Texas Transportation Poll Sentiment Analysis

Executive Summary

Poll Question

As part of the 2014 Texas Transportation Poll conducted by the Texas A&M Transportation Institute, the following open-ended question was asked: "What is the most significant transportation issue affecting you personally in your region?" (*1*). This question elicited over 3000 independent responses.

Using qualitative analysis software, researchers identified common themes within the responses to answer the following research questions:

- What are the most common transportation issues reported by Texans?
- Are their sentiments on these issues positive or negative?
- Are citizens giving us insight into what transportation policy makers should do to address transportation issues?

Frequently Cited Transportation Issues

Researchers found that the Texans who responded to the open-ended question most frequently identified the following issues that affected them, in order of most frequently cited to least frequently cited: congestion, gas prices, construction, public transportation, road maintenance, drivers (driver behavior), planning, bicycle and pedestrian, and toll roads.

Sentiments about Transportation Issues

Once researchers determined the most frequently cited issues, they evaluated the sentiment that respondents indicated for each of the transportation issues. The aim was to determine whether respondents' emotions were generally positive, generally negative, or generally neutral.

Respondents who indicated "drivers (driver behavior)," "road maintenance," or "toll roads" as the most significant transportation issue showed extremely negative sentiment. In fact, the topics of driver behavior and road maintenance both garnered almost seven times as many negative emotions as every one positive emotion.

Conversely, respondents who indicated either "public transportation," "bicycle and pedestrian," or "safety" as the most significant transportation issue showed slightly more positive than negative sentiment. These three issues were the only topics that garnered more positive comments than negative comments. However, since the open-ended question asked respondents to identify the most significant "transportation issue," which could be read as a transportation problem, the responses were inherently negative.

Implications for Transportation Policy in Texas

Researchers evaluated comments for each topic and developed the following broad implications for transportation policy in Texas:

- Congestion: The public is likely to well receive policies and practices that mitigate rural and urban traffic congestion. Policies should take into consideration public perceptions of the causes of traffic congestion, and use education and outreach to broaden public support.

- Gas prices: Drivers in Texas perceive the price of fuel as a significant issue. This perception is likely to fluctuate with the price of gas.

- Road maintenance: Respondents indicated that they are the most passionate about the inconveniences created by road maintenance or lack thereof. Policies that mitigate the impacts of road maintenance and/or improve road conditions may alleviate the negative sentiments about these issues.

- Diverse, safe transportation options: Respondents indicated that they are in favor of diverse, safe transportation options. Of the respondents who chose to discuss roadway safety and alternative modes of transportation, researchers noted positive sentiments toward the topics of adding bicycle and pedestrian facilities, expanding public transit opportunities, and improving road safety.

In addition, researchers further analyzed the comments of respondents who indicated that "congestion" was the most significant transportation issue that they faced in their region because it was the most frequently cited issue. Researchers evaluated these comments in order to identify what respondents believed was the perceived cause of congestion. The following is a summary of policy implications based on the findings of the top five perceived causes of traffic congestion:

- Road capacity: Many commenters attributed traffic congestion to inadequate road capacity, which could imply a need to better inform the public about the complexities of transportation planning and funding.

- Rush hour: The high number of comments about rush-hour traffic presents an opportunity for policy makers to mitigate the issue through policies that specifically address traffic congestion during peak periods, such as staggered or flexible work schedules, telecommuting, ridesharing, and other trip-reduction policies.

- Construction: Traffic congestion due to road construction aggravates Texas drivers and costs time. Outreach and educational efforts to inform the public about the duration of construction projects or how to get more information could be effective in reducing negative perceptions.

- Population growth: Accommodating new residents and the increased transportation demand they bring is a key challenge facing Texas leaders. Respondents believed that the growing population is leading to increased traffic but suggested differing perspectives on how best to address this challenge. This implies that policies that manage transportation demand are needed in high-growth regions and could be accompanied by clear information about what government is doing to address the mobility impacts.

- Traffic signals: As shown in the comments regarding traffic congestion, the public thinks traffic signal timing should be a high-priority issue. Programs that address traffic signal timing, such as 311 and other systems for gathering citizen input on traffic signal issues, at specific intersections could be especially helpful in rural counties with smaller transportation departments.

Introduction

This analysis provides insight into the transportation issues that are most significant to the public. Understanding the most commonly faced transportation issues and the degree of negative or positive sentiment toward these issues can allow policy makers to better deduce how palatable potential transportation policies will be to the public. Additionally, this analysis establishes a baseline process for researchers to conduct further qualitative analyses that will inform future quantitative studies, including the next Texas Transportation Poll.

This policy brief presents the following sections:

- A discussion of the most significant transportation issues reported.
- Respondents' perceptions of the causes of transportation issues.
- The potential implications that perceived transportation issues may have for transportation policy in Texas.
- Lessons learned about the utility of qualitative and sentiment analysis in transportation research and best practices for future statewide surveys.

Appendix A gives further information about the methodology used to conduct the sentiment analysis and qualitative analysis. Appendix B gives a review of current literature on the subject.

Transportation Issues

Analyzing the qualitative data received from the open-ended question, "What is the most significant transportation issue affecting you personally in your region?" in the 2014 Texas Transportation Poll provides insight into what Texans think are the most significant transportation issues affecting their respective regions. Evaluating both the volume of comments, the content of the comments, and the overall sentiment of those comments can provide transportation policy makers with unique insights into public opinion.

In order to do this, researchers pulled out the most commonly referred to topics in the comments and then looked at each topic individually to assess public sentiment. A detailed description of the methodology used is in Appendix A.

Frequency Analysis

Figure 1 shows that roads, traffic, and congestion were the most frequently mentioned topics in the responses to the open-ended question. Once researchers identified the most frequently referred to topics, they were able to isolate the comments that referred to each topic in order to understand if Texans felt positively or negatively toward that topic.

Figure 1. Word Cloud Displaying the Word Frequency for Responses to the Open-Ended Question: "What Is the Most Significant Transportation Issue Affecting You Personally in Your Region?"

Researchers organized the input from respondents into the following 10 topics:

- Bicycle and pedestrian.
- Congestion.
- Construction.
- Drivers.
- Gas prices.
- Road maintenance.
- Planning.
- Public transportation.
- Safety.
- Toll roads.

The most frequently mentioned topic was, by far, traffic congestion, garnering over 750 individual references throughout all of the unique responses. Gas prices, construction, and public

transportation were also among the most frequently referenced topics. Figure 2 shows the relative frequency of the 10 main topics that researchers identified.
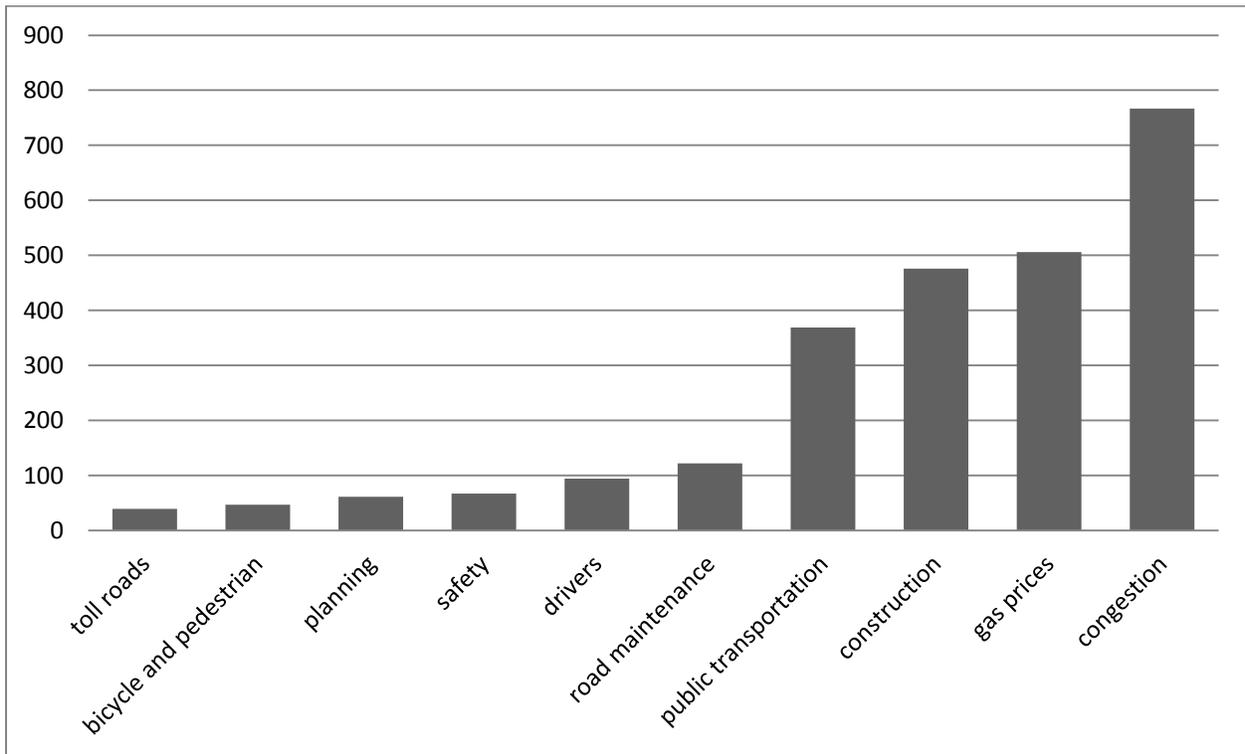

Figure 2. Frequency of Topics Mentioned in Responses to the Open-Ended Question.

Sentiment Analysis

Sentiment analysis allows for a more nuanced understanding of whether Texans feel positively or negatively toward a particular topic. For this analysis, sentiment was measured using Linguistic Inquiry Word Count (LIWC), an empirically validated linguistic software to infer psychological and/or emotional reactions. More detail about the LIWC software and methodology can be found in Appendix A.

The sentiment analysis has two caveats. First, the findings of the sentiment analysis are not necessarily representative of the positive or negative feelings of all Texans because the sentiment of each topic was analyzed only for respondents that chose to provide input regarding that particular topic. Second, the question elicited inherently negative responses because it asked respondents to indicate an issue that affects them personally. Therefore, this sentiment analysis reflects the degree of negativity used to describe transportation issues facing Texans in their regions.

Figure 3 illustrates the average sentiment score for each of the 10 main topics. Bicycle and pedestrian, safety, and public transportation were described using the most positive words, while the behavior of other drivers and road maintenance were described using the most negative

words. The sentiments expressed toward the top three issues (congestion, gas prices, and construction) were all neutral or slightly negative.
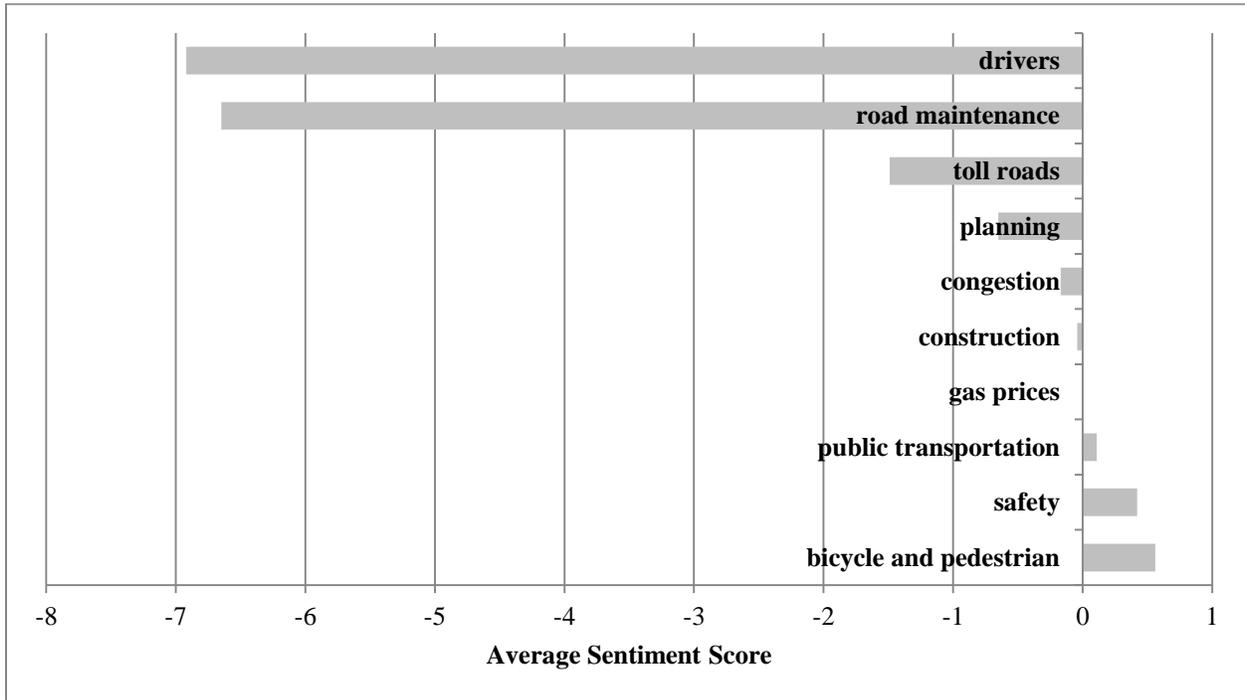

Figure 3. Sentiment Analysis by Topic.

Overall, the most negative sentiments did not always correspond to the most frequently mentioned topics. For example, sentiments toward road maintenance and drivers were considerably more negative than all other subjects, indicating that people have the strongest feelings toward these issues. Yet drivers and road maintenance were not the most frequently discussed topics. This indicates that feelings toward other drivers and toward road maintenance, while not the most frequently reported issue, evoke the strongest negative sentiments.

Since computer-based sentiment analysis considers the frequency of positive and negative words, this only implies associations with these topics and should not be considered a conclusive analysis. For instance, respondents mentioning "gas price" often included no further description of how or why this is a significant issue for them. At the time this poll was conducted, the average price of gas in Texas was $3.45 per gallon, the most expensive it had been since 2013 (*2*). Though one might assume gas prices could be negatively associated for respondents as an economic restriction on mobility, the lack of further description kept the average sentiment score neutral. Therefore, this analysis can be useful to identify important topics for the respondents but may not be conclusive about attitudes for all topics.

Most Significant Transportation Issues

This section provides an overview of public sentiment toward the first five of the 10 most significant transportation issues affecting Texans in their region and also briefly discusses other transportation issues.

Congestion

Respondents overwhelmingly cited traffic congestion as the most significant transportation issue affecting them in their region. Despite being the most frequently mentioned topic, average sentiment toward traffic congestion was just less than zero, indicating that, on average, comments were only slightly negative.

Respondents commonly named population growth, a lack of road space, an influx of college students, construction, and poorly timed traffic signals as causes of traffic congestion, as seen in the following comments:

> *"Too much traffic, especially during rush hours" (female, 25–34, Harris County).*

> *"too many cars and not enough lanes to drive in" (female, 45–54, Denton County).*

> *"Too many cars because everything is a significant distance from home to work or errands" (female, 55–64, Dallas County).*

Gas Prices

The high cost of fuel was the second most commonly cited transportation issue mentioned by respondents when the poll was conducted in spring 2014. As discussed previously, the average price of gas in Texas was $3.45 per gallon at that time (*2*). Because gas prices have dropped significantly over the last year, future Texas Transportation Poll surveys can help to track public sentiment toward this evolving issue.

Comments regarding gas prices were typically simple and to the point, for example, "gas prices" or "the price of fuel is too high." Due to the relative lack of description on this subject, overall public sentiment toward gas prices was found to be neutral. However, some respondents offered insight into how to ease the economic burden of gas prices, such as increased mobility and public transportation options, as seen in the following comments:

> *"If we had public transport trains in the area we could save on car wrecks, buying cars, gas & less traffic, pollution etc." (female, 55–64, Matagorda County).*

> *"price of gas coupled with the lack of options other than private vehicles. I would love to use public transportation, but it doesn't exist where I live" (female, 35–44, no location provided).*

Construction

Road construction was one of the most frequently mentioned topics by respondents, although sentiments toward road construction were only slightly negative. Most comments regarding construction pertained to the impact of construction on traffic delays and congestion. Many

respondents also commented on the amount of construction in their area or addressed issues at specific locations in their region, as seen in the following comments:

> *"Traffic due to (seemingly endless) construction" (female, 25–34, Tarrant County).*

> *"Multiple areas of the roads are being worked on at the same time making traffic worse" (male, 35–44, El Paso County).*

> *"Having trouble navigating the construction work on Airport Freeway from North Richland Hills to Dallas. Temporary lane changes and detours with poorly marked stripes and deflectors have made some areas dangerous" (male, 55–64, Tarrant County).*

Public Transportation

Public transportation was the fourth most frequently mentioned topic. Additionally, public transportation was one of the few topics that garnered a positive sentiment score; it averaged a score slightly higher than zero (see Figure 3).

The comments provided by respondents regarding public transportation called for additional facilities or increased access to public transit, while some also offered specific fixes to improve transit service for certain populations or locations. The following are several typical responses about public transportation:

> *"Not enough affordable public transit options" (male, 25–34, Williamson County).*

> *"Busses not operating late enough at night or early enough in the morning" (female, 35–44, Bexar County).*

> *"Lack of public transportation for the elderly" (male, 55–64, Kerr County).*

> *"No public transportation in suburbs/rural areas" (female, 25–34, Williamson County).*

> *"Lack of comprehensive public transport which presents a practical alternative to driving a car" (male, 35–44, Travis County).*

Road Maintenance

The fifth most frequently mentioned topic that respondents cited—and the issue that drew the most negative sentiment—was road maintenance. The subject of the comments regarding maintenance issues reflected two main issues: the inconvenience of road maintenance and poorly maintained roads. These comments were expressed using strong negative emotions, leading to an average sentiment score less than −6, as shown in the following comments:

> *"Potholes and terrible roads!!!!!!!" (female, 25–34, Bexar County).*

> *"Poor roads which haven't been maintained properly" (female, 25–34, no location provided).*

> *"The poor quality of surface roads in Dallas" (male, 45–54, Dallas County).*

> *"Poor road quality due to rural setting and excessive traffic (heavy vehicles and school buses) which slow traffic to major highways and cause damage to cars..." (female, 55–64, Kaufman County).*

Other Transportation Issues

Several respondents provided additional insights that were not among the most common topics but that describe how they perceive significant transportation issues.

One respondent was interested in greater investment in the state highway system:

> *"The fact that the state government hasn't put money into improving the state highways" (female, 65+, Robertson County).*

This comment could imply that regardless of the fixed funding streams devoted to transportation, this responder feels state government could reallocate funding to make highways a more significant priority.

Several respondents also noted issues with growing heavy truck traffic, sometimes in rural areas that may have limited routes designed to handle large trucks:

> *"A major influx of eighteenwheelers and oil field trucks on the little roads of our small town make it hard to get anywhere, especially at stoplights where giant trucks run lights or get stuck trying to turn" (female, 18–24, Dawson County).*

This respondent may be concerned about the potential for safety issues resulting from the increase in truck traffic in areas that usually did not have this burden.

Several respondents also pointed out mobility restrictions due to personal income:

> *"unable to buy own transportation, having to rely on public transportation that does not start running early enough, continue running late enough, or run in certain areas or on weekends/holidays at all" (female, 45–54, Harris County).*

This respondent is reporting financial difficulties as a barrier to mobility, and the transit system as an alternative does not offer the flexibility needed to meet her needs. This could indicate the challenge of mobility as it relates to job access and other goods and services outside of walking distance from housing. Another respondent relays a different set of challenges addressing public transportation:

> *"As an ecologically minded person, I would love to take Public Transportation, however, I don't feel safe where I work after dark" (female, 25–34, Collin County).*

Rather than noting restriction of personal funding, this respondent notes concerns with personal security. This could be related to a local workplace neighborhood or be a more general comment about her concerns about the security of transit systems.

Several people made connections between transportation and planning issues:

"housing too far away from jobs and services. I'd walk if things were closer" (female, 55–64, no location provided).

She notes the need to plan residential and employment destinations with access in mind. What some may see as transportation issues, this respondent may perceive as more of a land-use issue.

Public Perception of the Causes of Transportation Issues

The previous section described the most common transportation issues and their causes that respondents say they face in their region. These insights can assist policy makers to identify the public's perceptions of the cause, define palatable solutions, and determine appropriate messaging and language to use with the public.

To further understand public perception of the causes of transportation issues, researchers selected the most frequently mentioned transportation issue, traffic congestion, for additional qualitative analysis. The word cloud in Figure 4 reveals the most common words used to describe the issue of traffic congestion.

Figure 4. Traffic Congestion Word Cloud Results.

The majority of comments regarding traffic congestion were simple, stating "too much traffic" or "traffic congestion." Other comments were more descriptive, offering additional insight into the public's view of the causes of congestion. The top five most frequently cited causes of traffic congestion were:

- Road capacity.
- Rush hour.
- Construction.
- Population growth.

- Traffic signals.

Road Capacity

Many respondents blame traffic congestion on a lack of road space. Over 10 percent of total references to congestion mentioned road capacity, the number of lanes, or the number of roads. There appeared to be general consensus on this:

"too many cars and not enough lanes to drive in" (female, 45–54, Denton County).

"too many cars for the amount of lanes on the roads" (male, 35–44, Montgomery County).

These comments were also sometimes explained by compounding factors such as traffic signals and other drivers:

"Roads do not have the capacity to carry the traffic coupled with poorly timed traffic signals…" (male, 55–64, Travis County).

Rush Hour

Another area of consensus among respondents was that rush-hour traffic congestion is a major issue, comprising over 10 percent of total references to congestion. The pattern of responses reveals that respondents from all regions experience traffic delays during the morning and evening commute. For example, these respondents from rural Brazoria County and urban Dallas County mention rush hour as the biggest transportation issue in their respective region:

"Traffic delays in morning and evening commutes to work" (male, 35–44, Brazoria County).

"great volumes of traffic going and leaving work" (female, 25–34, Dallas County).

Construction

Respondents felt that traffic congestion is exacerbated when roads are under construction or maintenance. The overall consensus is:

"too much congestion and construction on roads" (female, 55–64, Coryell County).

Some comments were more specific about the location or type of construction, giving a more nuanced glimpse into the opinions of Texans and offering insight into what transportation policies they may or may not support. For example, this woman from Travis County would likely not support additional toll projects in the region:

"Traffic due to construction to build more toll roads we don't need" (female, 35–34, Travis County).

This woman from Denton County appreciates alternative work schedules for road crews:

> *"New road construction (adding lanes to the interstate), which causes back-up. Luckily the road crews work late night/early morning to assist with avoiding traffic delays" (female, 55–64, Denton County).*

Population Growth

Respondents recognized the demand placed on transportation infrastructure due to rapid population growth in many regions of the state. Many respondents expressed the sentiment that traffic congestion was due to population growth in their region, for example:

> *"road congestion due to increase in population in areas with insufficient lanes of traffic" (male, 35–44, Bexar County).*

Respondents were not always in agreement about the nature of the relationship between population growth and traffic congestion. Some simply mention population growth and the lack of roadway capacity growth, as seen in the previous comment, while others attributed traffic to their region's economic strength and resulting increase in the number of commuters, as seen in this comment:

> *"The economic growth and amount of businesses moving to my city bring in more residents that congest the highways during rush hour travel times" (female, 35–44, Bexar County).*

Traffic Signals

Traffic signal timing was the fifth most commonly cited cause of traffic congestion by respondents. For the most part, respondents blame congestion at intersections on the timing of traffic signals and/or on the people in charge, for example:

> *"Traffic lights that do not function properly. I have been stuck at several specific lights for 3–7 cycles during low traffic times" (female, 25–34, Nueces County).*

> *"The people in charge do NOT understand traffic flow. I have never seen traffic lights timed so poorly. I have watched poorly-timed traffic lights CAUSE congestion, day after day" (male, 45–54, Rusk County).*

Implications for Texas Transportation Policy

Policy Implications of Most Significant Transportation Issues

Mining the qualitative data gathered from the Texas Transportation Poll provides insight into implications for transportation policy in Texas. As noted previously, the findings of this analysis are based on input that was voluntarily provided by the respondents of the poll. The findings of this analysis are meant to provide context into what Texans feel are the most significant

transportation issues affecting them, but the results are not meant to be representative of how Texans as a whole feel. The following are broad implications for transportation policy in Texas:

- Congestion: The public is likely to well receive policies and practices that mitigate rural and urban traffic congestion. Policies should take into consideration public perceptions of the causes of traffic congestion and use education and outreach to broaden public support.

- Gas prices: Drivers in Texas perceive the price of fuel as a significant issue. This perception is likely to fluctuate with the price of gas.

- Road maintenance: Texans who responded to the open-ended question indicated that they are the most passionate about the inconveniences created by road maintenance or lack thereof. Policies that mitigate the impacts of road maintenance and/or improve road conditions may alleviate the negative sentiments about these issues.

- Diverse, safe transportation options: Texans who responded to the open-ended question indicated that they are in favor of diverse, safe transportation options. Of the respondents who chose to discuss roadway safety and alternative modes of transportation, researchers noted positive sentiments toward the topics of adding bicycle and pedestrian facilities, expanding public transit opportunities, and improving road safety.

The Texas Transportation Poll also shows that the public's reception of measures to mitigate these issues will be tempered if additional revenue is needed to pay for the measures.

Policy Implications of Perceived Causes of Transportation Issues

The analysis of the perceived causes of transportation issues provides contextual examples of what respondents believed to be the causes of traffic congestion. This context allows researchers to provide decision makers with information about not simply what Texans think the problem is (traffic congestion in this case), but also what they believe to be the *causes* of traffic congestion. This unique insight allows policy makers to consider potential policy approaches to addressing congestion.

The policy implications of the insight gleaned from qualitative analysis may point to specific fixes (focusing on signal timing) or may point to efforts focused on education and awareness. The following is a summary of policy implications based on the findings of the top five perceived causes of traffic congestion:

- Road capacity: Many commenters attribute traffic congestion to inadequate road capacity, which could imply a need to better inform the public about the complexities of transportation planning and funding.

- Rush hour: The high number of comments about rush-hour traffic presents an opportunity for policy makers to mitigate the issue through policies that specifically address traffic

congestion during peak periods, such as staggered or flexible work schedules, telecommuting, ridesharing, and other trip-reduction policies.

- Construction: Traffic congestion due to road construction aggravates Texas drivers and costs time. Outreach and educational efforts to inform the public about the duration of construction projects or how to get more information could be effective in reducing the negative perceptions.

- Population growth: Accommodating new residents and the increased transportation demand they bring is a key challenge facing Texas leaders. Respondents agreed that a growing population is leading to increased traffic, but suggested differing perspectives on how best to address this challenge, implying that policies that manage transportation demand are needed in high-growth regions and could be accompanied by clear information about what government is doing to address the mobility impacts.

- Traffic signals: As shown in the comments about traffic congestion, the public thinks traffic signal timing should be a high-priority issue. Programs that address traffic signal timing, such as 311 and other systems for gathering citizen input on traffic signal issues at specific intersections, could be especially helpful in rural counties with smaller transportation departments.

Utility of Sentiment Analysis and Qualitative Analysis in Transportation Research

While the Texas Transportation Poll provides an understanding of Texans' views on transportation issues and solutions in aggregate, the verbatim responses from the open-ended question allow for a more nuanced understanding of not just *what* Texans think about these issues but *why*, in the exact language they use to describe their concerns.

Most public engagement methods and surveys target specific questions and issues to seek answers from the public. This helps ensure direct responses but leaves little opportunity for asking other questions after the instrument is administered. Though people naturally think and relay information through narrative and storytelling, these methods do not always match well with methods that seek to quantify input. Qualitative methods, including sentiment analysis, allow exploration of new research questions with a static dataset. With an expanded qualitative analysis toolbox, researchers can continue to explore public input from different angles and methods, subject to the limits of the responses. This enables a more responsive research process that allows more than one iteration of analysis of a dataset, where policy makers working with researchers can glean insights not possible from more quantitative research designs.

Qualitative and sentiment analysis methods broaden the ability of researchers to draw insights into the public's perspective on transportation issues. This analysis demonstrates the potential of qualitative analysis to categorize responses and find areas of consensus or disagreement within

public responses. This analysis also demonstrates the ability of sentiment analysis tools such as LIWC to parse out the meaning and emotion behind responses to general open-ended questions. Tracking these sentiments over time can reveal further insights into the changing attitudes of the public.

One limitation of this study was due to the phrasing of the open-ended question, "What is the most significant transportation issue affecting you personally in your region?" By asking respondents to report on the most significant issue affecting them, the question invites respondents to consider only what is bothering them about transportation in the region. More neutral phrasing of the original question could have resulted in an entirely different set of responses and sentiments. To maximize the ability of qualitative and sentiment analysis to inform transportation policy, several key elements should be considered when crafting future open-ended survey questions:

- Does the phrasing of the question steer respondents to offer positive or negative sentiments on a topic?
- Are researchers seeking answers to specific research questions, for example, solutions to pressing transportation issues? Incorporating these specific queries into the survey will help to elicit relevant answers.

In tandem with the results from the larger Texas Transportation Poll, qualitative analysis of open-ended questions provides additional insight into the freely expressed viewpoints of Texans, in their own words. Qualitative and sentiment analysis add valuable context to quantitative results. The power of qualitative and sentiment analysis of open-ended survey results can be leveraged in future Texas Transportation Polls to elicit responses that address specific questions held by decision makers, lead to further understanding of public sentiment on specific issues, and offer the public an opportunity to weigh in with ideas for how to solve the most pressing transportation issues facing them in their regions.

Appendix A: Texas Transportation Poll Sentiment Analysis Methodology

Qualitative Analysis Software

Researchers analyzed the volume and content of the comments with the help of NVivo, a computer-assisted qualitative data analysis software that enables easy organization and categorization of qualitative data. NVivo can help conduct analysis on a variety of qualitative data sources, including open-ended questions, using search, query, and visualization tools within the software.

For this effort, researchers used NVivo to conduct a frequency analysis of all comments and an in-depth analysis of the key themes found in the comments. Comments were first manually coded into categories, allowing researchers to map common threads and begin to see patterns in

the responses. After initial manual coding, the NVivo software auto-coded the remaining responses based on the existing coding pattern detected.

Grounded Theory

The next step in the qualitative analysis process was to review the responses to the open-ended question, "What is the most significant transportation issue affecting you personally in your region?" to find key themes. The method used for reviewing the responses is based on grounded theory. Rather than seeking answers to predetermined research questions, the data are first examined for patterns, from which further research questions and hypotheses might arise.

Following this method, researchers reviewed the responses to find key themes to serve as nodes for topical coding. A word frequency query in NVivo revealed basic patterns in the responses.

Categorization of Comments

Comments were next manually coded into categories. Two separate researchers manually coded a random sample of 100 responses to the open-ended question into categories, or nodes, using NVivo software. This allowed researchers to map the common threads and begin to see patterns in the responses. Building on the categories that emerged from the sample of 100 responses and the word frequency query, researchers categorized the responses into the following 10 topical nodes:

- Bicycle and pedestrian.
- Congestion.
- Construction.
- Drivers.
- Gas prices.
- Road maintenance.
- Planning.
- Public transportation.
- Safety.
- Toll roads.

These nodes focus on identifying transportation issues.

Topical coding of the remaining responses, after the first 100, was then conducted using text search queries in NVivo. For example, a query for the words "gas price," including synonyms, brought up 875 responses from 541 respondents, such as "fuel prices" and "price of gas is too high," which were then coded into the topical node for gas prices. This process was conducted for each of the 10 nodes. Responses that contained multiple references, for example "gas prices and congestion," were coded into more than one node.

Researchers then conducted manual coding and/or un-coding of responses into appropriate nodes to ensure accuracy. For example, because the word "toll" is a synonym for price, many responses regarding toll roads were incorrectly coded into the gas prices node. Researchers manually reviewed the nodes for inaccuracies and ran text search queries to find and remove unrelated responses in each of the 10 nodes.

The final step to coding the data was to create a category, or node, for each survey respondent. Each respondent's node contains all of the responses to the Texas Transportation Poll survey questions for that participant. This enables researchers to create cross-tabulations of survey results based on demographic information or the geographic location of participants. Key attributes for each respondent, including age, gender, and county, were linked to the respondent nodes using the node classification wizard in NVivo.

Using the same methodology to evaluate the entire population of comments, researchers then manually coded the first 100 responses within the traffic congestion category into sub-categories that suggest causes of traffic congestion. After initial manual coding, the NVivo software was able to auto-code the remaining responses based on the existing coding pattern detected.

Accuracy Testing

The next step was to conduct accuracy testing to review the internal validity of coding results and resolve conflicts as necessary. Conducting a coding comparison helps to minimize the potential for researcher bias when categorizing responses. Using the *Coding Comparison Query* function in NVivo, researchers developed two separate measures of agreement. The first is the Kappa coefficient, and the second is the comparison of the percentage agreement between the two researchers.

The Kappa coefficient is a statistical measure that takes into account the amount of agreement that could be expected to occur through chance (*3*). A Kappa coefficient less than 0.40 indicates poor agreement between researchers, 0.40 to 0.75 indicates fair to good agreement, and over 0.75 indicates excellent agreement between researchers (*3*). The Kappa coefficient was between 0.2829 and 0.964 for each node, with an average Kappa score of 0.6307, indicating a good level of agreement between researchers.

The percentage agreement between researchers is the specific number of units of agreement divided by the total units of data, displayed as a percentage (*3*). Whereas the Kappa coefficient focuses on statistical probability, the percent agreement focuses on the actual occurrences where researchers coded words within the same nodes. The measure of agreement between researchers averaged 93.29 percent, while disagreement averaged 6.70 percent, indicating a high level of reliability between researchers. Agreement between researchers was highest for the public transportation, gas prices, and road maintenance categories, indicating that responses in those topical areas were the most clear cut. Researcher agreement was lowest for the bicycle and pedestrian, congestion, and planning categories.

Sentiment Analysis

Once coded, the responses within each node were analyzed using LIWC software to determine their overall sentiment. LIWC is an empirically validated linguistic software that infers psychological/emotional reactions in written text. LIWC analyzes text on a word-by-word basis, as opposed to analyzing entire phrases, and then calculates the percentage of words that are positive or negative.

Appendix B: Literature Review

How Can Qualitative Methods Inform Transportation Policy?

Transportation planning has 20th century roots in quantitative methods, but planners are increasingly using qualitative methods to incorporate personal narratives as part of the process. Richard Willson contends that placing language and discourse at the core of transportation planning forms a basis for "communicative rationality" that has become a dominant method for combining sound transportation planning methods and public policy (*4*).

Narrative and storytelling are valuable means of understanding the public, but the traditional planning process has constrained this information to preconceived public engagement processes included as portions of planning processes. Most public engagement methods and surveys target specific questions and issues to seek answers from the public. This helps ensure direct responses but leaves little opportunity for asking other questions after the instrument is administered. Though people naturally think and relay information through narrative and storytelling, these methods do not always match well with methods that seek to quantify input.

Qualitative methods open the door to explore public language (from comments, social media, and other sources) to understand what the public says about transportation issues, regardless of whether they exist within a specific public engagement strategy or not. Sentiment analysis allows policy analysts to gain insight on the public to arrive at better solutions.

Reducing Bias of Content Analysis Using Grounded Theory

There are generally two main approaches to classifying qualitative information:

- Using categories or themes previously created by the researchers or others to begin grouping information.
- Using the language of the subjects to form the basis of categorization.

The latter approach is termed *grounded theory*, where the researchers use the subjects' language to dominate the categorization to reduce bias from researchers' perspectives (*5*).

Software Advancements for Mining Public Sentiment

Recent research in linguistics and database management has overlapped to form qualitative data analysis software that combines the power of computer algorithms with the common sense power of personal narrative. Gaber and Gaber (*6*) explain that this software improves the process of analyzing written language in three ways:

- Quickly summarizing public comments according to coded attributes.
- Searching text for topical relationships.
- Organizing topics of discussion based on respondent attributes, such as demographics.

These advantages allow development of databases that can be further mined for linguistic characteristics, such as relative positive or negative sentiment about certain issues. For example, a recent sentiment analysis about transportation in Austin found social media discussions concerning drunk driving to be associated negatively, while a new car-sharing service was the most positive topic (*7*).

A more recent study underscored the value of sentiment analysis in transportation by evaluating discourse on Twitter about transportation agencies, recommending agencies engage with the public through new media to have a positive influence on agency communication with the public (*8*).

The sentiment software used in this study (Linguistic Inquiry and Word Count) has been validated in terms of estimating emotional expression found in text from a range of subjects (*9*). Methods from the previous two studies serve to guide this study.

References

1. C. Simek and T. Geiselbrecht. *Texas Transportation Poll: Final Report*. PRC 14-16-F, Transportation Policy Research Center, College Station, Texas, 2014, pp. 1–51. http://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/PRC-14-16-F.pdf.

2. AAA. Texas Fuel Prices. AAA Daily Fuel Gauge Report, May 2015. http://fuelgaugereport.aaa.com/states/texas/.

3. NVivo. Run a Coding Comparison Query. http://help-nv9-en.qsrinternational.com/nv9_help.htm#procedures/run_a_coding_comparison_query.htm.

4. R. Willson. Assessing Communicative Rationality as a Transportation Planning Paradigm. *Transportation*, Vol. 28, No. 1, Feb. 2001, pp. 1–31.

5. C. H. Yu, A. Jannasch-Pennell, and S. Digangi. Compatibility between Text Mining and Qualitative Research in the Perspectives of Grounded Theory, Content Analysis, and Reliability. *The Qualitative Report*, Vol. 16, No. 3, 2011, pp. 730–744.

6. J. Gaber and S. Gaber. *Qualitative Analysis for Planning and Policy*. Planner's Press, American Planning Association, Chicago, Illinois, 2007.

7. J. S. Evans-Cowley and G. Griffin. Microparticipation with Social Media for Community Engagement in Transportation Planning. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2307, Dec. 2012, pp. 90–98.

8. L. Schweitzer. Planning and Social Media: A Case Study of Public Transit and Stigma on Twitter. *Journal of the American Planning Association*, Vol. 80, No. 3, Dec. 2014, pp. 218–238.

9. J. H. Kahn, R. M. Tobin, A. E. Massey, and J. A. Anderson. Measuring Emotional Expression with the Linguistic Inquiry and Word Count. *The American Journal of Psychology*, Vol. 120, No. 2, Jul. 2007, p. 263.

Authors

Ben Ettelman

Jody Trendler

Greg Griffin

Tina Geiselbrecht