The Language of Closure: Examining Racial Differences in How A Community Discusses School Closure Metrics

Michael L. Chrzan
University of Maryland

Francis A. Pearman
Stanford University

School closures in urban districts disproportionately affect marginalized communities, yet community input often goes unanalyzed or is reduced to simple frequency counts. This study applies BERTopic, a neural topic modeling approach, to analyze 4,159 suggestions from 2,006 community members regarding school closure metrics in a large urban district. Through extensive hyperparameter tuning across 62 configurations, we identified 14 coherent topics that capture community priorities. Chi-square analysis revealed substantial variation in topic prioritization by race ($\chi2 = 152.0825$, $p < 0.0001$, $V = 0.1439$). Furthermore, an analysis of topic outliers revealed that White respondents were significantly more likely to provide suggestions that fell outside of community-wide themes ($z = 2.14$). These findings demonstrate that "neutral" community engagement processes may obscure the specific concerns of marginalized groups, and highlight the utility of computational methods in surfacing rigorous insights from large-scale text data.

The Language of Closure

Examining Racial Differences in How A Community Discusses School Closure Metrics

Michael Leon Chrzan[1]
Francis A. Pearman[2]

[1]Center for Educational Data Science and Innovation, University of Maryland
[2]Stanford University

March 1, 2026

Abstract

School closures in urban districts disproportionately affect marginalized communities, yet community input often goes unanalyzed or is reduced to simple frequency counts. This study applies BERTopic, a neural topic modeling approach, to analyze 4,159 suggestions from 2,006 community members regarding school closure metrics in a large urban district. Through extensive hyperparameter tuning across 62 configurations, we identified 14 coherent topics that capture community priorities. Chi-square analysis revealed substantial variation in topic prioritization by race ($\chi^2 = 152.0825$, $p < 0.0001$, $V = 0.1439$). Furthermore, an analysis of topic outliers revealed that White respondents were significantly more likely to provide suggestions that fell outside of community-wide themes ($z = 2.14$). These findings demonstrate that "neutral" community engagement processes may obscure the specific concerns of marginalized groups, and highlight the utility of computational methods in surfacing rigorous insights from large-scale text data.

Keywords: *School Closures, Racial Equity, Natural Language Processing (NLP), BERTopic, Educational Governance, Urban Education*

1 Introduction

The phenomenon of closing public schools in major urban centers constitutes one of the most contentious and racially fraught areas of modern education policy. District leaders often frame school closures through a technocratic lens, utilizing terms such as "right-sizing," "portfolio management," or "efficiency optimization" [Eddins et al., 2024]. This narrative posits that declining enrollments - driven by demographic shifts, lower birth rates, and the migration of families to suburbs or charter schools - render certain facilities financially unsustainable. From this perspective, closing a school is a fiscal imperative designed to redirect resources to better-performing institutions.

Empirical research overwhelmingly demonstrates that race is a primary predictor of school closures, often exceeding the explanatory power of academic performance or enrollment decline. A prior analysis by researchers at Stanford University found that majority-Black schools are more than three times as likely to be permanently closed as schools with other demographic profiles [Pearman and Greene, 2022]. This disparity remains statistically significant even when controlling for seemingly race-neutral factors such as standardized test scores, facility conditions, and enrollment trends.

These findings suggest that "neutral" policy criteria often function as proxies for race. For instance, a school in a gentrifying neighborhood might be closed due to "low enrollment" just before the land is repurposed for high-end development, while a similarly under-enrolled school in a stable white neighborhood is given a grace period to recover.

This paper analyzes a piece of the closure process from one such case occurring recently in the

US. For anonymity purposes, we refer to this district as Local United School District (LUSD), which is the traditional school district in the city and is facing the reality that they need to reorganize their school portfolio to better manage their funds to provide student services. As a part of this process, the district has gathered community input to decide the metrics they would use to identify schools for reorganization. In their communications with the community, the district noted that a school being identified for reorganization can result in things besides closure, such as a merger or co-location of schools.

The community input occurred in two pieces: town halls and surveys where all members of the community (parents, staff, students, and even general community members unaffiliated with a school) could voice what metrics they felt were important for identifying schools for reorganization. This study will use Natural Language Processing (NLP) techniques to look at the data gathered from the first survey to answer the question:

> How does the distribution of school reorganization themes in open-ended survey responses differ across racial groups?

We hypothesized that the suggestions for school reorganization metrics will differ across racial groups. We know from prior literature that the closure process has differential impacts on different communities and thus we assume that these communities concerns at the beginning of a closure process will speak to historical patterns they are likely aware of relating to school closures. By examining these differences, LUSD - and future districts - can better understand the needs and priorities of different stakeholder groups and use this information to create a more effective district reorganization process.

Thus, our primary contribution is to provide valuable insights into how different racial groups in a diverse U.S. school community view the school reorganization process and what metrics they believe are important. By understanding these differences and how to surface and analyze them, districts like LUSD can better take into account the needs and priorities of all community members. This, in turn, can help to build trust and collaboration between the district and the community and lead to better outcomes for all students, a crucial need given the rising rates of school closures across the United States, particularly in urban districts [Chrzan et al., 2025]. We also fill a gap in the literature in combining powerful, modern topic modeling techniques and community perceptions of school closures.

2 Related Work

2.1 School Closures and Racial Equity

Prior research into the community response to school closures has focused on the social costs of school closures and how communities resist them. For instance, Ewing researched large-scale closures in Chicago's Bronzeville district, finding that schools are not just places of learning but also function as historical landmarks and institutions that shape community identity [Ewing, 2018]. Ewing described "institutional mourning," or the social and emotional experience individuals and communities undergo when facing the loss of a shared institution, particularly when those communities are marginalized and rely heavily on that institution. Other scholars have found similar results in studies of community push-back against school closure initiatives within major cities such as Washington D.C. [Syeed, 2019], Philadelphia, Oakland, and Detroit [Green, 2017].

While much research has been done discussing the process of school closures typically employed by districts to identify schools for closure and the community's push-back against those processes, very few of those processes actively engage the community in determining the metrics themselves, despite research urging for these equitable practices [Hahnel and Marchitello, 2023]. Yet, this research suggests that communities are not passive recipients of school closure decisions and instead show that they actively engage in the process regardless, employing various strategies to voice their concerns and advocate for their schools. Communities challenge the data, highlight the social and cultural value of their schools, and emphasize the potential negative impacts closures can have on their communities, such as those seen in Philadelphia over a decade since they're 2012 mass closures [DeOrnellas, 2023]. Thus, it appears that this research reveals a discrepancy between what the community values and what the district values in determining which schools are worth keeping open.

A growing body of scholarship also positions school closures not just as a reaction to neighborhood change, but as a catalyst for it. Schools are community anchors; their removal signals to real estate developers and affluent outsiders that a neighborhood is being "reimagined", as examined in Pearman and Greene [2022]. Research on parental engagement reveals additional complexities. As urban areas gentrify, White middle-class parents increasingly engage with public schools, often acting as economic brokers through fundraising [Posey-Maddox, 2016]. However, their involvement can exacerbate resource disparities and marginalize families of color [Freidus, 2019]. Even when committed to diversity, White parents' perspectives may create "racial equity detours" that obscure ongoing inequities [Clonan-Roy et al., 2023].

Further, research indicates that racial disparities exist in how parents perceive and recommend school closures. Nuamah [2020] - in a study focusing on Chicago - found that when asked about the effectiveness of school closures as a means to improve the school system by removing low-performing schools Black and Latino parents were more likely to oppose school closures than White parents, despite Whites being less directly impacted by these policies. The researchers attribute this difference to the disproportionate burden of school closures borne by Black and Latino communities, leading to a perception of being targeted by these policies as opposed to supported by them. Notably, other research examining online school reviews reveal similar patterns, where language used in reviews often reflects racial and socioeconomic disparities in education [Gillani et al., 2021]. For instance, terms associated with affluence and Whiteness, like "private school" and "PTA," are positively correlated with schools' test scores, suggesting that reviews may implicitly encode demographic information and reinforce existing achievement gaps.

The Gillani et. al. study is one of less than a handful of studies that the authors were able to find that used NLP techniques to examine parental/community preferences for schools. Given their findings, it is clear that more research is needed on examining this crucial data source - text - in understanding the dynamics of district governance decisions, particularly in tense situations such as school closures. In this study we work to address this gap.

2.2 NLP in Educational Research

As school districts grapple with the volume of feedback generated by these controversies - often numbering in the thousands of emails, survey responses, and public comments - they increasingly need tools to analyze large amounts of text data at scale. Natural Language Processing (NLP) methodologies offer such tools and make it possible to extract and analyze themes and sentiments from large copra. While recent literature in using NLP to analyze teaching have demonstrated significant advances [Xu et al., 2024, Demszky et al., 2025, Avitabile, Andrew et al., Bartanen et al., 2025], the application of NLP to non-instructional educational contexts remains limited but growing.

For example, computational analyses have revealed systematic biases and structural trends in instructional materials. In high school history textbooks, NLP techniques found that Latinx groups are virtually absent, appearing in less than 0.3% of mentions despite representing over half of the student population in some regions [Lucy et al., 2020]. Furthermore, this research found that word associations show that Black figures are frequently described with lower levels of agency and power, often appearing as the objects of verbs like "owned" or "barred".

In the realm of school board governance, Anglin, Kylie L. et al. analyzed over 40,000 meeting minutes and revealed that race-related discourse is relatively rare, occurring in only 13% of meetings. However, these discussions are highly responsive to national events, with affirmative statements regarding racial equity surging after the murder of George Floyd in 2020 and oppositional critiques regarding "critical race theory" peaking in mid-2021.

While Anglin, Kylie L. et al. and Gillani et al. [2021] demonstrate NLP's capacity to detect racialized discourse in meeting minutes and online reviews, these sources represent only the public-facing or final administrative layer of district governance. They often miss the raw, high-volume influx of community feedback that occurs during the decision-making process - specifically the "institutional mourning" and resistance described by Ewing [2018]. To capture the dynamics of community sentiment *during* a consolidation crisis, this study applies these computational techniques to a novel corpus of direct constituent communications. The following section details the acquisition and pro-

cessing of this dataset.

3   Data

The survey LUSD administered was designed to be aligned to LUSD's values and commitments and focus broadly on specific categories of metrics, namely "Equity", "Excellence", and "Efficiency". This first survey in the process was designed to get community input on what metrics should be used by the district under each of these categories in order to create a composite measure for the district to use in the reorganization decisions. For example, under the category of excellence, a metric one might consider using are measures of academic achievement, such as state test scores.

The survey was administered by the district to all stakeholders, including the broader community. The survey also captured demographic data on the respondents, including their role in/relationship to the district and what schools they have relationships with, as well as their ratings on each of the categories and their subcategories. The scores were on a Likert scale from 1 to 5 with higher values indicating a higher belief in the importance that the criteria be considered for school portfolio reorganization decisions. Under each of these umbrella categories, respondents could also respond to an open-ended prompt: *"Please list any additional factors of [insert category] that you suggest should be considered."*. Since this prompt was given for each category as well as once just for general input respondents could answer this prompt up to 4 times. It is the responses to this prompt for each category that will be analyzed in this study.

This survey garnered 10,135 responses in just two weeks, which demonstrates what prior research has shown that school closures are often contentious events within communities. The survey was administered after the town halls mentioned earlier were completed so respondents could ask any questions to the district directly while they completed it, and it was also available online via the district's website. Given this survey distribution method, this constitutes a convenience sample.

Due to the ease of access for the survey and the contentious nature of the topic, there was some concern that people may re-take the survey or send the survey to others unaffiliated with LUSD. For this reason, any unknown computer addresses logged multiple times in the response submission metadata were removed along with 6 responses from individuals listing zip codes that were not in proximity to the district. The sample size listed earlier reflects those removals.

While over 10,000 respondents submitted data for the Likert items mentioned previous, only $n = 2,006$ respondents completed the open-ended response portion of the survey and had usable data. Usable data here was defined as data where the response was not copy-and-pasted across multiple prompts and where the data's language could be translated using Google's Compact Language Detector (CLD) [Ooms [aut et al., 2024] and the `polyglotr` package [Iwan, 2024]. Around 11% of responses were written in non-English languages. Of these 2,006 respondents, 131 of them copied and pasted their response across multiple of the open-ended prompts and 220 of them were unable to have their language detected by CLD. After removing these duplications and undetectable language statements, we were left with 4,159 suggestions from the community members who took the survey. The demographic breakdown the 2,006 respondents can be seen in Table 1 and other information on the response data can be seen in Appendix A.

One final adjustment was made after finalizing the sample. As can be seen in Table 1, certain racial groups had much lower representation in the resulting sample for analysis. For example, Only 7 respondents identified themselves as American Indian or Alaskan Native. This is a direct result from lower response rates to the survey for these groups and not due to any data manipulation mentioned previously. These groups with small representation were combined into one group, which we call "Underserved and Small Sample". While this is less than ideal - as there are few reasons to believe American Indian or Alaskan Native, Filipino, Pacific Islander, and African American respondents would all have similar suggestions that would warrant this combination - considering all these groups are ones that have historically been underserved by LUSD, removing their responses from the analysis was deemed less useful to the goal of the study. Thus, these groups were combined for our analysis. See Figure A.1 in Appendix A for absolute counts of all disaggregated groups.

Table 1: Demographic Breakdown

Variable	Category	Count	%
Education	Associates or technical degree	72	3.6
Education	Bachelor's degree	638	31.8
Education	Decline to State	6	0.3
Education	Graduate or professional degree	912	45.4
Education	High school diploma or GED	86	4.3
Education	Prefer not to say	114	5.7
Education	Some high school or less	81	4
Role	Central office staff	6	0.3
Role	Community member	43	2.1
Role	Community partner	22	1.1
Role	Decline to State	7	0.3
Role	Multiple roles	326	16.2
Role	Parent/Caregiver	1421	70.8
Role	School site staff	157	7.8
Role	Student	26	1.3
Survey language	English	1773	88.3
Survey language	Spanish	158	7.9
Survey language	Tagalog	1	0
Survey language	Vietnamese	2	0.1
Survey language	Simplified Chinese	74	3.7
Race	African American	64	3.2
Race	American Indian or Alaskan Native	7	0.3
Race	Asian	337	16.8
Race	Decline to State	447	22.3
Race	Filipino	39	1.9
Race	Hispanic/Latino	293	14.6
Race	Pacific Islander	13	0.6
Race	Two or More Races	149	7.4
Race	White	659	32.8
Gay	Decline to state	292	14.5
Gay	No	1542	76.8
Gay	Yes	174	8.7
Child gay	Decline to state	781	38.9
Child gay	No	1016	50.6
Child gay	Not applicable	80	4
Child gay	Yes	131	6.5

4 Methods

To identify and analyze thematic patterns in community input regarding school closures, we employed a systematic approach combining neural topic modeling with rigorous hyperparameter exploration and statistical validation. Our methodology prioritized both computational rigor and interpretability, ensuring that identified topics reflected genuine semantic structure in the data while remaining accessible to stakeholders and policymakers.

4.1 BERTopic and Neural Topic Modeling

We employed BERTopic [Grootendorst, 2022], a neural topic modeling framework that leverages transformer-based embeddings and density-based clustering, to identify thematic patterns in community input regarding school closures. BERTopic was selected over traditional approaches like Latent Dirichlet Allocation (LDA) for several key advantages:

- It captures semantic relationships through contextualized embeddings rather than relying on co-occurrence statistics alone,

- It automatically determines the optimal number of topics through density-based clustering rather than requiring manual specification, and

- It has been shown to produce more interpretable and coherent topics that align better with human judgment, particularly in social scientific settings [Egger and Yu, 2022, Kaur and Wallace, 2024]

4.2 Model Preparation

Prior to modeling, we constructed a custom stop word list combining sklearn's English stop words with domain-specific terms (e.g., "school") that appeared ubiquitously but carried minimal discriminative value.

Documents were encoded using the `ll-MiniLM-L6-v2` sentence transformer model [Reimers and Gurevych, 2019, Wang et al., 2020, noa, 2024], BERTopic's default embedding model, which produces 384-dimensional contextualized embeddings. This model was selected for its balance of computational efficiency and strong performance on semantic textual similarity tasks, making it well-suited for processing large collections of community-generated text. All computations were performed on Apple Silicon using Metal Performance Shaders (MPS).

The high-dimensional embedding space was reduced using UMAP (Uniform Manifold Approximation and Projection; McInnes et al. [2020]) to preserve both local and global structure. Following dimensionality reduction, we applied HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise; McInnes et al. [2017]) to identify dense regions corresponding to coherent topics. HDBSCAN's ability to identify clusters of varying density and designate noise points (outliers) made it particularly well-suited for community-generated text, which often contains off-topic or ambiguous contributions.

4.3 Hyperparameter Decisions

We conducted an extensive grid search across 62 hyperparameter configurations, systematically varying UMAP parameters (`n_neighbors`: 10-50,
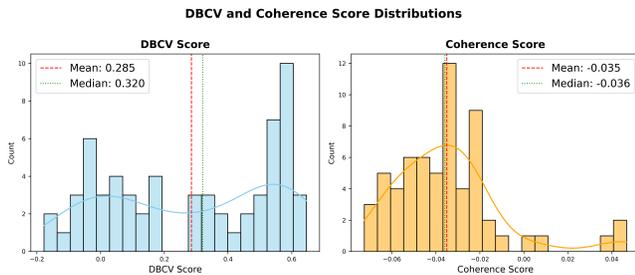
Figure 1: Evaluation Metric Distributions

Note: Hyperparameter Exploration Results. The histograms display the distribution of Density-Based Clustering Validation (DBCV) scores against Topic Coherence (NPMI) for all 62 tested hyperparameter configurations. The selected model represents a optimal trade-off, maximizing cluster separation (high DBCV) while maintaining interpretability.

`n_components`: 15-150, `min_dist`: 0.0-0.1) and HDBSCAN parameters (`min_cluster_size`: 10-50). For the `CountVectorizer`, we used an `n_gram_range` = (1, 2) and stops words described previously. We also incorporated BERTopic's built in functionality for determining the number of topics by setting the `nr_topics` parameter to "auto".

Each configuration was evaluated using two complementary metrics:

- DBCV (Density-Based Clustering Validation): A geometric measure of cluster quality that assesses both within-cluster density and between-cluster separation [Moulavi et al., 2014]. DBCV ranges from -1 to 1, with higher values indicating well-defined, separated clusters.

- Topic Coherence (NPMI): A semantic measure quantifying the co-occurrence of top topic words using Normalized Pointwise Mutual Information [Aletras and Stevenson]. Higher NPMI values indicate topics with semantically related words that frequently co-occur.

We ranked configurations separately on each metric and computed an average rank to identify parameter combinations that simultaneously optimized both structural clustering quality and semantic interpretability, as seen in Figure 1. The top-performing configurations (top 5th percentile by combined score) were manually examined by the authors for topic diversity, clarity, and similarity across their topics (as seen in Figure 2).

Lastly, as BERTopic relies on neural architectures trained on massive internet corpora, it is subject to the biases inherent in those training data, such as the under-representation of non-standard dialects or the association of certain demographics with lower-agency verbs [Garrido-Muñoz et al., 2021]. In the context of school closures, there is a specific risk that the 'coherent' clusters identified by the model might privilege 'Standard English' phrasing while relegating African American Vernacular English (AAVE) or other community-specific dialects to the 'outlier' category, effectively silencing them [Rozado, 2020, Nadeem et al., 2021]. To examine the potential of this 'algorithmic erasure,' we did not treat the generated outliers as discardable noise. Instead, we subjected the outlier category itself to rigorous statistical examination (see Section 5.4) to verify that the model's density-based exclusion did not systematically disproportionately impact marginalized racial groups.

4.4 Statistical Modeling

We examined racial disparities in topic distributions using chi-square tests of independence. To account for the nested structure of our data (multiple suggestions per respondent by survey category), we aggregated to the respondent level by assigning each respondent their most frequently occurring topic. This approach prevents pseudoreplication while retaining maximal information (respondents contributing suggestions across multiple topics were assigned to their modal topic).

We computed standardized residuals to identify which race-topic combinations showed significant over- or under-representation ($|residual| > 2$ indicating $p < 0.05$ for that cell). Effect sizes were quantified using Cramér's V, with values of 0.1, 0.3, and 0.5 corresponding to small, medium, and large effects respectively [Cohen, 1988, Cramer, 1946]. All statistical analyses were conducted in Python 3.11 using scipy.stats.

5 Results

5.1 Topic Model Selection and Quality

Our hyperparameter exploration procedure yielded a robust model that balanced structural density with semantic interpretability. As illustrated in
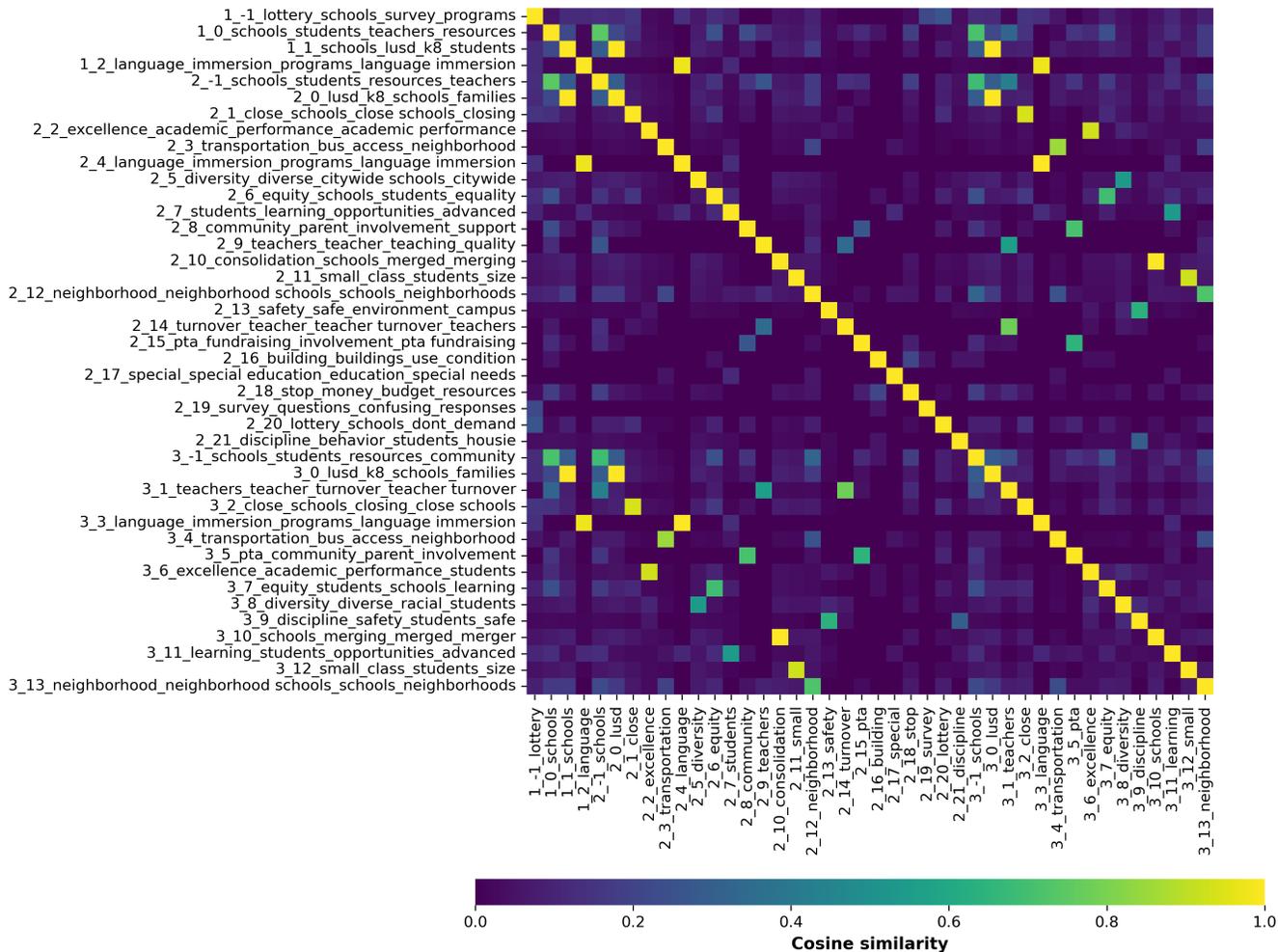
Figure 2: Semantic Similarity Between Topics within Models

Note: This heatmap visualizes the cosine similarity between the embeddings of the 14 identified topics. Darker regions indicate lower similarity. This diagnostic was used to ensure that the final topics represented distinct thematic concepts rather than duplicative clusters and to aid in model selection by examining topic similarity across the best performing models. Rows are labeled with different parts broken down by underscores as model_topic_representation.

Figure 1, the grid search revealed an inverse relationship between density-based clustering quality (DBCV) and topic coherence (NPMI) for our data. While some configurations achieved higher coherence, they often did so by collapsing distinct themes into overly broad categories. The selected model (`n_neighbors = 50`, `n_components = 75`, `min_dist = 0.1`) achieved a high DBCV score of 0.581, indicating well-separated clusters, while maintaining a coherence score of -0.018. This trade-off allowed for the emergence of 14 nuanced topics rather than fewer, more generalized ones.

Qualitative validation of these topics, assisted by the similarity matrix in Figure 2, confirmed

that the model captured distinct semantic spaces. For instance, while "Academic Performance" and "Advanced Learning Opportunities" share conceptual overlaps, the model successfully disentangled them into separate topics—one focusing on standardized metrics and the other on curricular enrichment—demonstrating the model's higher granularity.

This more nuanced topic distribution, however did come with a significant trade-off. In the best ranked model, only 113 suggestions (3%) were listed as outliers, however in our chosen model 45% of the corpus was listed as an outlier. We selected this model nonetheless because its retained clusters were

Label	#	Top Terms
K-8 schools	664	['lusd', 'k8', 'schools', 'families', 'students', 'k8 schools', 'district', 'model', 'rooftop', 'community']
Teacher Turnover	317	['teachers', 'teacher', 'turnover', 'teacher turnover', 'staff', 'resources', 'teaching', 'schools', 'district', 'need']
Alternatives to Closures	149	['close', 'schools', 'closing', 'close schools', 'dont close', 'closed', 'dont', 'students', 'children', 'closing schools']
Language Immersion	141	['language', 'immersion', 'programs', 'language immersion', 'program', 'bilingual', 'immersion programs', 'pathways', 'english', 'language programs']
Transportation	141	['transportation', 'bus', 'access', 'neighborhood', 'buses', 'schools', 'transit', 'public transportation', 'public', 'families']
PTA Inv. Funding	135	['pta', 'community', 'parent', 'involvement', 'support', 'fundraising', 'engagement', 'parents', 'participation', 'family']
Academic Perf.	132	['excellence', 'academic', 'performance', 'students', 'academic performance', 'test', 'scores', 'schools', 'student', 'important']
Equity	111	['equity', 'students', 'schools', 'learning', 'dont', 'need', 'equity equity', 'important', 'district', 'equal']
Diversity	108	['diversity', 'diverse', 'racial', 'students', 'schools', 'student', 'student body', 'citywide schools', 'body', 'citywide']
School Safety	104	['discipline', 'safety', 'students', 'safe', 'bullying', 'behavior', 'environment', 'student', 'campus', 'security']
Mergers	85	['schools', 'merging', 'merged', 'merger', 'consolidation', 'merge', 'mergers', 'students', 'think', 'merging schools']
Adv. Learning	84	['learning', 'students', 'opportunities', 'advanced', 'math', 'access', 'support', 'help', 'help students', 'programs']
Small Class Size	70	['small', 'class', 'students', 'size', 'sizes', 'class size', 'small schools', 'smaller', 'schools', 'ratios']
Neighborhood Sch.	62	['neighborhood', 'neighborhood schools', 'schools', 'neighborhoods', 'quality', 'families', 'need neighborhood', 'attend', 'community', 'schools available']

Table 2: Resulting Topic Model Themes

substantially more coherent, interpretable, and analytically actionable, making the higher outlier rate an acceptable cost for clearer thematic structure. The higher outlier rate reflected a stricter, more meaningful separation between genuinely coherent themes rather than a failure of the model to organize the data.

5.2 Thematic Landscape of Community Input

The 14 identified topics (detailed in Table 2) reflect a diverse array of community priorities. The most frequent themes extended beyond simple academic metrics to include "K-8 Schools," "School Safety," and "Language Immersion Programs." Notably, the topic "Alternatives to Closures" captured direct resistance to the reorganization premise, with terms

like "don't close," "moratorium," and "stop" appearing frequently. This suggests that for many respondents there was a *rejection* of the idea selecting metrics for reorganization and instead a desire for the district to prioritize other possible solutions.

Crucially, the range of topics illustrates that community stakeholders evaluate school selection through a lens much wider than the standard administrative metrics of enrollment efficiency and standardized test scores. While "Academic Performance" ($n = 132$) was a recognized concern, it was heavily outweighed by systemic and human-resource factors, most notably "Teacher Turnover" ($n = 317$), which appeared more than twice as frequently. Additionally, the emergence of logistical and social themes such as "Transportation" ($n = 141$), "PTA Funding" ($n = 135$), and "Equity" ($n = 111$) underscores that families view schools as complex community anchors rather than mere facilities to be optimized for capacity. This distribution signals that a decision-making framework relying strictly on traditional performance and enrollment data would likely overlook the qualitative stability and support structures - specifically regarding staff retention and equitable access - that the community prioritizes most.

5.3 Racial Differences in Priorities

The omnibus Chi-square test confirmed significant racial disparities in how community members prioritized these topics and the Cramer's V analysis revealed it to be a small to moderate effect size for 65 degrees of freedom ($\chi^2 = 152.0825$, $p < 0.0001$, $V = 0.1439$). As shown in Figure 4, specific topics drove this divergence more than others. The "K-8 Schools" topic was the largest contributor to the Chi-square statistic ($\chi^2_{contrib} = 46.8$), followed by "School Safety" ($\chi^2_{contrib} = 20.7$) and "Language Immersion Programs" ($\chi^2_{contrib} = 13.8$).

Figure 3 presents the distribution of closure topics by respondent racial/ethnic group, with bar lengths indicating within-group percentages and color denoting standardized chi-square residuals (red = over-represented; blue = under-represented; $|residual| \geq 2$).

Across all racial groups, K–8 schools dominated the discourse, accounting for the largest share of responses in almost every group ($\approx$10–22%). This topic is significantly over-represented among respondents who declined to state race (22%) and
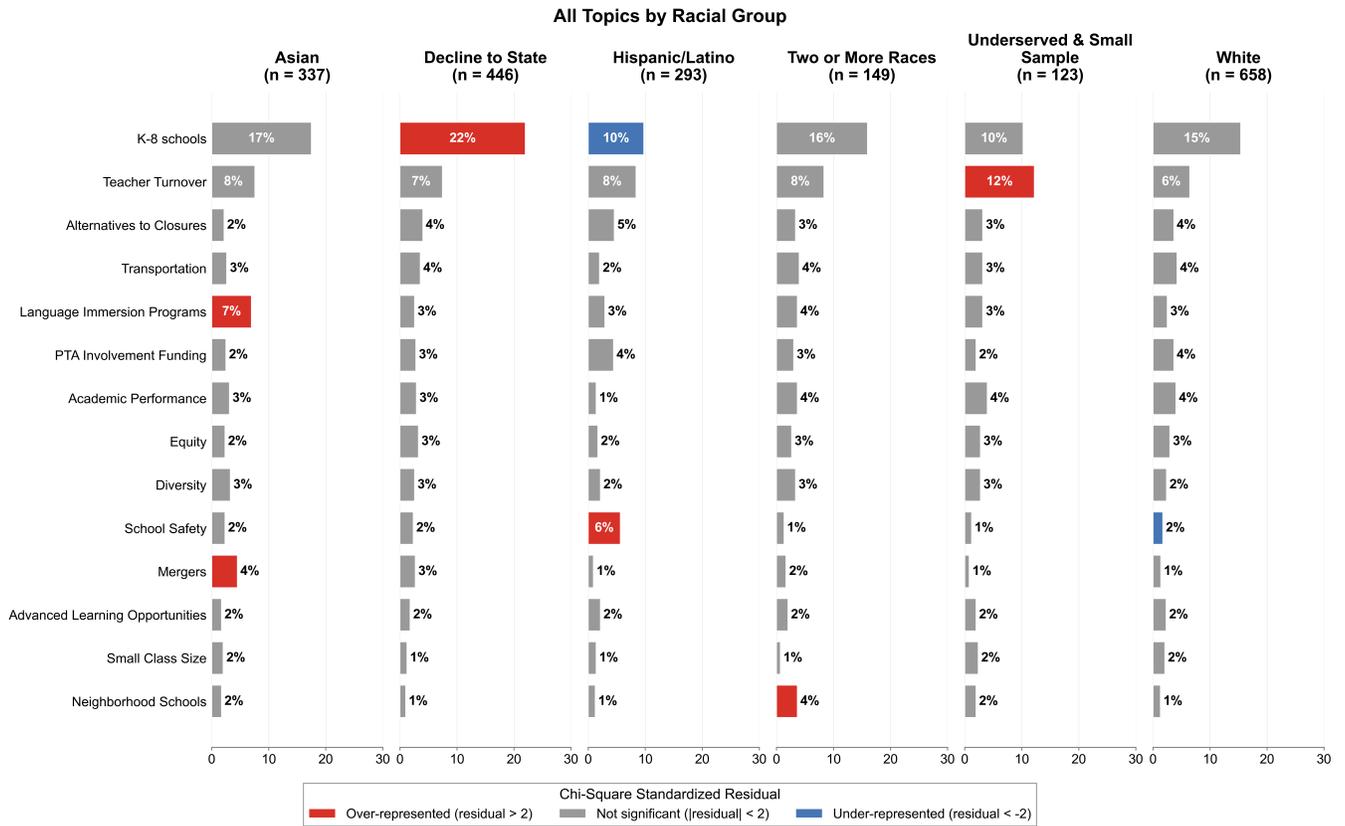
All Topics by Racial Group

Figure 3: Chi-Square Results

Note: Bars represent the percentage of responses within each racial group that reference a given topic. Bar colors indicate standardized chi-square residuals from tests of independence: red denotes over-representation ($residual \geq 2$), blue denotes under-representation ($residual \leq -2$), and gray denotes non-significant differences ($|residual| < 2$). Number of respondents, n, are shown in panel headers.

proportionally high among Asian (17%), White (15%), and multiracial respondents (16%), suggesting broad salience of elementary and middle school issues across groups.

Despite this and other similarities seen across groups, several group-specific emphases emerge. Asian respondents show a significant over-representation of language immersion program considerations (7%), distinguishing this group from others where the topic remains marginal (≈3–4%). Hispanic/Latino respondents display a significant over-representation of school safety (6%), while this topic is comparatively rare and sometimes under-represented among White respondents (2%). Respondents categorized as Underserved & Small Sample show a significant over-representation of teacher turnover (12%), indicating heightened concern about staffing stability in these communities, which is particularly concerning as it is also their most common topic discussed. Multiracial

respondents uniquely over-emphasize neighborhood schools (4%), a topic minimally discussed elsewhere.

Conversely, under-representation is limited but notable: Hispanic/Latino respondents discuss K–8 schools at a significantly lower rate (10%) relative to expectations, and White respondents show under-representation on school safety. Most remaining topics—such as equity, diversity, academic performance, transportation, mergers, and advanced learning opportunities—cluster around small percentages (≈1–4%) and do not differ significantly from expected distributions within most groups.

Overall, we find substantial cross-group convergence around a core set of issues, alongside targeted divergences that highlight how particular racial/ethnic communities disproportionately emphasize specific concerns with the reallocation framework provided by LUSD. These patterned dif-
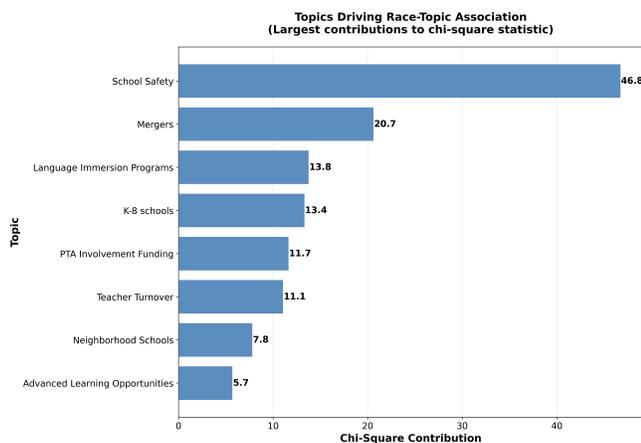
9

Figure 4: Topics Contributions to Racial Differences

Note: Bars represent each topic's contribution to the overall chi-square statistic from the test of independence between race/ethnicity and topic. Larger values indicate topics that account for a greater share of the observed race–topic association. Only topics with the largest contributions are shown.

ferences suggest that while many sentiments in the discourse are broadly shared, certain topics are especially salient for specific groups, with implications for inclusive agenda-setting and stakeholder engagement.

5.4 Model Performance

While the selected topic model achieved high semantic coherence and interpretability, this granularity came with a trade-off in coverage. As shown in Figure 5, 61.2% of respondents (but 45% of suggestions) had at least 1 of their suggestions labeled as an outlier. This indicates that a majority of suggestions contained idiosyncratic or unique phrasing that did not cluster densely enough to form coherent topics under the model's strict parameters. To ensure that this high outlier rate did not systematically silence specific community voices, we conducted a Chi-square analysis of outlier assignment by race.

The analysis revealed that White respondents were the only group to show a statistically significant over-representation in the outlier category (65.0%, standardized residual = 2.14). This suggests that while White respondents were active participants in the survey, their suggestions were significantly more likely to fall outside the core thematic clusters (such as "K-8 Schools," "School
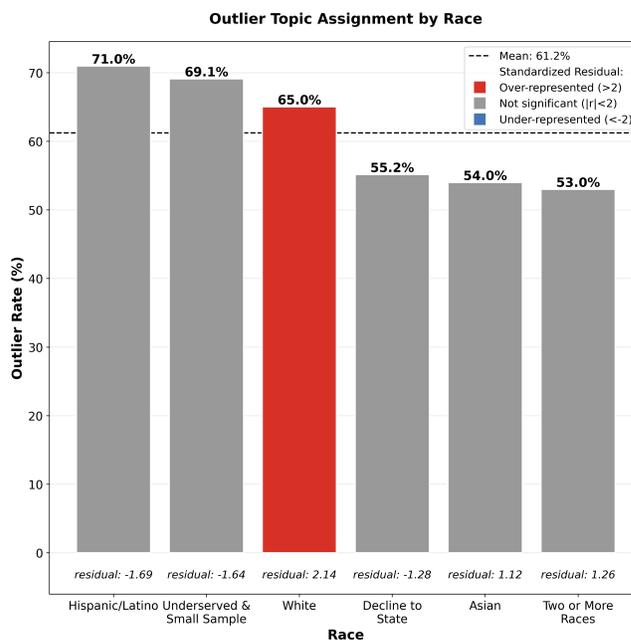

Figure 5: Outlier Representation by Race

Note: The dashed line represents the global mean outlier rate by respondents, in other words it shows that 61.2% of respondents had at least 1 suggestion labeled as an outlier. Bars represent the percentage of respondents classified as outliers within each racial group. Red bars indicate statistically significant over-representation (standardized residual > 2). Gray bars indicate non-significant differences.

Safety," and "Teacher Turnover") identified among the broader community. This finding aligns with prior literature suggesting that White parents in gentrifying urban districts may focus on niche concerns or leverage discourse that differs from the collective priorities of marginalized groups [Posey-Maddox, 2016, Clonan-Roy et al., 2023].

Descriptively, Hispanic/Latino respondents and those in the Underserved and Small Sample group exhibited the highest rates of outlier assignment (71.0% and 69.1%, respectively). However, these deviations did not reach statistical significance (standardized residuals < |2|), likely due to the smaller sample sizes of these groups relative to the White population in the dataset. Conversely, Asian respondents and those identifying as Two or More Races had the lowest outlier rates (54.0% and 53.0%, respectively), indicating that their suggestions were more consistently captured by the model's identified themes.

6 Discussion

The decision to close or reorganize schools is often framed by district leaders as a "technocratic" necessity driven by efficiency and enrollment. However, our analysis of over 4,000 community suggestions reveals that for the constituents of LUSD, the reorganization process is deeply tied to specific cultural, safety, and stability concerns that vary significantly by race. By leveraging neural topic modeling, we moved beyond surface-level keyword counting to uncover the distinct "languages of closure" spoken by different segments of the community.

Our primary finding confirms that racial identity strongly predicts the types of metrics community members value. While there was a convergence of interest around the "K-8 Schools" model - suggesting a district-wide preference for specific grade-span configurations - the divergences challenge the notion of a monolithic "community voice."

For Hispanic/Latino families, the significant over-representation of "School Safety" suggests that physical and emotional security is a prerequisite for any reorganization plan. This aligns with literature indicating that marginalized communities often view closures not just as a loss of a building, but as a disruption to a safe harbor in the neighborhood. Similarly, the focus on "Teacher Turnover" by the Underserved and Small Sample group highlights a concern with instability. For these communities, which have historically borne the brunt of district churn, the metric of success is not just test scores, but the retention of trusted adults. In contrast, Asian respondents' focus on "Language Immersion" points to specific programmatic priorities that risk being lost if schools are merged solely based on building capacity.

Perhaps the most novel finding emerged from our outlier analysis. White respondents were the only group significantly over-represented in the outlier category. This indicates that while White parents are often the most vocal in urban districts, their specific suggestions frequently fall outside the dense thematic clusters shared by the broader community. Drawing on Posey-Maddox [2016] and Clonan-Roy et al. [2023], this may reflect a tendency for advantaged parents to advocate for niche interests or individualized pedagogical preferences that diverge from the collective "survival" concerns - safety, staffing, and neighborhood access - voiced by families of color. This finding serves as a caution for administrators: a volume of feedback does not always represent a consensus of priority.

6.1 Policy Implications

These findings have immediate implications for how LUSD and similar urban districts solicit, analyze, and act upon community feedback.

First, districts must operationalize equitable closure practices, as described by Hahnel and Marchitello [2023] through scalable analysis. While research consistently urges districts to actively engage communities in determining closure metrics, the sheer volume of qualitative feedback often renders this daunting. Faced with thousands of responses, administrators may resort to simple frequency counts or cherry-picked quotes, effectively reducing deep community input to shallow metrics. This study demonstrates that modern NLP techniques provide a viable pathway to overcome this paralysis. By using tools like BERTopic to identify semantic patterns at scale, districts can move beyond performative listening and rigorously incorporate the "raw" institutional mourning and resistance of their constituents into the actual decision-making [Ewing, 2018, Ewing and Green, 2022]. Examples of such metrics LUSD might have been able to construct from these topics are available in Appendix Table A.1.

Second, engagement strategies must move beyond "one-size-fits-all" listening. The variation in priorities indicates that a standard town hall or survey may not capture the nuances of specific groups. If a district optimizes for the most frequent topic overall (e.g., K-8 models), they may inadvertently ignore the critical safety concerns of Hispanic/Latino families or the staffing concerns of Underserved and Small Sample groups because those groups constitute a smaller portion of the total response volume.

Third, data analysis must be stratified. As shown in this study, aggregate data can mask the distinct "languages of closure" spoken by different communities. If LUSD had looked only at the top topics for the whole district, "Teacher Turnover" might have appeared as a second-tier concern. Only by disaggregating by race did it emerge as the top priority for the district's most historically underserved populations. Policymakers utilizing broader, more robust metrics for evaluating schools for closures and

mergers during reorganization processes must explicitly weigh these stratified priorities rather than relying on a simple majority votes.

Finally, outliers should be analyzed. The high rate of unique, non-clustered suggestions from White respondents (65.0%) suggests that districts should be wary of letting idiosyncratic, highly specific demands drive systemic policy. Focused leadership requires distinguishing between the broad, survival-level concerns of marginalized groups—such as safety and stability—and the specific, often divergent preferences of a group with historically higher political capital.

7 Conclusion

School closures are among the most fraught events in education policy. This study demonstrates that the community response to these events is racialized. By applying NLP techniques to constituent feedback, we provided a method for districts to "listen" at scale without losing the signal of marginalized groups in the noise of the data. As LUSD moves forward with reorganization, utilizing these insights to create metrics that value safety, stability, and programmatic access—rather than just efficiency—will be essential to building trust and ensuring an equitable outcome.

7.1 Limitations

Our findings must be interpreted within the context of several methodological and sampling constraints.

First, the data represents a convenience sample gathered through open-access surveys. While we removed duplicate IP addresses to mitigate "ballot-stuffing", the demographic breakdown reveals an over-representation of White respondents relative to other groups, necessitating the aggregation of smaller racial groups (e.g., Native American, Pacific Islander) into a single "Underserved and Small Sample" category. This aggregation, while statistically necessary, risks masking distinct priorities within these unique communities.

Second, our reliance on automated translation tools (Google's Compact Language Detector) for the 11% of non-English responses introduces potential for nuance loss. While necessary for scale and the authors' ability to interrogate results, automated translation may fail to capture the specific idioms or emotional weight of terms used by non-native English speakers regarding their local schools. While we believe this to be somewhat less of a concern given the short nature of the text, we also acknowledge that Future work can and should examine the ability of multilingual neural embedding models that BERTopic could leverage, which remains an area for further research, to assist here in combining modern methodologies with district needs.

Finally, the strict density parameters of our BERTopic model resulted in a high outlier rate of 45% of suggestions. While this ensured that the 14 identified topics were semantically coherent and interpretable, it also meant that a majority of unique or idiosyncratic suggestions were excluded from the thematic map. Although our outlier analysis suggests this exclusion did not systematically bias against marginalized groups, it indicates that this topic modeling approach might only be able prioritize "consensus" narratives over niche or *highly* specific community concerns.

References

sentence-transformers/all-MiniLM-L6-v2 . Hugging Face, January 2024. URL https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2.

Nikolaos Aletras and Mark Stevenson. Evaluating Topic Coherence Using Distributional Semantics.

Anglin, Kylie L., Christian, Alvin, Jacob, Brian A., and Singleton, John D. Education Governance and Race: An Analysis of School Board Discourse Using Large Language Models. doi: 10.26300/EW42-T954. URL https://edworkingpapers.com/ai25-1175. Publisher: edworkingpapers.com.

Avitabile, Andrew, Bartanen, Brendan, and Kwok, Andrew. Using Large Language Models to Analyze Preservice Teacher Feedback and Reflections During Clinical Teaching. doi: 10.26300/SBDB-AE05. URL https://edworkingpapers.com/ai25-1203. Publisher: edworkingpapers.com.

Brendan Bartanen, Andrew Kwok, Andrew Avitabile, and Brian Heseung Kim. Why Do You Want to Be a Teacher? A Natural Language Processing Approach. *Educational Researcher*, 54(1):7–20, January 2025. ISSN 0013-189X, 1935-102X. doi: 10.3102/0013189X241276814. URL https://journals.sagepub.com/doi/10.3102/0013189X241276814.

Michael L. Chrzan, Francis A. Pearman II, and Benjamin W. Domingue. Deeper Roots Before the Storm: Utilizing Machine Learning to Alert School Districts of Permanent School Closures. *EdWorkingPapers*, June 2025. doi: 10.26300/QPVX-JV29. URL https://edworkingpapers.com/ai25-1210. Publisher: edworkingpapers.com.

Katherine Clonan-Roy, Rhiannon Maton, Charlotte E. Jacobs, Casey Matthews, Michael Kokozos, and Erika Kitzmiller. White Parent and Caregiver Perceptions of, and Resistance to, Equity and Anti-Racism Work in an Independent School. *Teachers College Record*, 125(7-8):77–109, August 2023. ISSN 0161-4681. doi: 10.1177/01614681231199558. URL https://doi.org/10.1177/01614681231199558. Publisher: SAGE Publications.

Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Routledge, New York, 2 edition, 1988. ISBN 978-0-203-77158-7. doi: 10.4324/9780203771587.

H Cramer. Mathematical methods of statistics, princeton, 1946, 1946. URL https://cir.nii.ac.jp/crid/1370290617608494736. Pages: 300 Publication Title: Math Rev (Math-SciNet) MR16588 Zentralblatt MATH Volume: 63.

Dorottya Demszky, Jing Liu, Heather C. Hill, Shyamoli Sanghi, and Ariel Chung. Automated feedback improves teachers' questioning quality in brick-and-mortar classrooms: Opportunities for further enhancement. *Computers & Education*, 227:105183, April 2025. ISSN 03601315. doi: 10.1016/j.compedu.2024.105183. URL https://linkinghub.elsevier.com/retrieve/pii/S0360131524001970.

Elizabeth DeOrnellas. The impact of the 2013 school closures, May 2023. URL https://philadelphianeighborhoods.com/2023/05/11/education-the-lasting-effects-of-the-2013-school-closures/.

Mary Eddins, Maja Pehrson, Kevin Burgess, Contributor Kate Callahan, Contributor David Lapp, Contributor DeQuendre Bertrand, and Contributor Brooke Ruoff. Revisiting Research on School Closings: Key Learnings for District and Community Leaders. Technical report, Research for Action, June 2024. URL https://eric.ed.gov/?id=ED660643. ERIC Number: ED660643.

Roman Egger and Joanne Yu. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, 7, May 2022. ISSN 2297-7775. doi: 10.3389/fsoc.2022.886498. URL https://www.frontiersin.org/journals/sociology/articles/10.3389/fsoc.2022.886498/full. Publisher: Frontiers.

Eve L. Ewing. *Ghosts in the schoolyard: racism and school closings on Chicago's South side*. The University of Chicago Press, Chicago, 2018. ISBN 978-0-226-52602-7.

Eve L. Ewing and Terrance L. Green. Beyond the Headlines: Trends and Future Directions in the School Closure Literature.

Educational Researcher, 51(1):58–65, January 2022. ISSN 0013-189X, 1935-102X. doi: 10.3102/0013189X211050944. URL https://journals.sagepub.com/doi/10.3102/0013189X211050944.

Alexandra Freidus. "A Great School Benefits Us All": Advantaged Parents and the Gentrification of an Urban Public School. *Urban Education*, 54 (8):1121–1148, October 2019. ISSN 0042-0859. doi: 10.1177/0042085916636656. URL https://doi.org/10.1177/0042085916636656. Publisher: SAGE Publications Inc.

Ismael Garrido-Muñoz, Arturo Montejo-Ráez, Fernando Martínez-Santiago, and L. Alfonso Ureña-López. A Survey on Bias in Deep NLP. *Applied Sciences*, 11(7):3184, January 2021. ISSN 2076-3417. doi: 10.3390/app11073184. URL https://www.mdpi.com/2076-3417/11/7/3184. Publisher: Multidisciplinary Digital Publishing Institute.

Nabeel Gillani, Eric Chu, Doug Beeferman, Rebecca Eynon, and Deb Roy. Parents' Online School Reviews Reflect Several Racial and Socioeconomic Disparities in K–12 Education. *AERA Open*, 7:2332858421992344, January 2021. ISSN 2332-8584, 2332-8584. doi: 10.1177/2332858421992344. URL https://journals.sagepub.com/doi/10.1177/2332858421992344.

Terrance Green. "We felt they took the heart out of the community": Examining a community-based response to urban school closure. *Education Policy Analysis Archives*, 25:21, March 2017. ISSN 1068-2341. doi: 10.14507/epaa.25.2549. URL https://epaa.asu.edu/index.php/epaa/article/view/2549.

Maarten Grootendorst. BERTopic: Neural topic modeling with a class-based TF-IDF procedure, March 2022. URL http://arxiv.org/abs/2203.05794. arXiv:2203.05794 [cs].

Carrie Hahnel and Max Marchitello. Centering Equity in the School-Closure Process in California. 2023.

Tomer Iwan. polyglotr: Translate Text, August 2024. URL https://cran.r-project.org/web/packages/polyglotr/index.html.

Amandeep Kaur and James R. Wallace. Moving Beyond LDA: A Comparison of Unsupervised Topic Modelling Techniques for Qualitative Data Analysis of Online Communities, December 2024. URL http://arxiv.org/abs/2412.14486. arXiv:2412.14486 [cs] version: 1.

Li Lucy, Dorottya Demszky, Patricia Bromley, and Dan Jurafsky. Content Analysis of Textbooks via Natural Language Processing: Findings on Gender, Race, and Ethnicity in Texas U.S. History Textbooks. *AERA Open*, 6(3): 2332858420940312, July 2020. ISSN 2332-8584. doi: 10.1177/2332858420940312. URL https://doi.org/10.1177/2332858420940312. Publisher: SAGE Publications Inc.

Leland McInnes, John Healy, and Steve Astels. hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11):205, March 2017. ISSN 2475-9066. doi: 10.21105/joss.00205. URL http://joss.theoj.org/papers/10.21105/joss.00205.

Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, September 2020. URL http://arxiv.org/abs/1802.03426. arXiv:1802.03426 [stat].

Davoud Moulavi, Pablo A. Jaskowiak, Ricardo J. G. B. Campello, Arthur Zimek, and Jörg Sander. Density-Based Clustering Validation. In *Proceedings of the 2014 SIAM International Conference on Data Mining (SDM)*, Proceedings, pages 839–847. Society for Industrial and Applied Mathematics, April 2014. doi: 10.1137/1.9781611973440.96. URL https://epubs.siam.org/doi/abs/10.1137/1.9781611973440.96.

Moin Nadeem, Anna Bethke, and Siva Reddy. StereoSet: Measuring stereotypical bias in pretrained language models. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5356–5371, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.416. URL https://aclanthology.org/2021.acl-long.416/.

Sally A. Nuamah. The paradox of educational attitudes: Racial differences in public opinion on school closure. *Journal of Urban Affairs*, 42(4):554–570, May 2020. ISSN 0735-2166. doi: 10.1080/07352166.2017.1360734. URL https://doi.org/10.1080/07352166.2017.1360734. Publisher: Routledge _eprint: https://doi.org/10.1080/07352166.2017.1360734.

Jeroen Ooms [aut, cre, and Dirk Sites (Author of CLD2 C++ library). cld2: Google's Compact Language Detector 2, October 2024. URL https://cran.r-project.org/web/packages/cld2/index.html.

Francis A. Pearman and Danielle Marie Greene. School Closures and the Gentrification of the Black Metropolis. *Sociology of Education*, 95(3):233–253, July 2022. ISSN 0038-0407, 1939-8573. doi: 10.1177/00380407221095205. URL https://journals.sagepub.com/doi/10.1177/00380407221095205.

Linn Posey-Maddox. Beyond the consumer: parents, privatization, and fundraising in US urban public schooling. *Journal of Education Policy*, 31(2):178–197, March 2016. ISSN 0268-0939. doi: 10.1080/02680939.2015.1065345. URL https://doi.org/10.1080/02680939.2015.1065345. Publisher: Routledge _eprint: https://doi.org/10.1080/02680939.2015.1065345.

Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, August 2019. URL http://arxiv.org/abs/1908.10084. arXiv:1908.10084 [cs].

David Rozado. Wide range screening of algorithmic bias in word embedding models using large sentiment lexicons reveals underreported bias types. *PLOS ONE*, 15(4):e0231189, April 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0231189. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0231189. Publisher: Public Library of Science.

Esa Syeed. "It just doesn't add up": Disrupting official arguments for urban school closures with counterframes. *Education Policy Analysis Archives*, 27:110, September 2019. ISSN 1068-2341. doi: 10.14507/epaa.27.4240. URL https://epaa.asu.edu/index.php/epaa/article/view/4240.

Wenhui Wang, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. MiniLM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers, April 2020. URL http://arxiv.org/abs/2002.10957. arXiv:2002.10957 [cs].

Paiheng Xu, Jing Liu, Nathan Jones, Julie Cohen, and Wei Ai. The Promises and Pitfalls of Using Language Models to Measure Instruction Quality in Education. In Kevin Duh, Helena Gomez, and Steven Bethard, editors, *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 4375–4389, Mexico City, Mexico, June 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.naacl-long.246. URL https://aclanthology.org/2024.naacl-long.246/.

Appendix

A   Additional Tables and Figures

This appendix provides supplementary visualizations detailing the foundational characteristics of the survey data utilized in this study. Figure A.1 breaks down the absolute counts of unique respondents by their self-identified racial and ethnic categories, offering transparency into the sample's composition prior to the necessary statistical aggregations employed in our primary analyses. Additionally, Figure A.2 examines the distribution of word counts across the submitted open-ended suggestions. By visualizing response lengths both in aggregate and disaggregated by racial group, this figure demonstrates that textual verbosity remained largely consistent across demographics, mitigating concerns that response length systematically skewed the subsequent topic modeling outputs.
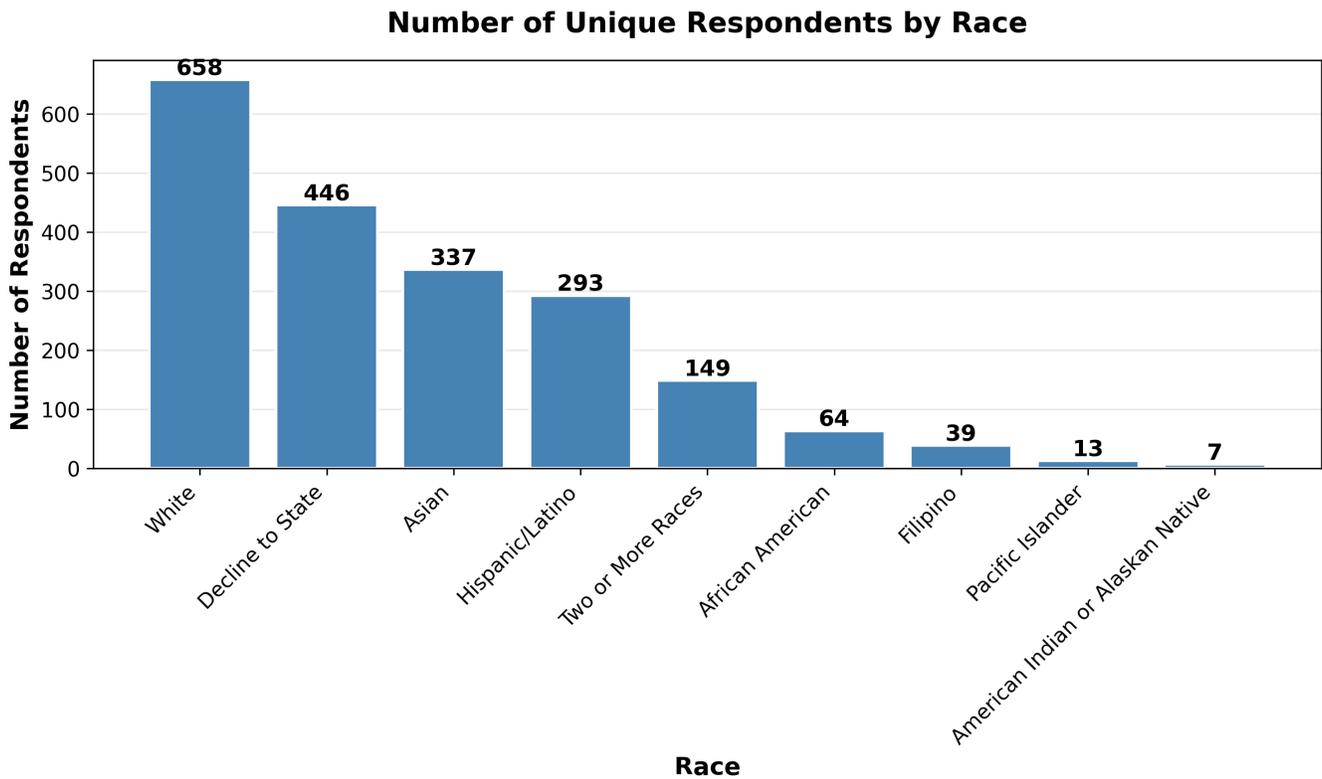

Figure A.1: Respondent Counts by Race

Distribution of Unique Survey Respondents by Racial and Ethnic Group. This bar chart illustrates the absolute frequencies of the unique respondents categorized by their self-identified race or ethnicity. To ensure statistical validity in the primary chi-square analyses, demographic groups with limited sample sizes—specifically African American, Filipino, Pacific Islander, and American Indian or Alaskan Native—were subsequently aggregated into an "Underserved and Small Sample" category.
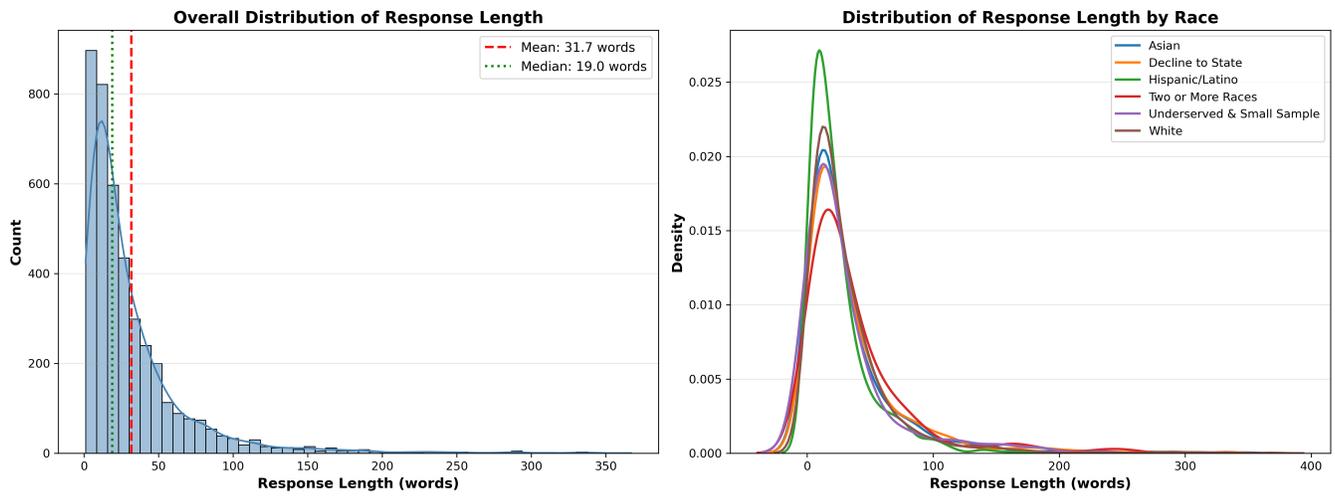
Figure A.2: Response Length Distributions

Distribution of Open-Ended Response Lengths. The left panel displays the overall distribution of word counts for the 4,159 analyzed suggestions, highlighting a right-skewed distribution with a mean of 31.7 words and a median of 19.0 words. The right panel presents density plots of response lengths disaggregated by racial group, demonstrating that the distributions of suggestion verbosity are largely uniform across all demographic categories, thereby limiting the influence of response length on topic assignment.

Table A.1: Example Metrics

Topic	Example Metrics (school-level)
K-8 Schools	**Grade-span stability**: Whether school is a K–8. **Feeder disruption index**: number of feeder changes closure would cause.
Teacher Turnover	**Annual teacher turnover rate**: percent leaving per year. **3-year stability index**: percent of staff with 3+ years at school.
Alternatives to closures	**Feasibility of alternatives score**: number of viable options (co-location, boundary shifts, program moves) assessed and rated before closure.
Language Immersion	**Program uniqueness index**: number of comparable immersion seats within X miles. **Continuity risk**: percent of immersion students who'd lose pathway.
Transportation	**Average student commute time change**: minutes added if reassigned. **Access burden index**: percent of students losing walkable access (¡1 mile).
PTA involvement/funding	**Community capacity index**: PTA participation rate. **Fundraising equity index**: per-pupil PTA funds vs district median.
Academic performance	**Growth-focused metric**: median student growth percentile. **Support need index**: percent students below proficiency with trend over 3–5 years.
Equity	**Displacement equity index**: which racial/SES groups are disproportionately affected (CA law). **Service equity index**: impact on special populations (SPED, EL, homeless).
Diversity	**Enrollment diversity index**: e.g., Simpson or Theil index. **Segregation impact score**: change in racial/SES balance at receiving schools.
School safety	**Safety climate index**: survey-based safety scores. **Incident-adjusted metric**: incidents per 100 students, trend over 3 years.
Mergers	**Merger viability score**: program compatibility, building capacity fit, projected class sizes. **Disruption index**: number of students changing school and program.
Advanced learning	**Access to advanced opportunities**: percent of students with access to advanced/accelerated courses. **Program loss risk**: number of unique advanced offerings at risk.
Small class size	**Student–teacher ratio**: by grade band. **Class size equity index**: comparison to district averages, especially for high-need students.
Neighborhood schools	**Neighborhood anchor index**: percent of students living within 1 mile. **Community reliance score**: number of community orgs using the building, evening/weekend use hours.

Note: This table describes an example multi-criteria framework our partner district could have used to assess school-level options based on the results of our topic model, shifting from a technocratic lens to an explicitly equitable one. The framework treats strong performance on safety, stability, community anchoring, and program uniqueness as protective signals against closure, even when enrollment is low. Metrics are organized into the thematic domains identified by our topic model and are computed where possible by racial/ethnic subgroup to detect disproportionate impacts.