject–year observation shares over teacher–subjects within each school. Under conditionally mean-zero, mutually uncorrelated record-level errors, $\mathbb{E}[\hat{\beta}'A\hat{\beta} \mid X] = \beta'A\beta + \sum_r \sigma_r^2 B_{rr}$ with $B_{rr} = x_r'(X'X)^{\dagger}A(X'X)^{\dagger}x_r$, so the plug-in variance is upward biased in expectation. The observation-level leave-out correction subtracts the second term:

$$\hat{\theta} \;=\; \hat{\beta}'A\hat{\beta} \;-\; \sum_r B_{rr}\,\hat{\sigma}_r^2, \qquad \hat{\sigma}_r^2 \;=\; \frac{y_r\,\hat{e}_r}{1-P_{rr}}. \tag{B.1}$$

Under these conditions and $P_{rr} < 1$, $\hat{\sigma}_r^2$ is unbiased for the record's error variance without requiring homoskedasticity, although an individual estimate $y_r\hat{e}_r/(1-P_{rr})$ can be negative because unbiasedness concerns the expectation. This property makes (B.1) valid when the number of estimated effects grows with the sample (Kline et al., 2020). To estimate the model, I absorb pupil effects by demeaning within pupil and then estimate the remaining two-way model, including the pupil term $1/n_{i(r)}$, where $n_{i(r)}$ counts that pupil's records, analytically in the leverage. Otherwise the correction would be too small.

The unrestricted leave-out estimator needs adjustment because errors are correlated across observations. Write the transitory error as $e_r = \zeta_{c(r)} + \eta_{j(r),t(r)} + \xi_r$, where $c(r)$ indexes classroom-year cells, ζ_c is their common shock, η_{jt} is a school-year shock, and ξ_r is the remaining record-level disturbance. These common shocks correlate errors across records, so the diagonal correction in (B.1) is insufficient on its own, while the mover graph is too sparse to leave out whole clusters. I therefore calibrate the two common-shock components and report results over their joint range, following the logic of the extension of the leave-out estimator to clustered dependence. The leave-one-out teacher effect also absorbs a teacher–school match that is constant within the cell, requiring the match correction to subtract the entire match component. The required trace factors, computed over 20,663 teacher–school cells, are 1.686 for the teacher component, 0.945 for the school component and 0.890 for the school-average teacher component. Because the match is bounded near zero (Section 6.2.3), this correction is negligible for the reported components.

I also compare school effects estimated under two testing regimes to assess the time-invariance restriction on ϕ_j. Their correlation is 0.78, and imposing time invariance reduces the estimated dispersion by about 6 percent. The restriction therefore makes the estimated spread of school effects smaller.

The classroom-year input used in the reported run is $\sigma_\zeta^2 = 0.100$. An internal estimator of the sum of classroom and school-year variance returns 0.0625, 0.0624, 0.0636 and 0.0621 on four independent half-samples, compared with an input sum of 0.1224. However, this estimator recovers the sum of the two components and omits a leverage denominator, which can lower the estimate. The estimates near 0.062 recur across splits, but this diagnostic alone establishes neither a population lower bound nor the contribution of sampling error to the discrepancy. Across 14 specifications the implied classroom-shock standard deviation averages 0.269 with a median of 0.277 and a range of $[0.204, 0.339]$, against an independent stayer-autocovariance benchmark of 0.307. Allowing the school-year component to be heterogeneous changes its effective value by 7 to 24 percent, depending on which component is being estimated: 0.0237 for the teacher component, 0.0252 for the school component, 0.0238 for the school-average teacher component and 0.0276 for the composite, compared with a common value of 0.0222. Table 4 reports the leave-out point estimate of each component, the estimates across combinations of the two common-shock inputs, and the estimates from independent classroom splits. The split-sample bootstrap intervals behind that column are $[0.239, 0.267]$ and $[0.224, 0.253]$ for σ_μ, $[0.301, 0.376]$ and $[0.292, 0.352]$ for σ_ϕ, and $[0.094, 0.177]$ and $[0.060, 0.151]$ for $\mathrm{sd}(q)$. The school-level estimates are observation-weighted and use 500-resample school bootstraps on intersections of 1,166 and 1,175 schools. The teacher estimates are unweighted over teacher–subjects and use 500-resample teacher–subject bootstraps. Unweighted over schools, the split gives 0.375 and 0.373 for σ_ϕ and 0.151 and 0.144 for $\mathrm{sd}(q)$; the leave-out estimates are also unweighted over schools or teacher–subjects. For each split estimate I use cross-half covariances from a pair of disjoint classroom samples, with no calibrated noise inputs in either pair. The cross-moments remove estimation error if the errors are mean-zero and uncorrelated across halves, although separating classrooms does not by itself rule out shared school-year shocks.

The split-sample estimates also show the uncertainty in relationships between the components. The covariance between school effects and the school-average teacher component is -0.0078 with a bootstrap interval of $[-0.0215, +0.0042]$ under one split draw and $+0.0020$ with an interval of $[-0.0103, +0.0125]$ under the other, both with standard errors near 0.006. The implied *correlation* changes sign across four statistically identical half-sample estimates and spans a range of roughly seven tenths. Since its denominator is small and imprecisely estimated, I do not report the correlation. Under the split-sample error decomposition, the estimated shared error component, which enters the school and school-average teacher estimates with opposite signs, accounts for at least 85 percent of the observed variance of the estimated school-average teacher component in every half-sample estimation—0.857, 0.868, 0.926 and 0.879. Reliabilities are 0.640, 0.627, 0.621 and 0.640 for the school effect across two draws and two weightings, with an independent within-municipality estimate of 0.624, and 0.364 and 0.347 for individual teacher–subject effects. Only about 11,230 of roughly 15,000 teacher–subjects appear in both halves, and their effects are measured more precisely, so the teacher reliability estimates may exceed reliability in the full sample.

B.2 The components against the literature

To assess whether low estimated teacher dispersion explains the ordering in Table 4, I compare it with estimates from other settings and with the richer pupil specification below. At 0.225, teacher-by-subject value-added here is of the same order as the dispersions reported in the United States literature, whose designs are typically within-school and therefore yield lower bounds on the total dispersion of teacher quality (Rivkin et al., 2005); the ordering of the two components holds across this range of teacher estimates. The pupil-by-subject specification reported below raises teacher dispersion and is more likely to inflate the teacher estimate, so the pooled specification used in the text is conservative in this respect.

The comparison across schools uses teacher quality averaged over each school's roster, since this is the teacher component that differs across schools in the family's choice model. At the reported leave-out calibration it accounts for about 6 percent of the sum of the school and school-average teacher component variances. There is a related finding in other settings: using nearly 3,500 North Carolina transfers, Mansfield (2015) finds teaching quality fairly evenly distributed both across schools and across pupil backgrounds.

B.3 Lagged-score ANCOVA

The pupil-fixed-effect specification controls for persistent individual differences, but does not explicitly condition on prior scores. As a robustness check, I also estimate an ANCOVA specification, replacing pupil effects with grade-interacted cubic terms in lagged mathematics, English, and Spanish scores and individual demographic controls. Conditioning on prior achievement follows the value-added approach of Chetty et al. (2014), while I retain both teacher–subject and school effects to distinguish the two contributions to achievement. Let $\mathbf{x}_{ist}^{\text{anc}}$ collect these lagged-score and demographic controls, with coefficient vector γ^{anc}. The estimating equation is

$$Y_{ist} = (\mathbf{x}_{ist}^{\text{anc}})'\gamma^{\text{anc}} + \mu_{k(i,s,t),s}^{\text{anc}} + \phi_{j(i,t)}^{\text{anc}} + e_{ist}^{\text{anc}}. \tag{B.2}$$

The superscript anc denotes the ANCOVA specification; the outcome and assignment indices are those in (1). Teacher effects remain specific to subjects, school effects are pooled across subjects, and both are held constant over time. There is no pupil fixed effect in this equation, and e_{ist}^{anc} collects the remaining disturbance.

Interpreting the ANCOVA effects as contributions to achievement requires the remaining disturbance to have conditional mean zero given the included controls, current teacher and school assignments, and the teacher and school effects. Assignment may respond to prior scores, but the controls must account for the student heterogeneity that is related to assignment and current achievement. Measurement error in prior scores or unobserved changes in family circumstances can leave this restriction unsatisfied (Rothstein, 2010). It is a different conditioning restriction from Assumption 1, so a comparison of estimates cannot establish either assumption on its own. Since prior scores also reflect educational inputs received in previous years, the ANCOVA effects measure contributions conditional on prior achievement and need not equal the effects in the levels specification.

The ANCOVA estimation uses the 2011–2019 panel and grades 4–8 with the required lagged scores. Grade 3 is excluded because there is no preceding tested grade, whereas the pupil-fixed-effect specification retains it. The teacher–subject estimates available under both specifications have a correlation of 0.849. This is a correlation of estimated effects from the respective estimation samples, without an adjustment for estimation error; differences can therefore reflect sample composition and precision as well as the conditioning specification. A comparison of component dispersion would also require common weighting and comparable corrections for classroom-year and school-year shocks. I do not interpret this correlation as evidence that the corrected variance components in Table 4 are unchanged under ANCOVA.

Table B1 provides a further comparison using residuals from both specifications. Under ANCOVA as under pupil fixed effects, the match estimator returns larger dispersion for teachers who never moved than for closure movers in every stratum. This result concerns the behavior of that estimator, rather than establishing robustness of the school-effect estimates or the school-choice and incidence results.

B.4 Other specification checks

Pupil-by-subject effects. Absorbing subject-specific pupil ability rather than a single pupil effect raises the estimated dispersion of teacher effects, from 0.2950 to 0.3283 corrected and from 0.4370 to 0.4828 uncorrected, on 3.17 million observations. This increase makes the pooled specification reported in the text conservative with respect to teacher dispersion, but the richer specification has 930,000 parameters, a mean leverage of 0.293 compared with 0.120, and 14 unidentified directions compared with two. With 0.29 parameters per observation, the leave-out correction becomes demanding, which is why I use this specification as a sensitivity check producing higher dispersion rather than as an established upper bound or a preferred estimate.

Restricting the graph to closure-forced moves. A connected graph of 1,235 school nodes requires at least 1,234 distinct edges. Closure movers with at least two scored years on both sides of the move supply 514 edges, so their largest possible connected component contains at most 515 schools. A looser forced-mover definition supplies 1,585 edges; that count alone neither establishes nor rules out connectivity, which depends on how the edges link schools. The graph counts differ from the samples used to estimate match variance, the 1,626 and 469 movers described in Section 6.2.3. The strict closure-mover sample cannot connect the full school network on its own.

Why individual school effects are not ranked. For a rural closing school with about four teachers and two and a half scored years, the standard error of $\hat{\phi}_j$ decomposes into a classroom-shock term of 0.086 to 0.137, an idiosyncratic term of 0.060, a match term of 0 to 0.071 and a school-year term of 0.098, for a total of roughly 0.15 to 0.20—the size of the paper's headline contrast. The school-year component averages only over years, whereas classroom and idiosyncratic components also average over observations within a year. For this reason I do not rank individual closed schools, map their estimated effects, or plot them in a scatterplot, although averaging over the 212 rural closed schools reduces some of the error. In the menu-gap regressions, conditional mean-zero measurement error enters the outcome and affects precision without causing errors-in-variables attenuation.

The plug-in teacher–school relationship. In the production estimates, the plug-in correlation between estimated teacher and school effects is -0.310, with variance shares of 0.398 for pupils, 0.138 for teachers, 0.180 for schools and 0.368 for the residual. This is a limited-mobility artifact and not an economic correlation (Andrews et al., 2008). The same artifact is visible in levels: regressing the estimated school effect on the estimated school-average teacher effect within municipality across 1,257 schools gives -0.2035 with a standard error of 0.0381; estimating the two components on disjoint classroom halves, under the assumption that estimation errors are independent across halves, gives -0.0452 (standard error 0.0361) and -0.0720 (standard error 0.0404), so that roughly 70 to 78 percent of the naive coefficient is shared estimation error.

Table B1. Teacher–School Match Placebo Estimates

Specification	Stratum	Closure movers	Teachers who never moved
Pupil fixed effects	All	0.169	0.196
Pupil fixed effects	≥ 2 classrooms	0.120	0.171
Pupil fixed effects	≥ 3 classrooms	0.098	0.161
ANCOVA	All	0.217	0.256
ANCOVA	≥ 2 classrooms	0.145	0.221
ANCOVA	≥ 3 classrooms	0.140	0.220

Notes: Each entry is the match standard deviation from an estimator that scales the classroom-year shock per pupil rather than per classroom-year. Mover selection for the placebo is 1,368 movers, and 439 under the stricter rule; the bound in the text rests on the 1,626 and 469 movers of Section 6.2.3. True match variance is zero by construction in the right-hand column. The comparison is in levels and is not matched on the number of years separating a teacher's two observation blocks, which is why the differences are not interpretable as estimates and only the ordering is claimed. ANCOVA uses the lagged-score specification in Appendix B.3. All figures in standard deviations of achievement.

B.5 The teacher–school match: placebo and mechanism

The match variance is estimated from teachers observed at two schools; the identifying samples are 1,626 closure movers with a scored year on each side and 469 with at least two scored years on both sides. The reported bound is $\sigma_\psi \in [0, 0.10]$, with three estimation approaches that treat estimation error differently, yielding upper limits of 0.105 (a placebo-differenced gradient estimator), 0 (a pooled specification) and 0.095 (a cross-subject covariance estimator, which has no subtracted noise term at all because a teacher's idiosyncratic components are independent across her two subjects).

For the placebo in Table B1, I apply the estimator that scales the classroom-year shock per pupil rather than per classroom-year to teachers who never changed school, dividing their observed years into two blocks. The estimator returns larger match variance for non-movers than for closure movers in all six strata. Since true match variance is zero for a teacher who never moved, this shows that the estimator measures something beyond the loss of a match when a teacher moves. However, the time gap between the two blocks is not matched, so the negative differences do not indicate a negative variance.

The pattern is related to the gradual decline in teacher value-added autocovariance. Among stayers, the autocovariances are 0.176, 0.165, 0.158, 0.147, 0.145 and 0.135 at lags one through six. Therefore, two observation blocks g years apart have an additional difference variance of $2[c(1) - c(g)]$ relative to a one-year gap. Dividing this difference variance by two to express it as a per-match variance gives an apparent match standard deviation of $\sqrt{c(1) - c(g)}$: 0.104 at a two-year gap, 0.136 at three, 0.169 at four and 0.202 at six. These magnitudes are comparable to, but do not encompass all of, the placebo estimates in Table B1. The decay is small enough not to threaten the time-invariant treatment of μ and large enough to generate apparent match variance even without a change in the teacher–school match.

The pooled specification raises a further concern: it returns a negative mover premium at t-statistics between -2.6 and -8.4, which a variance cannot be. One possible explanation is that school effects absorb different amounts of variation for movers and stayers: a stayer's two observation blocks belong to the same school and its effect cancels from the contrast, while a mover's blocks span two schools whose effects are estimated partly from her own pupils and therefore absorb some of the variation being measured. The bias would be toward zero for movers, inflating apparent portability. This concern motivates the upper-bound interpretation and the emphasis on the cross-subject method, in which school effects absorb both subject contrasts for the same mover in the same way. Further calculations from the estimation factorization would be needed to evaluate this explanation and refine the pooled estimate; the upper-bound interpretation still relies on the assumptions of the other methods.

B.6 The forecast test and its correct label

Table B2 shows an earlier forecast exercise in Panels A and C and a later joint regression in Panel B. In the joint regression, I include $\Delta\hat{\phi}$ and the change in school-year teacher quality, $\Delta\hat{q}$, and obtain coefficients of 0.97 and 1.05, but $\Delta\hat{\phi}+\Delta\hat{q}\equiv\Delta\hat{\Psi}$ by construction. If realized changes equal $\Delta\hat{\Psi}$ plus a residual orthogonal to both regressors, the population coefficients are $(1,1)$ regardless of how that composite is divided between them. The accounting identity does not establish the orthogonality condition on its own, so coefficients near one cannot validate the structural interpretation of the two components. In the earlier exercise the coefficient is 0.742 when $\Delta\hat{\phi}$ enters alone. Within any one estimation sample, a univariate coefficient combines that sample's joint forecast coefficients with cross-regressor covariance; it is not a direct estimate of the covariance. The different samples and quality constructions prevent combining the coefficients in Panels B and C through this identity. A related pattern appears in levels. Regressing the estimated school effect on the estimated school-average teacher effect returns -0.2035 with a t-statistic of -5.3, and estimating the components on separate classroom halves substantially reduces it, conditional on independent estimation errors across halves, with 70 to 78 percent of the apparent relationship being shared estimation error (Section B.4).

B.7 Local School Comparisons: Robustness Checks

Table B3 gives every specification behind Section 6.3.3. Figure B1 plots the empirical distribution of the menu gap by cell; the shares above the local benchmark in Table 7 are the heights of the four curves at zero, and the ordering of the curves is monotone where the cell means are not. Two comparisons explain the specifications that control for testing participation. Closed schools tested a larger share of their pupils than their surviving alternatives did, by 33 percent in rural markets and 52 percent in urban markets, and the rural–urban difference in participation change is -0.108 with a standard error of 0.074. The external coefficient of participation on the school effect, estimated on the school-level cross-section, is -0.119 with a standard error of 0.029, while the in-sample coefficient in the gap regression is $+0.060$ with a standard error of 0.049—a difference of 0.18 with a t-statistic near 3.2. Since the coefficient in levels does not carry over to the menu gap, I use the specification without this control as the primary estimate, retaining the controlled specification as a robustness check.

Table B2. Forecasting Achievement Changes

Regressor	Coefficient	Standard error
Panel A. Historical composite forecast		
$\Delta\hat{\Psi}$, attenuation-corrected	0.940	0.011
$\Delta\hat{\Psi}$, uncorrected	0.916	—
Panel B. Joint components, school-year quality		
$\Delta\hat{q}$	1.053	—
$\Delta\hat{\phi}$	0.969	—
Panel C. Historical separate-component forecasts		
$\Delta\hat{\phi}$ alone	0.742	—
$\Delta\hat{q}$ alone	0.157	—

Notes: The dependent variable is the displaced pupil's realized change in test score. Panels A and C reproduce an earlier exercise on 24,263 closure-displaced pupils with a realized score change and both endpoints in the connected set, 40.3 percent of the displaced. For Panel A the reported mean variance of the forecast change is 0.2617 and the mean noise component is 0.0066. The archived summary does not document the precise time aggregation of teacher quality in this earlier exercise. These historical diagnostics therefore cannot be treated as estimates using the same construction as Panel B. Panel B uses 20,775 displaced school-change observations and school-year teacher quality: for a move from origin o_i in year t_i to destination d_i in year t_i+1, $\Delta\hat{q}_i = \hat{q}_{d_i,t_i+1} - \hat{q}_{o_i,t_i}$ and $\Delta\hat{\Psi}_i = \Delta\hat{\phi}_i + \Delta\hat{q}_i$. Panel A's standard error is a bootstrap standard error; the correction is for attenuation from noise in the regressor and is *not* a sample split—these pupils' own scores contributed to estimating the effects on the right-hand side, so the exercise is an in-sample forecast coefficient rather than a held-out validation. **Panel B is reported for completeness and is not a test of the decomposition.** Since $\Delta\hat{\phi} + \Delta\hat{q} \equiv \Delta\hat{\Psi}$ by construction, the population coefficients are $(1,1)$ if realized changes equal the estimated composite change plus a residual orthogonal to both regressors. Under that condition, unit coefficients cannot distinguish alternative splits of the composite. The accounting identity by itself does not impose unit coefficients. Panel C's univariate coefficients do not directly identify cross-component covariance and cannot be combined algebraically with the joint coefficients from Panel B's different sample. Panels B and C are conventional least squares with no adjustment for generated regressors or clustering.

B.8 Absorption: specification and estimates

The model treats ϕ_j as time-invariant over the observed periods. If effectiveness changes when the school receives displaced children, $\hat{\phi}_j$ includes that change with weight equal to the period's share of the school's observations. Thus, $\hat{\phi}_j \approx \phi_j^{\text{pre}} + \theta \bar{A}_j$, where A_{jt} is school j's absorption intensity in year t and θ the shift per unit of absorption. A closed school has $\bar{A}_o \approx 0$, so $\text{gap}_\phi(o)$ inherits $-\theta\bar{A}_{M(o)}$ and the bias in $\hat{b}_{\text{rural}}$ is $-\theta$ times rural's coefficient in a regression of menu absorption on the right-hand side of (12). I estimate both quantities using pupil–subject–year residuals from the production model, which retain the within-school time variation that the time-invariant school effect does not absorb:

$$\hat{e}_{ist} \;=\; \theta A_{j(i,t),t} \;+\; \sum_{h=-3}^{0} \delta_h \mathbf{1}[t - cy(j(i,t)) = h] \;+\; \nu_{j(i,t)} \;+\; \lambda_t \;+\; \varepsilon_{ist}, \tag{B.3}$$

where $cy(j)$ is school j's closure year and the event-time indicators are defined to be zero for schools that never close. Earlier periods form the omitted event-time category. The regression includes school and year effects and standard errors clustered on the school. Estimated on all pupils, $\hat{\theta} = -0.0629$ with a standard error of 0.0186; estimated on the receiving school's own incumbent pupils, dropping the arrivals,

Table B3. Robustness of the Rural Contrast

Specification	Coefficient	t
Baseline specification	+0.148 (0.046)	3.22
Pre-María, schools with limited score data retained	+0.156 (0.046)	3.40
Adding municipality fixed effects	+0.126 (0.074)	1.70
Leaving out closure year 2013	+0.147 (0.046)	3.20
Leaving out closure year 2014	+0.163 (0.050)	3.23
Leaving out closure year 2015	+0.105 (0.053)	2.00
Leaving out closure year 2016	+0.160 (0.050)	3.18
Leaving out closure year 2017	+0.176 (0.067)	2.61
Adding a testing-participation control	+0.154 (0.046)	3.34
Subtracting the external participation bias	+0.135 (0.048)	2.83
Adding the menu-absorption control	+0.144 (0.046)	3.13
Subtracting the absorption bias	+0.148 (0.046)	3.22
Both corrections entered as controls	+0.151 (0.046)	3.25
Both corrections subtracted	+0.135 (0.048)	2.83
Menus in the bottom quartile of absorption ($n = 81$)	+0.238 (0.078)	3.06
Adding log school size	+0.099 (0.047)	2.08
Menu weighted by road proximity rather than enrollment	+0.110 (0.047)	2.36
Menu rebuilt on corrected total-enrollment weights	+0.145 (0.046)	3.15
Full window, including the 2018 wave	+0.070 (0.043)	1.63

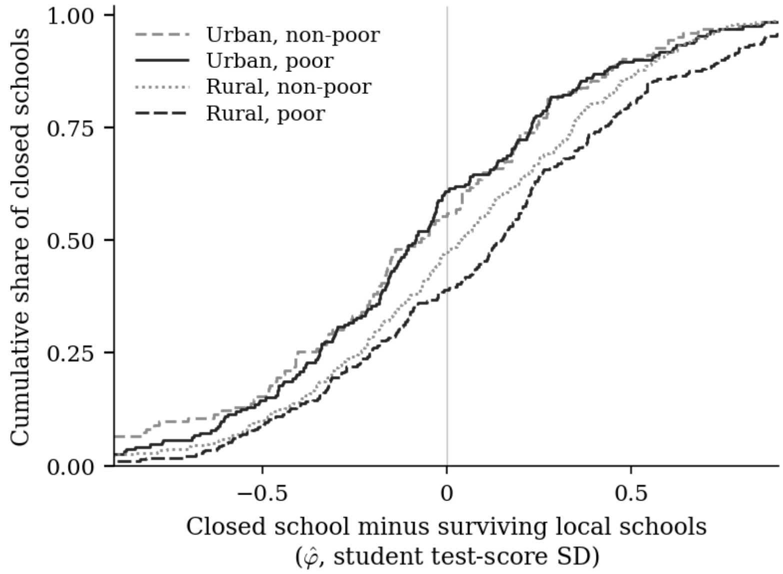
Notes: Each row reports the rural coefficient from a closed-school gap regression. Table 7 reports the rural-by-poverty interaction. The unit of observation is a closed school. Unless a row states otherwise, I use the same sample, estimand, sign convention, and estimator as in Table 7: 317 pre-María closed schools in 76 municipalities, with child-weighted least squares and closure-year fixed effects. Parentheses report standard errors clustered by municipality. After correcting for testing participation and absorption, the estimate ranges from +0.135 to +0.154 and remains distinguishable from zero. Dropping one wave at a time gives positive estimates with a t-statistic of at least 2.00, ranging from +0.105 to +0.176. The 2015 wave has the greatest influence. With municipality fixed effects, the coefficient remains similar but is less precise: only 36 of the 76 municipalities contain both rural and urban closures. Including the post-María 2018 wave in the full-window row reduces the estimate. Schools closed in that wave were above their local alternatives on average, with no rural gradient. The level term is +0.310 (0.092), and the rural interaction is −0.164 (0.085), $t = -1.94$, which is not statistically distinguishable from zero. I did not use the contaminated school-size variable to construct the menus. Rebuilding the menus with corrected weights gives a similar estimate and changes the cell means by at most 0.008.

$\hat{\theta} = -0.0735$. Menu absorption is nearly equal across geography, 0.0324 in rural markets against 0.0288 in urban markets, so rural's coefficient in $\bar{A}_M$ is +0.0029 with a standard error of 0.0022 and the implied bias in $\hat{b}_{\text{rural}}$ is +0.0002. The terminal-year coefficients δ_h are +0.021 ($t = 2.31$) two years before closure, +0.029 ($t = 3.32$) one year before, and +0.014 in the final year, so the time-invariance restriction understates a closing school at the moment of closure and the menu comparison is conservative on this margin.

B.9 Observables and scale

Table B5 decomposes the menu-gap coefficients and the child-level coefficients of Section 6.3.5 into the part predicted by three school characteristics and a residual. To construct the predicted part, I regress the school effect on these characteristics jointly. Table B4 reports this regression and a demographic regression showing that school effects contain variation beyond school composition. Figure B2 plots the municipality-

Figure B1. Distribution of Local School-Effect Gaps



Notes: The figure uses 317 schools that closed before Hurricane María in 76 municipalities. Each observation is a closed school. The plotted variable is $\text{gap}_{\phi}(o)$ as defined in the notes to Figure 3; positive values mean the closed school had a higher estimated effect than the enrollment-weighted mean of its surviving local alternatives. The curves show empirical cumulative distributions by origin-school rurality and poverty, with poverty split at the median. Each school receives equal weight in these curves. The cell means in Figure 3 instead use child weights. Units are student test-score standard deviations, normalized within subject, grade and year. The share of closures with a positive gap rises monotonically across the four cells: 0.246 urban poor, 0.338 urban non-poor, 0.539 rural non-poor and 0.610 rural poor. These shares use estimated gaps. Measurement error can change the sign of a gap, so the shares do not provide bounds on the shares of true gaps. The curves are descriptive and are reported without inference.

demeaned distribution of estimated school effects for closed and surviving schools and, for closed schools, by cell after deflating for estimation error: the four cell means span 0.10 and dispersion is widest among rural-poor schools.

In the decomposition, total = predicted + residual holds exactly, so applying the same estimating equation to each part reproduces the total. The choice of sample for the gradients is less straightforward. Using the 664 surviving schools excludes closed-school observations but requires the relationship estimated on open schools to apply to closed schools as well, whereas using all 1,257 schools includes the closers. I report both sets of estimates without choosing a preferred column. Because the predicted component is a generated regressor, I also use two inference procedures: a wild cluster bootstrap-t treating the gradients as known, and a two-step pairs cluster bootstrap that resamples municipalities and re-estimates them. The two-step procedure changes the additive rural residual from $p = 0.042$ to $p = 0.066$ and leaves the interaction essentially unchanged, because the interaction's predicted component is near zero.

B.10 The child-level change in school effect

Table B6 shows the four cell means of the realized change in Section 6.3.5, which I divide into the removal component, evaluated at the enrollment-weighted mean of the surviving menu, and the destination component, the remaining difference between that mean and the actual destination. Table B5 reports the

Table B4. School Effects, Size, and Demographics

	Coefficient		t
Panel A. The three scale terms, entered jointly			
Log school size	-0.1137	(0.0379)	-3.00
Log tested-grade span	$+0.0210$	(0.0491)	$+0.43$
Log catchment breadth	-0.0672	(0.0474)	-1.42
Panel B. School demographics			
Urban	-0.0183	(0.0286)	-0.64
School poverty rate (standardized)	-0.0132	(0.0151)	-0.87
Tested-grade span	$+0.0057$	(0.0071)	$+0.80$
Ever closed	-0.0173	(0.0305)	-0.57
R^2, raw		0.058	
R^2, deflated by the reliability of $\hat{\phi}$		0.093	

Notes: The unit of observation is a school. The sample contains 1,257 schools in 78 municipalities. Both panels use the estimated school effect $\hat{\phi}_j$ as the dependent variable, measured in student test-score standard deviations normalized within subject, grade, and year. I include municipality fixed effects, so the comparisons are within municipalities. Parentheses contain standard errors. *Panel A.* A log point of enrollment is associated with a school effect lower by 0.11 standard deviations, holding tested-grade span and catchment breadth fixed. I include the three terms jointly. Log school size and log catchment breadth have a correlation of $+0.63$, so separate regressions do not isolate their associations. Catchment breadth also correlates $+0.22$ with urban status and -0.35 with closure. The median school draws from 2.0 residence markets per school-year; the ninetieth percentile is 3.9. I do not report a univariate size gradient. In simulated data with known coefficients, that specification gave -0.171 when the true coefficient was -0.125. The recorded enrollment variable repeats the school total on every grade row, so its logarithm combines size, span, and breadth gradients. *Panel B.* School demographics explain little of the estimated school effect. Urban status, the school poverty rate, and grade span each predict essentially none of it. After deflating the raw R^2 by the reliability of $\hat{\phi}$ (0.624), observables explain about 9 percent of the true variance, essentially all through scale. Conditional on municipality and scale, closure was not selected on the school effect overall. This is what the sign reversal between urban and rural markets in Table 7 implies.

additive-plus-interaction regression and its decomposition into predicted and residual parts. The cell means are computed on 34,772 moves across 325 origin schools and the regressions on 34,118 children in 317 origin schools, with close-year fixed effects and standard errors clustered on the origin school. A wild cluster bootstrap-t with 999 replications, the null imposed and Rademacher weights over origin clusters corroborates the clustered inference for the interaction, since simulation at this design puts the clustered rejection rate at 0.062 against a nominal 0.05. Across the four cells the components predicted by school size, tested-grade span and catchment breadth span only 0.067, against a realized span of 0.210: the decomposition therefore compares how much the predicted and realized components vary across cells.

There is a further limitation to the predicted component: the cross-sectional relationships in Table B4 do not fully explain the changes associated with moves. Regressing the realized change on the fitted index gives a slope well below one, whereas fully applying those relationships to the moves would give a unit slope. Realized changes therefore vary less than one-for-one with the index, but this does not by itself establish upper or lower bounds on the observable and residual shares of every group contrast. Possible explanations include correlation between destination choice and unobserved destination quality, the selection of origins onto closing schools, and remaining shared estimation error on the moves that cross municipality boundaries.

Table B5. Decomposition of Local School-Effect Gaps

Panel A. School-level menu gap

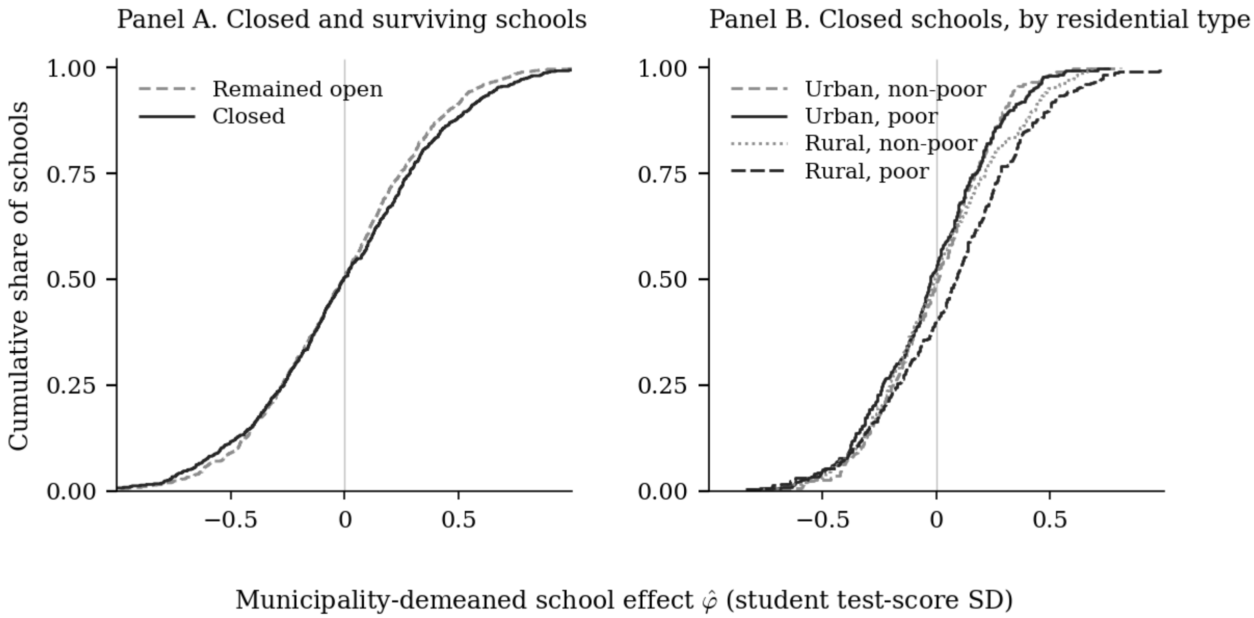
	Rural		Rural × poor	
	Survivor gradients	All-school gradients	Survivor gradients	All-school gradients
Total	+0.148	+0.148	+0.201	+0.201
Predicted by observables	+0.062	+0.046	−0.021	−0.012
	(0.011)	(0.010)	(0.021)	(0.023)
Residual	+0.085	+0.102	+0.223	+0.213
	(0.048)	(0.046)	(0.099)	(0.096)
Share of the total predicted	42%	31%	−11%	−6%
Two-step p on the residual	0.134	0.066	0.028	0.040

Panel B. Child-level realized change in the school effect

	Rural	Poor	Rural × poor
Total	−0.006	+0.051	−0.204
	(0.078)	(0.084)	(0.103)
Predicted by observables	−0.045	−0.008	−0.014
	(0.017)	(0.018)	(0.021)
of which school size	−0.038	−0.012	−0.006
of which tested-grade span	+0.003	+0.001	−0.001
of which catchment breadth	−0.010	+0.004	−0.007
Residual	+0.039	+0.058	−0.191
	(0.081)	(0.088)	(0.106)

Notes: Panel A. The unit of observation is a closed school. The sample contains 317 pre-María closures in 76 municipalities, with the same estimand and estimator as Table 7. I construct the predicted component using school-effect gradients in log school size, log tested-grade span, and log catchment breadth, applied to the corresponding differences between each closed school and its surviving menu. I estimate these gradients either on the 664 surviving schools or on all 1,257 schools. The decomposition uses the canonical menu file, whose coefficients differ from the primary estimates in Table 7 by no more than 0.11 standard errors. I report totals at the primary values; total = predicted + residual holds to rounding. Parentheses contain standard errors. I use two inference procedures: a wild cluster bootstrap-t that holds the gradients fixed and a two-step pairs cluster bootstrap that re-estimates them. The reported p-value uses the two-step procedure and includes uncertainty in the estimated gradients. Appendix B.9 discusses limitations of both gradient samples. I therefore report both without selecting a preferred column. School size, tested-grade span, and catchment breadth explain 31 to 42 percent of the additive rural coefficient, but none of the interaction. The predicted interaction components are within one standard error of zero and provide no clear evidence that these observables contribute to the interaction. *Panel B.* The unit of observation is a forcibly displaced child. The sample contains 34,118 children from 317 origin schools before María. The regressions include closure-year fixed effects and cluster standard errors by origin school. The outcome is the realized change in the estimated school effect. The separate rural and poverty coefficients cannot be distinguished from zero; the contrast is concentrated in their interaction. Observables explain 7 percent of that interaction, leaving 93 percent in the residual. Across children, they explain 5.3 percent of the variance in the realized change, with a bootstrap interval of 4 to 7 percent. A wild cluster bootstrap-t over origin clusters corroborates the clustered inference: the simulated rejection rate is 0.062 at a nominal level of 0.05. The interaction is imprecisely estimated. The absolute coefficient-to-standard-error ratios are 1.98 for the total and 1.80 for the residual. Power and expected exaggeration conditional on significance depend on assumptions about the true effect and its sampling distribution. These ratios do not identify bias-corrected true effects, and I apply no such correction to the decomposition.

Figure B2. School Effects by Closure Status



Notes: Each observation is a school. Panel A shows the empirical cumulative distribution of the municipality-demeaned estimated school effect $\hat{\phi}_j$ separately for schools that closed and schools that remained open. Panel B shows the distribution for closed schools by school rurality and poverty, after adjusting for estimation error in $\hat{\phi}$. The panel's label "residential type" refers to these school groups. Units are student test-score standard deviations, normalized within subject, grade and year. Because $\hat{\phi}$ is estimated, the raw distributions in Panel A are wider than the underlying distribution of school effects: the estimated reliability of $\hat{\phi}$ is 0.62 using within-municipality variation, so the raw distributions are about 1.27 times too wide. I therefore use Panel A to compare the locations of the distributions. Panel B adjusts their dispersion for estimation error. The adjusted cell means and standard deviations are -0.024 and 0.261 for urban non-poor ($n = 243$), -0.040 and 0.306 for urban poor ($n = 318$), -0.003 and 0.288 for rural non-poor ($n = 386$) and $+0.064$ and 0.336 for rural poor ($n = 310$); the four cell means span only 0.104. The figure is descriptive. Estimated effects for individual small rural schools have standard errors of 0.15 to 0.20.

B.11 Displaced teachers: exit, destination and the composition of exit

Table B7 reports the teacher outcomes described in Section 6.3.7: exit from the tested panel, the destination school's preclosure mean teacher value-added, and movement with a teacher's own displaced pupils. Panel D shows the composition of exit. Of pre-hurricane exits, 26.6 percent are consistent with retirement on age or tenure and 73.4 percent are not; the retirement-consistent group has mean value-added 0.17 standard deviations below the mean against stayers' $+0.09$, while non-retirement exit is quality-neutral. I retain this retirement classification as provisional, since the flags have not been verified against the territory's pension statute.

Table B6. Decomposition of Destination School-Effect Changes

	Urban non-poor	Urban poor	Rural non-poor	Rural poor
Realized change in the school effect	$+0.0057$	$+0.0516$	-0.0165	-0.1587
	(0.0720)	(0.0538)	(0.0390)	(0.0459)
Removal component	$+0.0257$	$+0.0736$	-0.0330	-0.1563
	(0.0706)	(0.0562)	(0.0403)	(0.0436)
Destination component	-0.0200	-0.0219	$+0.0165$	-0.0025
	(0.0163)	(0.0180)	(0.0179)	(0.0191)

Notes: The table uses 34,772 moves by children displaced from 325 origin schools in the pre-María period. Each observation is a move from a closed school to a destination school. The outcome is $\Delta\hat{\phi}$, the estimated school effect of the destination minus that of the closed school, in student test-score standard deviations normalized within subject, grade and year. A negative value means the child moves to a school with a lower estimated school effect. I define cells using the *origin school*: rurality follows the school's urban/rural tag, and poverty divides schools at the median poverty rate. These groups differ from the household groups in Table 9. Parentheses report standard errors clustered by origin school. To calculate the removal component, I replace the realized destination with the enrollment-weighted mean of the surviving local menu. The destination component is the remaining difference between the actual destination and this benchmark. The two components sum to the realized change by construction. The destination component never exceeds 0.022 in absolute value in any cell and is statistically indistinguishable from zero. The removal component accounts for most of the rural-poor mean contrast. This decomposition does not identify the causes of destination choice. For rural-poor children, the change in school size comes from the removal component. Families choose schools slightly smaller than the enrollment-weighted menu, so their destination choices partly offset the loss. The rural-poor removal component of -0.1563 and the school-level rural-poor menu gap of $+0.1367$ in Table 7 have consistent interpretations under opposite sign conventions. Their samples and weights differ, so the two averages need not be equal. The estimated change in the school effect across a forced move has a reliability of 0.80. Table B5 decomposes these cells into rurality, poverty, and their interaction, and examines the school characteristics that may explain the differences.

Table B7. Teacher Mobility after School Closures

Outcome	Regressor	Coefficient	t	Semipartial R^2
Panel A. Leaving the tested panel ($n = 879$)				
Probability of exit	Own value-added	−0.0486 (0.0228)	−2.13	—
Probability of exit	Tenure	+0.0336 (0.0160)	+2.11	—
Panel B. Destination-School Teacher Value-Added ($n = 541$)				
Mean teacher value-added	Own value-added	+0.0358	1.25	0.0044
Mean teacher value-added	Seniority	+0.0000	0.00	0.0000
Panel C. Moving with one's own displaced pupils ($n = 541$)				
Followed own pupils	Own value-added	−0.0260	−0.62	—
Followed own pupils	Seniority	+0.0598	+2.19	—
Panel D. Composition of exit (pre-María, $n_{\text{exit}} = 428$)				
Reassigned in-system and detected		0.0% (a detection limit)		
Retirement-consistent		26.6%; mean value-added −0.172 ($n = 114$)		
Other exit		73.4%; mean value-added −0.060 ($n = 314$)		
Stayers, for comparison		mean value-added +0.094 ($n = 752$)		

Notes: The unit of observation is a displaced teacher of a tested subject. Closures displaced 2,145 such teachers, and 999 (46.6 percent) left the tested panel. I restrict all regressions to pre-María closures and include closure-year fixed effects. Panels B and C also include origin-school fixed effects, comparing colleagues displaced from the same school. Regressors are measured in standard-deviation units. Parentheses contain standard errors clustered by origin school. I omit them where the source reports only a t-statistic. Mean teacher value-added is the destination school's mean teacher value-added before closure. The arriving teacher therefore does not mechanically affect this measure. The seniority coefficient on mean teacher value-added is exactly zero to four decimals, with a t-statistic of zero. A standard error cannot be recovered from these values. I report the point estimate and semipartial R^2 instead; a permutation test using random reassignments gives a p-value of about one. The destination's urban, poverty, and distance margins are also null ($|t| \leq 1.4$). Panel C asks whether teachers move to the school that receives their own displaced pupils. The results do not support the interpretation that teachers with higher value-added were less likely to move with their pupils. In Panel D, leaving the tested panel does not necessarily mean leaving teaching. The staff file covers tested grades only, so a move to an untested grade or administrative position enters the other exit category. The zero in the first row reflects a detection limit, rather than a measured zero. I provisionally classify retirement by age of at least 55 or tenure of at least 30. This classification has not been checked against the governing pension statute. Age is observed for 66.8 percent of exiters. Quality selection among exiters is concentrated in the provisional retirement category; non-retirement exit is quality-neutral and more common among junior teachers.

C Demand Estimation: Moments, Calibration, and Validation

I describe here the moments and calibrations used in Section 5.2, including the exit target, nesting parameter, and distance transformation, before examining model fit, residence measurement, and identification of the rural distance increment. I also explain why market population differs from observed enrollment and derive the auxiliary travel calculation for teacher quality.

C.1 Moment conditions and identification

Let $\theta = (\lambda_0, \lambda_{\text{rural}}, \lambda_{\text{pov}}, \beta_{\text{rural}}, \beta_{\text{pov}}, \pi_{\text{pov}})'$. For every candidate θ, mean utilities solve the share equations

$$s_{mj} = \int p_{mj}(z; \mathscr{J}_m, \theta)\, dF_m(z), \qquad j \in \mathscr{J}_m. \tag{C.1}$$

Here s_{mj} is school j's observed enrollment share in market m, and the vector $\delta_m(\theta)$ contains the school–market mean utilities $\delta_{mj}(\theta)$. As in Section 5.2, I evaluate the probability function at the mean utilities that solve these equations, where F_m is the maintained joint distribution of household characteristics and M_m is resident school-age population. I retain these mean utilities when reducing the choice set, together with the calibrated p and ρ.

Write $\mathscr{C}_m$ for the closing origins in market m. Equation (10) defines the joint first- and second-choice probability $T_{moj}(z;\theta)$. Including the outside alternative,

$$\sum_{j \in (\mathscr{J}_m \setminus \{o\}) \cup \{0\}} T_{moj}(z;\theta) = p_{mo}(z; \mathscr{J}_m, \theta). \tag{C.2}$$

Since the joint probability T already incorporates selection into the origin school, conditioning further on inside re-enrollment requires summing T over inside destinations only in the denominator.

Index the four cells by $c \in \{\text{UN}, \text{UP}, \text{RN}, \text{RP}\}$, using this order when stacking their moments, where U and R refer to origin catchment exposure and N and P to administrative family poverty. Write $z = (b, P, I)$, so that P is the administrative family-poverty component of household type z. If A^R_o is the origin's rural exposure, the cell weights are

$$(a_{\text{UN}}, a_{\text{UP}}, a_{\text{RN}}, a_{\text{RP}})(o,z) = \big((1-P)(1-A^R_o),\ P(1-A^R_o),\ (1-P)A^R_o,\ PA^R_o\big).$$

They use exposure and its complement, rather than assigning an observed residential rurality to each child. For a destination statistic $h(m,o,j)$, let $r_h(m,o,j)$ indicate the required data coverage. Its model counterpart is

$$H_{ch}(\theta) = \frac{\sum_m M_m \sum_{o \in \mathscr{C}_m} \int a_c(o,z) \sum_{j \in \mathscr{J}_m \setminus \{o\}} r_h(m,o,j) h(m,o,j) T_{moj}(z;\theta)\, dF_m(z)}{\sum_m M_m \sum_{o \in \mathscr{C}_m} \int a_c(o,z) \sum_{j \in \mathscr{J}_m \setminus \{o\}} r_h(m,o,j) T_{moj}(z;\theta)\, dF_m(z)}. \tag{C.3}$$

For the pooled distance moment, I set $a_c = 1$. The distance moments require measured origin-to-destination distance, while the quality moments require both distance and quality coverage; I use the same restrictions and cell weights in their empirical counterparts, giving each student unit weight before applying the exposure weights. All denominators must be positive.

In (C.3), I use the origin-year inside menu after removing o and exclude destinations that are absent from this grade-specific menu. Since the pairwise calculation does not separately remove other schools that close at the same time, Assumption 2 requires these changes to leave the represented destination choices unchanged. The data-coverage restrictions must also select the same population for the observed and predicted moments.

Distance statistics consist of indicators for ranks 2, 3, 4–5, and 6+, measured from the origin, and pooled relocation kilometers. The quality statistics consist of the four retained quality-rank indicators and $q_{j,t-1}$ less the unweighted mean among schools in $\mathscr{J}_m \setminus \{o\}$ with measured quality, including j. For each quality statistic, collect the four predicted cell means in $H_h(\theta)$, and let H_h^{data} denote the corresponding population targets. Sample averages estimate these targets in the GMM criterion. Define

$$B = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ -1 & 0 & 1 & 0 \end{pmatrix}, \qquad g_h^q(\theta) = B\big[H_h^{\text{data}} - H_h(\theta)\big]. \tag{C.4}$$

The first two rows compare poor and non-poor students within each exposure cell, while the last compares rural and urban exposure among non-poor students, giving three independent contrasts for each of five quality statistics. I omit the first-rank share and fourth group contrast because adding up implies both.

Let θ_0 denote the population parameter vector satisfying the maintained moment restrictions. Assumption 2, together with the maintained population and choice distributions and matched sampling rules, is sufficient for these moment conditions. A weaker sufficient restriction for the quality block alone permits

$$H_h^{\text{data}} - H_h(\theta_0) = a_h \mathbf{1}_4, \qquad h = 1, \dots, 5, \tag{C.5}$$

where a_h is the same across the four normalized cells. Since $B\mathbf{1}_4 = 0$, the retained contrasts remove this discrepancy, although school mean utilities that are common to households do not imply that it is equal across cells. This relaxation applies only to the quality block: the distance, composition, and exit moment conditions must still hold at the same θ_0.

Let $g(\theta)$ stack the seventeen distance conditions, fifteen quality contrasts, composition condition, and exit condition after share inversion. Under valid population moments, $g(\theta_0) = 0$. With differentiable moments and an interior parameter vector, a sufficient condition for local identification is

$$\operatorname{rank}\left[\left. \frac{\partial g(\theta)}{\partial \theta'} \right|_{\theta=\theta_0} \right] = 6. \tag{C.6}$$

This derivative includes the response of $\delta_m(\theta)$ to each coefficient, so the condition requires the coefficients to produce locally distinct substitution responses after enrollment shares have been matched. Having thirty-four moments and removing algebraically redundant contrasts does not establish the rank condition, nor does the condition imply global uniqueness. I minimize the sample GMM criterion and use the specification comparisons below to examine sensitivity to the maintained demand structure.

C.2 Estimation moments

At each candidate parameter vector, I invert mean utilities to match enrollment shares and use a micro-moment criterion to match features of displaced students' destinations along with the composition and exit targets. I estimate the coefficients by generalized method of moments in PyBLP (Conlon and Gortmaker, 2020, 2025), verifying the optimum by replicating it from different starting values rather than using a gradient norm. Table C1 lists the moment blocks.

Table C1. Demand Estimation Moments

Block	Count	Target and population	Main parameters
Distance rank	16	Exposure-weighted share of displaced re-enrollers choosing another origin-year school in each retained distance-rank bin, for four type cells $\times$ rank bins $\{2,3,4\text{–}5,6+\}$; the first-rank share is dropped in each cell by adding up	$\lambda_{\text{rural}}, \lambda_{\text{pov}}$
Mean relocation distance	1	Mean realized origin-to-destination distance, target 3.90 km, effective $N = 13{,}666$, unit weight per student, no cross-cell content	λ_0 (scale)
Differenced quality	15	Rank and cardinal-gap contrasts in lagged q across three independent type contrasts; the fourth is algebraically redundant	$\beta_{\text{rural}}, \beta_{\text{pov}}$
Composition (COMP)	1	$P(\text{income-poor} \mid \text{public enrolled}) = 0.6706$ (0.0036), ACS/PUMS 5-year 2013–2017, Puerto Rico, children 8–14 with poverty status determined	π_{pov}
Extensive margin (EXIT)	1	Calibration target for the closure-induced change in outside probability, target 0.030	Calibrates ρ
Total	34		

Notes: Estimation frame: origin years 2013–2017 and destination years 2014–2018; 3,120 municipality-by-grade-by-year markets, 1,281 schools, 31,767 product-market rows, 285,800 agent rows, 26,675 displaced micro records, 78 municipality clusters. Rank-moment type cells are exposure weights, not a partition: each displaced child enters both the rural and the urban cell with weights A^R and $1-A^R$, where A^R is the population-weighted, distance-decayed rural share of the closed school's catchment (mean 0.498 over 1,467 closures). I estimate the coefficients jointly from all moment blocks; the final column indicates the main source of information for each block. The composition target and the outside-share benchmark in Appendix C.9 use the same census extract and frame. They therefore do not provide independent checks on one another. The exit target informs the calibration of ρ and remains in the GMM criterion when ρ is fixed. It continues to restrict the other coefficients, so its fit is not independent validation.

C.3 The exit target and its composition

I construct the calibration target of 0.030 from two components. The first is the causal impact effect of closure on public enrollment, -2.23 percentage points with $t = -4.08$, which I estimate in an interaction-weighted event study using a matched control group, with coarsened exact matching on prior public enrollment, grade, region, and urban/rural status (Iacus et al., 2012; Sun and Abraham, 2021). I restrict this sample to grades below 10 and move years through 2016. The second component adds approximately 0.7 percentage points for cross-municipality switching, a rough increment rather than a separately estimated causal effect. Since the model's inside menu contains only same-municipality schools, it records these switches as movement to the outside option, so the target concerns leaving the origin municipality's public-school choice set rather than exit from public school alone. The baseline demand estimation fits 0.028.

C.4 What the fit checks establish

I use three exercises to assess model fit and estimation. First, the predicted exit response is 0.028, compared with the calibrated target of 0.030, but since ρ was calibrated to reproduce this target, agreement between them checks the consistency of the solution and does not independently show that the model matches the causal exit rate. I do not use this fit to validate the preference estimates in Section 4.2 or the incidence estimates in Section 6.4.

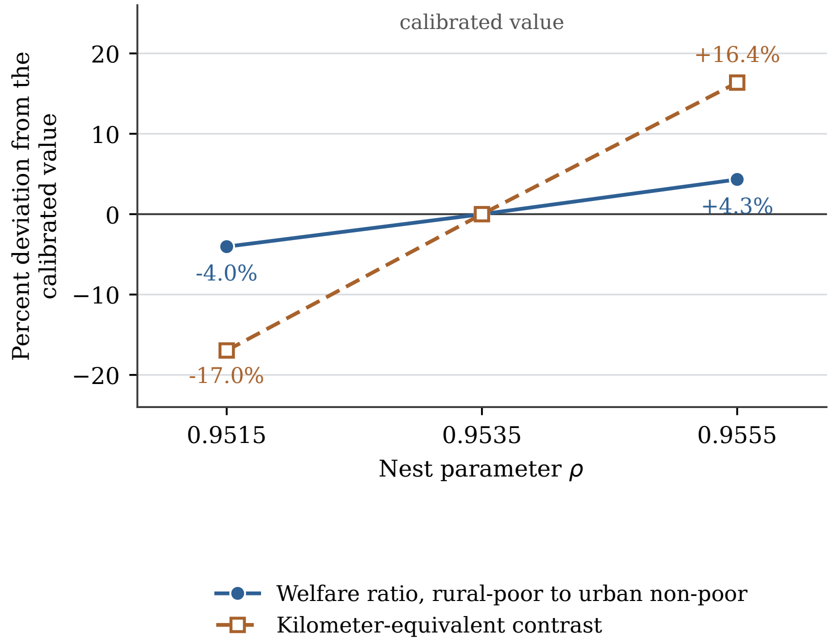
Second, the moments do not target the *ordering* of post-closure exit across groups. The model produces more exit among rural than urban displaced children and more among non-poor than poor children, with the same signs as the causal event-study responses in Section 6.1, although the gaps are roughly half as large. Since the group split comes from an earlier distance transformation and a different value of the inside-option shifter, I use it only as evidence on the direction of the comparisons and not their magnitudes.

Third, the Monte Carlo in Appendix C.7 shows that the rank-plus-exposure design recovers the data-generating rural increment while ZIP-centroid imputation gives the wrong sign, supporting the design as a way to address residence measurement error. It does not validate the ecological identification of the rural increment, which I examine in Appendix C.8. Matching the data also does not establish predictive accuracy, since school-choice predictions can be wrong because of auxiliary population inputs as well as preference estimates (Pathak and Shi, 2021); the population discrepancy in Appendix C.9 concerns one such input.

C.5 Nesting Parameter: Calibration and Sensitivity

For a household with given location and characteristics, the odds of choosing between two inside schools depend on utility differences scaled by $1-\rho$. Since the integrated destination moments also reflect selection into the origin school and subsequent inside re-enrollment, their composition can respond to ρ and the inside-option shifter, with the outside margin providing additional information about the utility scale. I obtained $\rho = 0.953524$ from a joint (λ, ρ, π) generalized method-of-moments run with 19 moments and five nonlinear parameters; the run used 27 iterations, returned a standard error of 0.0021, and had a numerically

Figure C1. Sensitivity to the Nesting Parameter

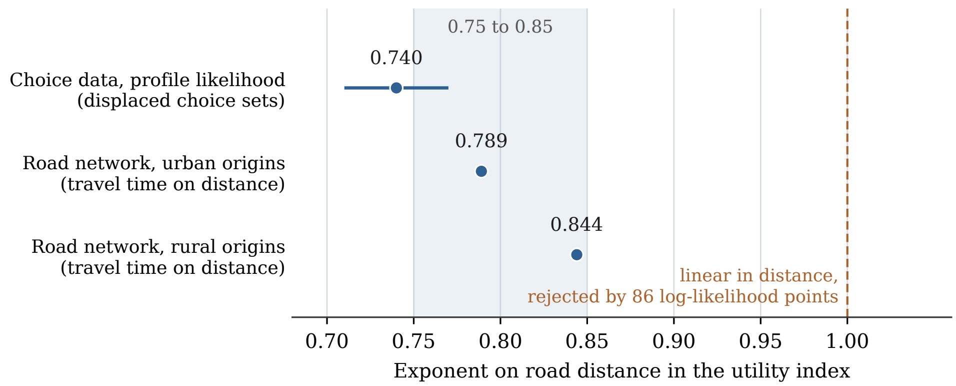


Notes: I calibrate the nesting parameter ρ using an earlier joint GMM estimation of the distance, nesting, and income-poverty parameters. That estimation gives 0.953524 using 19 moments and 27 iterations, with a standard error of 0.0021 and a numerical check of the GMM objective function. I then hold the parameter fixed at this value to match the calibration target of 0.030 for leaving the local public-school choice set. The horizontal axis is a three-point grid around that value; the middle point is the calibrated value 0.953524. I index each series to its value at the calibrated point. Both series therefore share an axis and show changes relative to that point rather than levels. The welfare ratio is the rural-poor to urban-non-poor ratio of mean welfare losses and moves by approximately ± 4 percent around its calibrated-point value; the kilometer-equivalent contrast moves by approximately ± 17 percent, which is why the main results emphasize access losses in utility units. The rural-poor loss share increases between adjacent grid points (changes of $+0.0017$ to $+0.0067$). These model-implied results use Population A of the earlier distance-only welfare calculation. This specification includes only the three distance parameters: the base coefficient and its rural and poverty increments. It omits the income-poverty inside-option shifter and the quality interactions. The incidence *shares* are ratios and change comparatively little across the grid. This does not imply that they are invariant to ρ.

checked slice of the GMM objective function. I then held this value fixed in the baseline demand estimation with $d^{0.8}$. This estimation therefore provides no standard error for ρ, and I report none, since assigning a standard error to ρ from this estimation would contradict its treatment as a fixed calibration input. The within-nest scale factor is $1/(1-\rho) = 21.52$, with a delta-method standard error of 0.93. This factor scales the inside utility index and does not multiply every welfare loss or kilometer equivalent by the same amount. The standard errors for λ and β condition on the calibrated ρ and exclude its first-stage uncertainty.

Figure C1 reports sensitivity of access losses to $\rho \in \{0.9515, 0.953524, 0.9555\}$. For this exercise, the welfare evaluation uses only the three distance parameters, excluding the type-specific inside-option and quality shifters. Across the band, the welfare ratio between the most and least affected types changes by about 4 percent, and the kilometer-equivalent contrast changes by about 17 percent. Between adjacent grid points, the rural-poor loss share increases by $+0.0017$ to $+0.0067$, so increasing ρ makes the result slightly more regressive. Welfare shares are not invariant to ρ, although the distance-versus-distance and distance-versus-quality *ratios* are invariant by construction because they do not contain $(1-\rho)$.

Figure C2. Estimates of the Distance Exponent



Notes: Distance enters utility as d^p in raw kilometers. The two road-network rows show within-origin elasticities of log travel time with respect to log road distance, calculated from OSRM routes: 0.789 (SE 0.0004) over 174,467 urban origins and 0.844 (SE 0.0007) over 93,699 rural origins. Minutes per kilometer are similar across zones (1.434 urban, 1.373 rural), so the exponent does not need to account for a rural speed penalty. For the choice-data row, I estimate a conditional logit on displaced students' choice sets with $\hat{\delta}$ as an offset. The sample contains 14,092 students and 260,527 rows. The profile-likelihood estimate is $\hat{p} = 0.740$ with a 95 percent profile interval of $[0.710, 0.770]$. Horizontal bars are 95 percent intervals; on the two routing rows they are narrower than the marker. Shading marks the interval 0.75 to 0.85, which contains both routing estimates. The choice estimate, 0.740, lies just below this band and outside the urban routing interval, although its profile interval overlaps the band; the implied change in marginal disutility at 3.4 km is about 1 percent, so the economic difference is negligible. Relative to the estimated form, linear distance loses 86 log-likelihood points, $\log(1+d)$ loses 410 to 467 and $\log d$ loses 1,143. I use the external calibration of 0.8 for the exponent rather than fit it to choices.

The comparison with an earlier specification also shows the role of scale: after removing nest scaling, the current base distance coefficient is -0.92, compared with -1.07 in the earlier non-nested, linear-distance specification, and the rural gap is -1.32, compared with -1.05. The nest absorbs the difference in scale without changing the qualitative comparison.

C.6 The distance functional form

Figure C2 brings together three sources of evidence on the exponent p in d^p. First, within-origin routing regressions of log travel time on log road distance give an exponent of 0.789 (standard error 0.0004, $n = 174{,}467$) at urban origins and 0.844 (0.0007, $n = 93{,}699$) at rural origins. Minutes per kilometer are similar across geography, at 1.434 for urban origins and 1.373 for rural origins, leaving no rural speed penalty for the functional form to absorb. Second, a pooled calculation using 279,759 school-to-school pairs gives 0.803. Third, I profile p in a conditional logit using displaced students' choice sets, with 14,092 students and 260,527 rows and $\hat{\delta}$ entering as an offset, obtaining $\hat{p} = 0.740$ and a 95 percent profile interval of $[0.710, 0.770]$.

I *calibrate the exponent externally* using road geometry rather than fitting it to the choices that identify λ, so the reported λ standard errors condition on this calibration. The choice-based estimate lies just outside the routing value, although the difference has a small economic effect: moving from $p = 0.75$ to $p = 0.80$ changes marginal disutility at the median displaced move by about one percent. The alternative functional forms fit less well. Relative to $d^{0.8}$, linear distance loses 86 log-likelihood points, $\log(1+d)$ loses 410 to

467, and $\log d$ loses 1,143. A spline with knots at 2, 4, and 8 kilometers improves fit by 155 points at the cost of nine additional parameters, but since its rural segment is not robust to the definition of the rural indicator, I keep the spline as a robustness comparison rather than the primary specification.

C.7 Residence measurement: the ZIP Monte Carlo

Since the micro frame records residence at ZIP level, I examine the resulting measurement error by running the estimator on synthetic geography calibrated to the observed distribution of menu size and the spread of catchment rural exposure, 0.71 compared with an observed 0.709. I use 26 replications across three seeds, with starting values chosen independently of the data-generating parameters and a rural coefficient starting at exactly zero. The data-generating coefficients are (-0.352, -0.104, -0.050) for ($\lambda_0, \lambda_{\text{rural}}, \lambda_{\text{pov}}$), and an oracle estimator that observes each child's block group recovers (-0.354, -0.105, -0.045). The rank-plus-exposure design used in the paper recovers (-0.349, -0.097, -0.057), with Monte Carlo standard errors of 0.008, 0.012, and 0.007, whereas assigning each child to the ZIP centroid gives (-0.211, $+0.019$, $+0.004$). All coefficients from the rank-plus-exposure design are within about one Monte Carlo standard error of the truth, and its rural increment is negative in 100 percent of replications, compared with 23 percent for the ZIP-centroid estimate. This improvement comes at a cost in precision, particularly for the rural coefficient, whose sampling dispersion is 3.2 times the oracle estimator's, compared with 1.6 for poverty.

The exercise has seven limitations. The synthetic model has neither an outside option nor a nest, its mean utilities are known rather than estimated, and there is no weighting matrix. Since distance enters linearly, the exercise does not assess the functional form, and binary rurality simplifies the measurement problem. Moreover, the synthetic samples are smaller than the estimation frame, and the geography matches menu size and exposure spread rather than Puerto Rico's map. The relative dispersions describe the precision cost within this simulation; neither these ratios nor the absolute dispersions replace standard errors for the empirical estimates. The exercise addresses measurement error in residence and does not address the ecological identification of the rural increment, which I examine next.

C.8 The rural increment: three diagnostics

I identify the rural increment using variation across closure catchments in kernel-weighted rural exposure, where the distance-decay kernel approximates the rural composition of a catchment without assigning a residential block group to each pupil. This comparison is between catchments rather than households with observed rural indicators, so it may reflect other catchment differences or sensitivity to the distance transformation. I examine the latter concern with linear and logarithmic specifications. An observed student-level rural indicator is unavailable because the micro frame has no student residence tag, and the Monte Carlo in Appendix C.7 shows how ZIP-centroid imputation would affect the sign.

I use three diagnostics to examine this comparison. First, the algebra of the poverty and rurality contrasts implies that the ratio of their standard errors should lie between 2.8 and 3.4, compared with the observed ratio of 3.8. Since the poverty contrast compares agents within a market and origin, its Jacobian contains the full behavioral response, whereas the rural contrast's Jacobian contains that response multiplied by a covariance attenuator. The observed precision pattern is within about 15 percent of the algebraic prediction. Second, most identifying variation is between municipalities. Recomputing the rank-one rural-minus-urban contrast within municipality gives $+0.0262$, compared with a pooled $+0.0669$, for a ratio of 0.39 across the 51 municipalities with mass on both sides and effective samples of 10,839 urban and 9,933 rural observations in the poor cells. The sign remains the same and the magnitude is about two-fifths as large; the non-poor comparison uses only four municipalities and is not informative. I treat the pooled figure as the main descriptive contrast, since the within-municipality estimate uses a different source of identifying variation and is not an established lower bound. The bandwidth of the exposure kernel remains an open issue. At a five-kilometer bandwidth, the exposure measure has mean 0.439 and standard deviation 0.275, compared with $(0.498, 0.277)$ in the baseline demand specification, and it correlates 0.92 with the baseline values, placing 18.2 percent of its variance within municipalities across the 64 municipalities with meaningful spread. Third, the magnitude depends on functional form: the rural-to-urban ratio of marginal disutility is 2.4 under $d^{0.8}$ and 1.8 under $\log(1+d)$, although the sign is stable in every structural run.

Restricting the menu to schools in the same municipality can remove nearby alternatives for students living close to municipal boundaries, an omission that is correlated with distance and may attenuate estimated $|\lambda|$. However, neither the coefficient bias nor its effect on access costs has a general sign. Out-commuting is 11.65 percent among rural children and 13.82 percent among urban children, so the omission may also affect the groups differently. Transport constraints provide another possible interpretation of the rural increment, since a distant school that is unreachable without a vehicle or scheduled service can produce greater estimated distance sensitivity among rural families without a difference in tastes. In that case, the coefficient partly measures constraints, and its implications for incidence depend on the transport alternatives available. Block-group vehicle availability is the missing input needed to distinguish the two explanations.

C.9 Market Size and Public-School Participation

I construct market size from the total resident school-age population in each municipality-by-grade-by-year cell by aggregating Puerto Rico Community Survey block groups. Since agent weights sum to one within each market, I weight the agent frame by market size when aggregating shares across markets; otherwise the calculation gives equal weight to each market. The population-weighted model inside share is 0.5481, compared with 0.7645 for $P(\text{public} \mid \text{population})$ in the same census extract, a difference of 21.6 percentage points. Of this gap, 7.8 points concern children attending schools across a municipal boundary, which the same-municipality filter removes from the choice set (12.46 percent of rows), and another 13.8 points reflect population counts that exceed realized enrollment. The discrepancy increases across the size distribution.

The discrepancy affects how I interpret the outside option and may also affect comparisons across types, although aggregate inside shares do not identify a common adjustment to the mean utilities of a heterogeneous model. I cannot remove this discrepancy within the estimation frame because the cross-border rows do not record municipality of residence, which prevents me from subtracting out-commuters from the market-size denominator. Leaving schooling accounts for only part of the outside option. Non-enrollment is 1.76 percent among the income-poor and 1.79 percent among the income-non-poor, compared with an overall outside option of 0.0870 for the income-poor and 0.4257 for the income-non-poor, so the full poor/non-poor gap is private school enrollment, 0.0694 compared with 0.4078. For a separate composition check, I use a different extract. The model places 46 percent of inside enrollment in rural block groups, compared with a census benchmark of 47.3 percent for public-enrolled children in grades one to eight; private enrollment is more urban, at 38.2 percent rural, consistent with the pattern predicted by the affordability mechanism.

C.10 An auxiliary travel calculation for teacher quality

In Section 5.2, I distinguish the structural common-quality coefficient β_q from the projection coefficient b_q, which uses mean utilities recovered from the heterogeneous demand model. Here I use a homogeneous nested-logit index for an auxiliary calculation, suppressing the market index within each market:

$$\widetilde{\delta}_j = (1-\rho)\ln s_j + \rho \ln s_h - \ln s_0, \qquad s_h = \sum_{j \in \mathscr{J}_m} s_j, \quad s_0 = 1 - s_h. \tag{C.7}$$

The closed form recovers mean utility from homogeneous nested-logit shares, whereas the heterogeneous baseline model requires numerical inversion. At the calibrated ρ, I regress the auxiliary index on q, cluster the standard errors, and drop imputed-quality rows (5.28 percent), obtaining a slope of $+0.002089$ (0.002539), $t = 0.82$, with market fixed effects (3,744 groups), and $+0.042199$ (0.056736), $t = 0.74$, with region-by-year fixed effects (42 groups). The sample contains $N = 34{,}780$ observations and 78 clusters. I report the market-fixed-effect specification, whose fixed effects absorb $\rho \ln s_h - \ln s_0$ but leave dependence on ρ:

$$\widetilde{\delta}_j - \overline{\widetilde{\delta}}_m = (1-\rho)(\ln s_j - \overline{\ln s}_m).$$

The slope is therefore scaled by $1-\rho$, which cancels in a quality–distance tradeoff only if both coefficients have the same normalization.

An upper-bound interpretation needs further restrictions. Let Q be the intended latent counterpart of lagged teacher quality and $\hat{q}$ its measured value, written as q when I suppress hats in the demand analysis. Define measurement error as $\nu = \hat{q} - Q$ and structural mean utility as $\delta^* = \beta_q Q + X'\Gamma + \xi_{\text{aux}}$. Here, X contains the auxiliary regression's controls, with coefficients Γ, and the remainder ξ_{aux} includes the unobserved component of school demand along with any components of the baseline school controls and region-year effects that X does not absorb. The coefficient β_q retains its structural interpretation from Section 5.2. I use $a = \widetilde{\delta} - \delta^*$ for the difference between the auxiliary index and structural mean utility, and superscript $\perp$ for residualization on the auxiliary regression's controls using its sample and weights. The population projection slope $\widetilde{b}_q$ comes from regressing $\widetilde{\delta}$ on $\hat{q}$ and those controls; it differs from b_q because both the auxiliary index and the regression specification differ from the baseline. The following measurement and sign restrictions are sufficient:

$$\begin{aligned} &\operatorname{Cov}(\nu^{\perp}, Q^{\perp}) = \operatorname{Cov}(\nu^{\perp}, \xi_{\text{aux}}^{\perp}) = \operatorname{Cov}(\hat{q}^{\perp}, a^{\perp}) = 0, \\ &\operatorname{Cov}(Q^{\perp}, \xi_{\text{aux}}^{\perp}) \geq 0. \end{aligned} \tag{C.8}$$

Let $r_X = \operatorname{Var}(Q^{\perp}) / \operatorname{Var}(\hat{q}^{\perp}) > 0$ be reliability after residualization. The population slope then satisfies

$$\widetilde{b}_q = r_X \beta_q + \frac{\operatorname{Cov}(Q^{\perp}, \xi_{\text{aux}}^{\perp})}{\operatorname{Var}(\hat{q}^{\perp})}, \qquad \beta_q \leq \frac{\widetilde{b}_q}{r_X}. \tag{C.9}$$

Classical measurement error in quality does not establish the restriction on a, since even under classical error a homogeneous share index need not recover mean utility from the heterogeneous choice model. Using estimated quality in nonlinear share inversion also requires restrictions on how mean utility is recovered.

For a utility change associated with one standard deviation of true quality, define $r_0 = \operatorname{Var}(Q) / \operatorname{Var}(\hat{q})$ in the same sample and with the same weights, before residualization. Then

$$\beta_q \operatorname{sd}(Q) \leq E^{\text{ub}} \equiv \frac{\widetilde{b}_q \operatorname{sd}(\hat{q}) \sqrt{r_0}}{r_X}. \tag{C.10}$$

The inequality becomes an equality if residualized true quality and omitted determinants of school utility also have zero covariance. If $r_X = r_0 = \text{rel}$, the right-hand side is $\widetilde{b}_q \operatorname{sd}(\hat{q}) / \sqrt{\text{rel}}$, and the auxiliary sensitivity analysis uses this common-reliability convention with $\text{rel} \in [0.125, 0.25]$ and a central value near 0.18. However, the calibration does not establish reliability after controlling for the regression covariates, equality of r_X and r_0, or the restrictions on omitted determinants of school utility and recovered mean utility, so I interpret the exercise as a sensitivity calculation. It does not establish an upper bound for the estimated demand system.

To express a utility increment E in kilometers, I solve $\lambda_\tau \left[(d+\Delta)^{0.8} - d^{0.8}\right] = -E$ at the mean displaced distance, with $\lambda_\tau < 0$ and both terms on the same utility scale. The smallest $|\lambda_\tau|$ gives the largest distance equivalent. Using $|\lambda| \in \{0.06, 0.10, 0.18, 0.35\}$, $\text{rel} = 0.125$, and the market-fixed-effect slope in the auxiliary calculation gives a Δ that ranges from a few meters at the steepest distance coefficient to about 45 meters at the weakest. Using the upper endpoint of the 95 percent interval for the slope raises the weakest case to about 150 meters. This endpoint combines the upper interval endpoint, the smallest $|\lambda|$ on the grid, and the lowest reliability in the reported range; the point figures are 15 to 45 meters. Since the grid does not include the baseline urban non-poor coefficient, $|\lambda_0| = 0.042877$, which is smaller than 0.06, the reported endpoint is not an upper bound over the baseline household types. Interpreting the calculation as structural willingness to travel would also require a comparison using the mean utilities recovered from the heterogeneous model.

To describe the units of the baseline variables, I report their dispersion within markets: mean utility has a standard deviation of 0.1158 utils, and lagged quality has a standard deviation of 0.2201. The within-market correlation between δ and q is $+0.2064$, so measured teacher quality accounts for about 4.3 percent of the within-market variance in mean utility, with the remaining 96 percent concerning other observed characteristics and the unobserved component of school demand.

D Incidence and Fiscal Accounting: Details

The calculations behind Tables 9, D4, D3, D2, D1 and 11 are described here, beginning with the welfare measure and the reporting populations, then walking access and the monetary valuation of additional travel. I also explain how I construct the fiscal components and where the available data limit this account.

D.1 The welfare object and the reporting populations

For household i of type $\tau(i) \in \{$urban non-poor, urban poor, rural non-poor, rural poor$\}$, suppress the market and time subscripts and write the systematic utility as

$$\begin{aligned} v_{ij} &= \delta_j + \lambda_{\tau(i)} d_{ij}^{0.8} + (\beta_{\text{rural}} R_i + \beta_{\text{pov}} P_i) q_{j,t-1} + \pi_{\text{pov}} I_i, \\ \lambda_{\tau(i)} &= \lambda_0 + \lambda_{\text{rural}} R_i + \lambda_{\text{pov}} P_i. \end{aligned} \tag{D.1}$$

Here δ_j is the mean utility recovered by inversion in Section 5.2, and the specification is the six-parameter model used for the baseline incidence calculation, with the common quality term already included in δ_j. Using expected maximum utility from (7), I calculate the access loss as $L_i(C) = W_i(\mathscr{O}^{\text{pre}}) - W_i(\mathscr{O}^{\text{pre}} \setminus C)$, holding surviving-school utilities, distances, weights and parameters fixed. The welfare change reported in

Table D1. Government Projections of Closure Savings

Date	Schools	Figure	What the statement says, and what it covers
11 May 2014	≈100	\$27 m	A point estimate: spending would be reduced by \$27 million, with no dismissals and no stated horizon.
23 May 2017	≈180	\$7 m	A floor: savings of *more than* \$7 million, described as essentially utility costs; a contemporaneous report gives \$7–10 million over 179 schools.
5 April 2018	283	\$150 m	An approximate target inside a fiscal plan: the Department *aims to save some* \$150 million through lower operating costs and attrition, with no layoffs and teachers reassigned.

Notes: I verify all three statements verbatim against the wire and newspaper sources published on the stated dates. The statements give a point estimate, a floor, and an approximate target for different cost categories. None specifies the time horizon. They therefore cannot be compared directly or used to calculate a common per-school estimate. The 2017 statement is consistent with reported electricity spending. The Department paid \$53 million for electricity in the 2017–18 fiscal year across about 1,100 schools, or roughly \$48,000 per school-year. Applying that rate to 179 closures gives about \$8.6 million, within the reported \$7–10 million range. Plant operation and maintenance also includes maintenance, custodial services, grounds, and security. The electricity figure therefore cannot be compared directly with the plant component in Table 11. Two further public records describe what happened to the buildings after closure. In a 2020 field study by the Othering and Belonging Institute and the Centro para la Reconstrucción del Hábitat, which visited a random sample of buildings, 69 percent of the 119 closed schools in the sample were vacant. A Comptroller audit of eight closed schools in one region found electricity or water still connected for between 61 and 426 days after closure, at a cost of \$167,747 over a fourteen-month billing window. Both findings apply to the samples studied. I report their sample sizes rather than treat them as population estimates.

Table D2. Travel Costs per School Closure

	Present value
Median	\$37,103
Mean	\$65,659
Ninetieth percentile	\$178,314
Minimum	−\$76,247
Closed schools	334

Notes: The table uses 334 pre-María closed schools and the central valuation assumptions in the notes to Table D3. Each observation is a closed school. I calculate the monetary value of the change in expected travel when that school is removed from the preclosure choice set and all other schools remain available. It differs from the nonnegative logsum loss in (7). The distribution is highly skewed and includes negative values. Some closures *reduced* expected travel because the displaced children's next-best alternatives were closer than the closed school. The sum of the travel costs from removing schools individually can be greater or smaller than the cost of removing them jointly. I therefore do not sum school-level figures across a market. I do not report a break-even share or divide these travel costs by the fiscal account in Table 11, for the reasons given in Section 7.3. Table 6 reports the separate comparison of closed schools with local survivors using estimated school effects and access-cost ranks.

Table D3. Monetized Travel Changes after School Closures

	Urban non-poor	Urban poor	Rural non-poor	Rural poor
Present value per displaced student	\$135	\$253	\$1,026	\$1,141
Sensitivity				
Income relative to urban non-poor	1.000	0.989	0.868	0.857
Range across the sensitivity grid	2.2× (rural poor: \$285 to \$943)			
Change in the rural-poor to urban-non-poor ratio	0.9%			

Notes: The table uses the displaced-equivalent student population, Population B in Table 9, for 334 pre-María closed schools. I value the model-implied change in expected one-way commuting distance and time, then discount these costs over the student's remaining grade window. I assume a parent escorts the child on four legs a day over 180 school days. I value travel time at 50 percent of hourly household income, use vehicle operating costs of \$0.11 per kilometer, and apply a 3 percent discount rate. The income measure is the agent-weighted mean of block-group median household income across all markets: \$20,455 in 2017 dollars. Dividing by 2,080 hours gives \$9.83 an hour and a value of travel time of \$4.92. Block-group margins of error are about 35 percent, so I do not assign these income values to individual agents. In the primary specification, I use one pooled value of time to value travel time equally across income groups, following transport-appraisal practice. The row using a separate value for each type is a sensitivity check. This alternative reduces the rural-poor to urban-non-poor ratio by construction. I assess sensitivity by varying the value-of-time fraction over $\{0.35, 0.50, 0.65\}$, the vehicle operating cost over $\{0.08, 0.11, 0.15\}$ and the discount rate over $\{0.03, 0.05, 0.07\}$. Across these combinations, the levels change by a factor of 2.2, while the ratio between the extreme types changes by under 1 percent. The valuation assumptions therefore affect the levels more than the relative burdens. This sensitivity exercise holds the estimated parameters fixed and does not establish robustness to parameter uncertainty. This calculation values additional travel for families that already drive. It may understate costs for households that must change transport modes, and it does not place a monetary value on the full logsum access loss. In particular, the share of rural-poor displaced-equivalent children choosing a school within walking range falls by 20.8 percentage points, so reaching schools beyond that range may require motorized transport (Table D4); the data do not record vehicle availability by block group. The calculation measures the burden shifted onto households without subtracting the government's transport savings. It therefore describes incidence and cannot be interpreted as a social cost-benefit calculation. I keep it separate from the fiscal account in Table 11. I report costs separately by household type, without aggregating them.

Table D4. School Access within 1.6 Kilometers

	Urban non-poor	Urban poor	Rural non-poor	Rural poor
Share within 1.6 km, before	0.213	0.256	0.253	0.267
Share within 1.6 km, after	0.170	0.185	0.065	0.060
Change (percentage points)	−4.3	−7.1	−18.8	−20.8
Reduction relative to the pre-closure share (%)	20	28	74	78

Notes: The unit of observation is an agent–market cell in the displaced-equivalent population. I use the same 992 markets, 122,350 agent rows, and 334 pre-María closed schools as Table 9. The shares are model-implied point estimates. I calculate them from choice probabilities over the pre-closure and post-closure menus, weighting by the probability of having attended a closing school. The 1.6-kilometer threshold is a round radius, not a policy threshold. The ordering of the four cells does not depend on this radius. Because the quantity is a share of children relative to a fixed radius, it requires no value of time, discount rate, or earnings assumption. The four pairs use the distance-only welfare evaluation in Appendix D: the three distance parameters in Table 5, comprising the base coefficient and its rural and poverty increments. The calculation excludes type-specific outside-option and quality shifters.

Table 9 uses the opposite sign: $\Delta W_i(C) = -L_i(C) \leq 0$, following the discrete-choice welfare measure of Small and Rosen (1981). Although the factor $1/(1-\rho) = 21.5$ rescales systematic utility inside the nest, it does not multiply plain-logit welfare losses to produce nested-logit losses, because those losses also depend on the utilities and on which options are removed through (7).

For Population A, I give each agent weight $\tilde{w}_i = w_i M_m$, multiplying the integration weight by the market's resident school-age population. Population B is the displaced-equivalent population, for which I use $\tilde{w}_i s_{i,C}^{\text{pre}}$, multiplying the population weight again by the model-implied probability of attending a closing school before closure, $s_{i,C}^{\text{pre}} = \sum_{j \in C} p_{ij}^{\text{pre}}$. The integration weights satisfy $\sum_{i \in m} w_i = 1$ in each market m. For a group g the mass share is $\sum_{i \in g} h_i / \sum_i h_i$ and the loss share is $\sum_{i \in g} h_i L_i(C) / \sum_i h_i L_i(C)$ under the same weights h. Thus Population B is an ex ante calculation weighted by attendance probabilities, rather than a count of observed displaced pupils or a measure of their losses conditional on realized displacement. The calculation covers 992 closure markets of 3,120, 122,350 agent records of 285,800, and 334 closed schools, with no market in which every school closes.

In Table 9, I report standard errors for the aggregate statistics in the second panel, which summarize the distribution of access losses, while the individual cells contain point estimates. The levels also depend on the calibrated nesting parameter $\rho = 0.953524$, which I hold fixed at the value from the earlier estimation using the target for exit from the local public-school choice set, without re-estimating it in the baseline run. As described in Appendix C.3, the target combines estimated public-school exit with an allowance for cross-municipality switching. Across $\rho \in \{0.9515, 0.953524, 0.9555\}$, the mean-loss ratio changes by about 4 percent. The changes in the loss share between adjacent grid points are positive, between $+0.0017$ and $+0.0067$, so a higher nesting parameter makes the distribution of losses slightly more regressive. I computed this grid and the walk-zone shares below using only the three distance parameters in Table 5: the base coefficient and its rural and poverty increments. These calculations therefore differ from the six-parameter specification in Table 9, since they omit the type-specific inside-option and quality shifters.

D.2 The walk zone

Table D4 reports the model-implied share of displaced-equivalent students whose chosen school lies within 1.6 road kilometers. For household i and open set $\mathscr{O}$, the probability is $a_i(\mathscr{O}) = \sum_{j\in\mathscr{O}} p_{ij}(\mathscr{O})\mathbf{1}\{d_{ij} \leq 1.6\}$. The reported group share is $\sum_{i\in g} h_i a_i(\mathscr{O})/\sum_{i\in g} h_i$, using the same preclosure Population-B weights before and after closure. After closure, this share falls by 78 percent for rural-poor displaced-equivalent children and 74 percent for rural non-poor children. The corresponding decreases are 20 percent for urban non-poor and 28 percent for urban poor children. The four cells start with similar shares, between 0.213 and 0.267, so unequal starting points do not account for the share lost. I use 1.6 kilometers as a round radius rather than a policy threshold, and the ordering does not depend on this choice. For the four before-and-after pairs, I use the choice probabilities from the specification with only the three distance parameters in Table 5: the base coefficient and its rural and poverty increments, leaving out the type-specific inside-option and quality shifters. I report the shares without standard errors. The calculation concerns the share of children choosing a school within a fixed radius and therefore requires no assumption about the value of time, discount rates, or earnings.

D.3 Valuing additional travel

In Table D3, I place a monetary value on the model-implied change in expected travel using transport-appraisal conventions. This does not monetize the logsum access loss, which also reflects school attributes and the value of having alternatives. I value time at 50 percent of an hourly wage derived from the agent-weighted mean of block-group median household income, \$20,455 in 2017 dollars, corresponding to a wage of \$9.83 an hour and a value of time of \$4.92. I use a marginal vehicle operating cost of \$0.11 per kilometer, 180 school days, a parent escorting the child on four legs a day, and a 3 percent discount rate. The horizon is the child's remaining grades-3–8 window. Under these assumptions, the present value of additional travel after closure is about \$1,141 per displaced-equivalent rural-poor child and about \$135 per urban non-poor child. The difference is roughly \$1,000. I report the levels and their difference, omitting their ratio because it is imprecisely estimated. This imprecision comes from the base distance coefficient rather than the valuation assumptions. The income measure is a median within each block group, and I take its mean across agents. Block-group margins of error are approximately 35 percent, so I do not assign these income values to individual agents.

The calculation values additional kilometers for a family that already drives, which may understate the burden for a family that has to change transport modes, although it does not provide an unconditional lower bound on logsum access losses. For a rural-poor family that must travel beyond walking range after closure, escorted travel adds 20.3 minutes each way, or 244 hours a year. The same federal guidance values this time at 100 percent of the wage rather than 50. The implied cost is roughly \$2,398 a year, compared with \$255 for a family that already drives, or about nine times as much. The model-implied share of rural-poor displaced-equivalent children choosing a school within walking range falls by 20.8 percentage points.

Reaching schools beyond that range may require motorized transport that their households may not have, and vehicle availability by block group is not observed. For these families, the rural distance coefficient may partly reflect transport constraints rather than tastes, and evaluating the welfare consequences requires knowing which transport modes are feasible.

The valuation assumptions affect the levels more than the relative burdens. I compare escorted and unescorted travel, with the value of time in $\{0.35, 0.50, 0.65\}$ of the hourly wage, vehicle cost in $\{0.08, 0.11, 0.15\}$ dollars per kilometer and discount rate in $\{0.03, 0.05, 0.07\}$. Across these combinations, the rural-poor cost changes by a factor of 2.2, while the comparison across types changes by under one percent. I evaluated the grid using the specification with only distance parameters described above, so it shows sensitivity to the valuation assumptions; the levels in Table D3 come from the full specification. The relative burdens are more sensitive to the estimated parameters than to these valuation assumptions. The comparison across types is much less precisely estimated than the ordering of the levels, so I report only the levels and their difference. Federal guidance sets a maximum value-of-time fraction of 0.60 for local personal travel. The 0.65 value in the top row of the grid therefore exceeds this guidance. Valuing each type's time at its own income attenuates the gap by construction and does not change the ordering; the primary calculation uses a single pooled value, as standard transport appraisal does, to assign the same value to travel time across income groups.

I keep the levels separate for each type. For the transport comparison in Section 7, the displaced-equivalent count is 24,451, the model-implied mass of the six-parameter run. The dollar figure measures the burden on families without subtracting the territory's transport savings, so it describes incidence and cannot be interpreted as a social cost-benefit calculation.

D.4 Travel costs of individual school closures

For Table D2, I remove each closed school individually from the preclosure choice set, retain the other schools, and monetize the resulting change in expected travel. Unlike the logsum loss $L_i(C)$, which cannot be negative when alternatives are removed with all other inputs fixed, the travel cost can be negative because a closure may reduce expected travel. The distribution is highly skewed, with a negative minimum: few children living near the schools whose removal reduced travel had been choosing them. On single-closure markets the individual-removal calculation reproduces the joint run exactly. With multiple closures, individual-removal travel costs need not add up to the joint cost, and there is no general ordering between the two.

The table shows the variation in travel costs across schools, but I do not calculate the share of closures whose travel cost exceeded a savings benchmark, nor do I report a marginal value of public funds, for the reasons given in Section 7.3. The framework of Hendren and Sprung-Keyser (2020) permits comparisons of policy benefits and net fiscal costs, including for policies that generate fiscal savings, such as closures. Applying it here would require a numerator measuring the monetary value of aggregate access loss, which is not reported, and a denominator represented by a partial-identification interval. I therefore report the two accounts separately.

D.5 The three official savings announcements

I reproduce the three public statements and their sources in Table D1, after verifying their wording verbatim. One statement gives a point estimate, another a floor, and the other an approximate target, for different cost categories and without a specified time horizon, preventing a direct comparison either across statements or with annual facility costs. School counts also differ across sources for the 2017 round (179 and 184) and for the 2018 round (259, 262, 281 and 283 across conventions and universes). The 2017 claim is consistent with utility savings. The Department paid $53 million for electricity in fiscal 2017–18 across about 1,100 schools, or roughly $48,000 per school-year. Applying this rate to 179 closures gives about $8.6 million. Utility payments were $53 million, $47 million and $40 million in fiscal 2017–18, 2018–19 and 2019–20.

D.6 The salary file and its limits

The administrative salary file covers 867,338 employee-years and 71,518 people between 2010 and 2024. Its coverage is the whole departmental workforce, 71,518 people in 2024 compared with a teaching workforce of about 30,000, but only employees observed at the end of the period enter the file. The year-on-year exit rate is 0.00 percent in all 14 adjacent-year pairs, and every person's last observed year is 2024. At least 29 percent of the 2013 workforce is therefore missing, and this selection is correlated with closure exposure. Since the modal monthly salary points apply to everyone on the same schedule, I can recover pay rates from this file, although I cannot use it to measure the workforce, total payroll, or changes in payroll. The crosswalk from the teaching roster to the salary identifiers covers 69.8 percent of teacher-school-year cells, from an 84.3 percent identifier match on 59,345 cells over 15,574 teachers and 1,291 schools, and its coverage declines from 75.5 percent in 2013 to 64.0 percent in 2019. Every pay statistic in Section 7 therefore holds among staff observed through 2024.

D.7 Constructing the fiscal components

To construct the plant component in Table 11, I use a facilities-spending relationship of $18,657 plus $121.16 per student, fitted on 127 surviving schools, which predicts $35,498 at the median closing school. From this amount I subtract mothballing costs of $6,460, derived from legislative testimony reporting roughly $1 million a year across 113 disused schools, or about $8,850 each, on 17 March 2017. I adjust this amount by a 27 percent disposal rate implied by a utility letter classifying 201 accounts as 55 disconnected, 39 transferred and 70 active. Under an alternative assumption, closure saves only the fitted intercept because the per-student expenditure follows the child. The fit is a cross-section on surviving schools that observes no closed school, and a cross-sectional relationship between school size and facilities spending need not equal the expenditure response when a network contracts, which is one reason the plant term enters Table 11 as a band rather than a point.

To estimate the staffing component, I use the 2016–2018 cohorts, which contain 467 of the 610 closures and are the only cohorts observed in the staff roster. Excess exits total 496 under the clipped calculation and 403 under the netted calculation, or 1.06 and 0.86 per school. I value these exits at the median wage of exposed staff, $2,475 a month. However, the data do not identify the refill rate, the flow test fails its lead placebo, and the absorption interval is an order of magnitude wider than the effect it would need to detect, so I set the lower bound of the staffing component to zero. The extrapolation from those cohorts to all 610 schools is an assumption, and both components must be included when comparing policies.

For transport, the reduction in bus routes needed to produce the proposed route-based offset exceeds the island's entire number of active routes. There is also a 2014 statute that capped departmental transport spending through price and contract terms between June 2014 and July 2017 without reference to closures. Transport spending could therefore fall without a change in the network. In the one region with audits before and after closure, contracted transport spending roughly doubled, from $10.8 million to $20.5 million. I do not estimate the transport term. A reduction in public transport expenditure need not represent a resource saving if the same journeys are instead financed by households.

To assess one-time costs, I use a Comptroller audit that visited eight closed schools in one region and found electricity or water still connected for between 61 and 426 days after closure, at a cost of $167,747 over a fourteen-month billing window. The audit covers eight of the 84 schools consolidated in the region, so its per-school figures are averages over the schools visited. Audited regional inventories elsewhere imply roughly $94,000 to $121,000 of stranded movable property per closed school; because this is the registered asset value rather than a payment at closure, I exclude it from the account and do not use it to calculate a payback period.