TRB 17-04615

Enabling Transit Service Quality Co-monitoring Through a Smartphone-Based Platform

Corinna Li[1]
(347) 863-5765; clli@mit.edu

P. Christopher Zegras[1, 2]
(617) 452-2433; czegras@mit.edu
(Corresponding author)

Fang Zhao[2]
(65) 6516 8603; fang.zhao@smart.mit.edu

Zhengquan Qin[2]
(65) 6516 8603; zhengquan@smart.mit.edu

Ayesha Shahid[1]
(617) 909-7316; ayeshas@mit.edu

Moshe Ben-Akiva [3, 2]
(617) 253-5324; mba@mit.edu

Francisco Pereira [4]
(617) 253-9729; camara@transport.dtu.dk

Jinhua Zhao[1, 2]
(617) 324-7594; jinhua@mit.edu

[1]Massachusetts Institute of Technology
Department of Urban Studies and Planning
77 Massachusetts Avenue 7-346 Cambridge, MA 02139

[2]Singapore MIT Alliance for Research and Technology
1 CREATE Way, #09-01/02 CREATE Tower;
#01-13 Enterprise Wing, Singapore 138602

[3]Massachusetts Institute of Technology
Department of Civil Engineering
77 Massachusetts Avenue 1-290 Cambridge, MA 02139

[4]Technical University of Denmark (DTU)
Transport Department
115 Bygningstorvet, 2800 Kgs. Lyngby, Denmark

5,998 words + 1 figures + 5 tables = 7,498 words

Submitted for publication in the Transportation Research Record

March 15 2016

ABSTRACT
The growing ubiquity of smartphones offers public transit agencies an opportunity to transform ways to measure, monitor, and manage service performance. We demonstrate the potential in a new tool for engaging customers in measuring satisfaction and co-monitoring bus service quality. The pilot adapted a smartphone-based travel survey system, Future Mobility Sensing (FMS), to collect real-time customer feedback and objective operational measurements on specific bus trips. The system uses a combination of GPS, Wi-Fi, Bluetooth, and accelerometer data to track transit trips, while soliciting users' feedback on trip experience. While not necessarily intended to replace traditional monitoring channels and processes, these data can complement official performance monitoring through a more real-time, customer-centric perspective. The pilot operated publicly for three months on Boston's Silver Line (SL) bus rapid transit. Seventy-six participants completed the entrance survey, half of whom actively participated, completing over 500 questionnaires while on board, at the end of a trip and/or at the end of a day. Participation was biased towards frequent SL users, who were majority White and of higher income. Indicative models of user reported satisfaction reveal some interesting relationships, but the models can be improved by fusing the app-collected data with actual performance characteristics. Broader and more sustained user engagement remains a critical future challenge.

INTRODUCTION

The digital age is shifting customer preferences and urban mobility business models, posing challenges and opportunities for public transit operators and regulators. Today's transit riders, increasingly accustomed to the on-demand economy and a customer-centric user experience, are holding transit services to ever higher standards. Service providers need to pay more attention to proactively engaging customers to gather feedback and build more personal relationships (*1*). These evolving new norms are increasingly at odds with traditional methods and metrics for monitoring and evaluating transit service quality. Customer satisfaction surveys, though common in the industry, are administered infrequently and by-and-large ask subjects to provide only general, overall ratings. This not only reduces the possibility to pinpoint which trips are satisfactory or unsatisfactory – and what characteristics might influence this outcome – it also shortchanges the potential to use riders as higher-resolution sources of information on specific dimensions of experienced service quality, hindering agencies from obtaining more spatially and temporally precise results.

In recent years, technological advances and institutional changes have catalyzed the adoption of public service co-production (*2*). Government-citizen relationships are changing from paternalistic, provider-customer dynamics to more collaborative interactions (*3*). One type of co-production is service co-monitoring – agencies using public feedback to supplement official monitoring and regulation (*4*). The growing ubiquity of Internet-connected mobile devices and civic engagement platforms, makes citizens ever-more equipped to generate and submit feedback without time or geographic constraints (*5*). Smartphones epitomize the potential: smartphone penetration rates have been increasing steadily over the past decade, and in the USA transit riders' smartphone ownership tends to be higher than the national average (*6*). In theory, smartphones enable service providers to gather more dynamic, low-cost, and real-time details from the users' perspective. Providers' responsiveness and working partnerships with customers could also be enhanced. Customers would consequently benefit from improved service quality and strengthened sense of citizenship and empowerment.

This paper describes a pilot testing the capacity of transit service co-monitoring. In collaboration with the Massachusetts Bay Transportation Authority (MBTA), we engaged bus riders in Boston as co-monitors of service quality combining Bluetooth beacons with Future Mobility Sensing (FMS), a smartphone-based travel survey platform. A Bluetooth beacon continuously broadcasts its unique ID to the surrounding area, allowing for co-locating devices such as smartphones (*7*). FMS uses a combination of GPS, Wi-Fi, and cellphone accelerometer data to track user trips. Combining these two technologies, we aim to detect FMS app users' bus trips in real-time, survey them on their transit experience during or shortly after their trips, and collect objective bus operational data via smartphone sensors. An initial version of this approach was tested on a small scale in Singapore in 2015, after which technical and design improvements were implemented in the Boston pilot.

This pilot aims to test three key technical capabilities deemed necessary to crowdsource bus service monitoring and passenger satisfaction: (1) accurately and reliably detect a passenger's bus trip with minimum user input; (2) gather meaningful, high-resolution feedback at the level of a particular stage of an individual passenger's transit trip; and, (3) fuse collected data with actual vehicle performance measured through official automatic vehicle location (AVL) data, to associate service quality feedback with specific bus runs and operators.

MEASURING PUBLIC TRANSIT SERVICE SATISFACTION

Public transport literature commonly defines “quality of service” as “perceived performance from passengers’ point of view” (*8,9*). Service quality is, by definition, customer-centric, so assessing it typically involves customer satisfaction measurements (*10*). Hence, the term “customer satisfaction” has become synonymous with “service quality” in the public transport sector. Results of customer satisfaction surveys have likewise become the de facto service quality indicators for many public transport agencies (*11-13*). The exact dimensions that constitute service quality, however, vary considerably among scholars and transit organizations. Studies have included as few as six (*9*) and as many as 31 attributes (*14*); most fall between 8 and 22 attributes (*8,15-19*), including, but not limited to: accessibility, stop and station conditions, service frequency and reliability, onboard comfort, safety, and service rendered by staff

Within the transit industry, the most prevalent channels of gauging service quality are intercept surveys (onboard or at stops/stations), telephone surveys, and web-based surveys (*TABLE 1*). Many agencies use multiple, complementary channels to capture a wider respondent pool. Of the 27 U.S. transit agencies and Metropolitan Planning Organizations surveyed in 2006, three-quarters reported using intercept surveys, with two-thirds supplementing these with telephone interviews (*20*). Onboard and intercept surveys are especially advantageous in gaining direct access to transit customers and obtaining relatively representative samples (*21*). On the other hand, conducting intercept surveys for an extensive system can be resource-intensive (*22*). An increasing number of agencies have adopted web-based tools (*13,23)*. Web-based surveys have advantages in time and cost efficiency (*20,22*), but suffer biases against population with no, or limited, access to the Internet (*21*).

These conventional methods suffer from a difficulty in providing ubiquitous, detailed assessment and feedback on specific trips and their characteristics. Considering that public transport customer satisfaction surveys are often done annually (if that) (*20*), questionnaires often ask respondents for general service assessments (*24*). This puts regulators and system administrators at a disadvantage in obtaining high-resolution data – information that can reveal performance variations by driver, route, and time of day, and precise areas for commendation and targeted improvement (*25*). Traditional in-person questionnaires, administered during or after a ride, can also solicit trip-specific assessment, but their scopes are limited without a major commitment to staffing. Public transport experiences can vary from trip to trip, but deploying survey teams to every bus, stop, and station every day would be unrealistic.

Some agencies have gone beyond general service assessment by asking customers about specific trips (*23,26*). However, such recalled assessment may be inaccurate, as people’s actual and recalled travel experiences often differ due to psychological heuristics or unobserved events (*27-29*). This phenomenon implies that the sooner a rider has the chance to report her level of satisfaction, the more accurate the result. Recently, some public transport agencies have partnered with academics to attempt to shorten the lag between a user’s ride and subsequent survey solicitation, using new technological tools.

Recent attention has turned to smartphones as a new medium for data collection. Carrel et al (*24*) used an Android app (San Francisco Travel Quality Study) to examine the relationship between objectively measured service quality (e.g. travel and wait times) and transit riders’ satisfaction, emotions, and modal choice. During the month-long study, participants were asked to use transit service on at least five days and fill out the corresponding daily in-app surveys (for which they received a reminder every day). The app used Wi-Fi and cell tower positioning to automatically record the user's location information, which was then matched with AVL data to infer transit trips. The system generated multi-day data for understanding customers' ride experiences, with feedback at the daily rather than trip-specific level. Users taking more than one public transport trip on a given day were allowed to submit only one set of ratings, even though their experiences could vary widely from trip to trip. In contrast, Dunlop et al (*30*) used a BlackBerry OS-based survey app (TOES) to measure riders' emotional state before, during, and after each bus trip. Participants, however, needed to manually signal the beginning of their trip and begin each survey in the sequence; otherwise, they would be prompted by the app to complete the next stage of the trip every six minutes.

Other research examples have not been able to attribute reported satisfaction to specific trips and their service characteristics. Most studies focus largely on identifying attributes that most strongly influence riders' satisfaction with local public transport services, asking survey respondents to rate their satisfaction with local transit services; many also ask to rate each service attribute's importance (*15,16,31*). Others quantify the importance of attributes through stated preference experiments (*8,17,32,33*). While these studies have collectively identified many service attributes important to customer evaluation of transit service quality, they do not provide insights into the relationships between reported satisfaction and objectively measured service attributes -- wait time, onboard crowding, travel time, etc. Such insights could have important practical implications, helping transit operators and regulators improve service aspects directly related to measured customer satisfaction.

In light of the ongoing practical and research gap between linking reported satisfaction with measured transit performance characteristics, we developed a transit service quality measurement platform aiming to allow transit agencies to practically collect trip-specific user feedback in real-time. The trade-off between high information resolution and user burden is one key motivator of our effort, which leverages Bluetooth beacons to administer trip-specific surveys with minimal manual input from users. The platform was first piloted in Singapore in Summer 2015: a proof-of concept involving four Bluetooth beacon-equipped bus stops in downtown Singapore and 26 volunteers recruited from the government transportation authority. Using FMS, participants were automatically prompted to rate their bus experience while waiting at the stop, onboard a bus, or after alighting a bus. The pilot demonstrated the basic capability to collect trip-based feedback in real-time, but suffered from several weaknesses, including a high percentage of false positives in trip-detection, sampling bias, and inability to match responses to official trip records. The subsequent pilot, conducted in Boston, aimed to overcome some of those shortcomings.

FUTURE MOBILITY SENSING-TRANSIT QUALITY (FMS-TQ)

We designed FMS-TQ with three objectives: (1) automatically detect users' bus trips as they happen; (2) more accurately capture riders' transit experience with real-time surveying; and (3) explore relationships between customer's reported satisfaction with transit service on a specific trip and objectively measured performance metrics of that same trip.

FMS

FMS was developed for high-resolution travel surveys, to non-intrusively gather individual data on trips, modes, routes, and activities (*34*). Its original system architecture contains three components: a smartphone app, backend, and web interface.

The smartphone app (iOS or Android) continuously collects location (GPS, WiFi, GSM), accelerometer data, and some other information (e.g. battery level). FMS can produce detailed and rich data for travel surveys, based on sensing what people do, rather than asking people to report what they do. This intends to eliminate problems with traditional self-reported surveys, such as under-reporting short trips and reporting inaccurate locations and times (*34*). FMS attempts to capture all relevant trip and activity data and upload the data to the backend when the phone is charging. The backend server receives and stores the data and turns them into travel and activity information (stops, trips, modes) using machine learning algorithms. These data then feedback to a web interface that presents the user with a sketch of her day for validation through a map and an editable timeline with sequences of stops and trips and relevant characteristics. FMS was piloted in Singapore in 2012 and has since been commercialized and is being deployed in several metropolitan areas around the world for household travel surveys.

FMS-TQ builds on the core app and database, but involves an additional Bluetooth beacon component. These beacons, as elaborated below, enhance the FMS platform’s real-time transit trip detection without excessively compromising phone battery performance.

Bluetooth Low-Energy (BLE) Beacon Extension

The BLE beacon is a small device that leverages low power Bluetooth transmission to broadcast packets of data, at regular intervals. Smartphones, with Bluetooth enabled and within the beacon signal range, receive the data, which include the beacon’s unique ID and the distance between the phone and the beacon. Our Boston pilot used this micro-local positioning function by installing beacons on buses and at bus stops. When a user of the FMS-TQ app enters or leaves a bus stop beacon’s signal range, the app registers the arrival/departure times. Likewise, onboard beacons indicate when a user has boarded a bus, is travelling on a specific bus, or has alighted. Given this ability for real-time bus trip detection, FMS-TQ can be programmed to solicit user feedback in a variety of formats at any point during or after the bus trip. Each beacon’s unique ID further allows for matching the survey response to a particular bus. The app records the information of when the user is first detected at the stop, providing the possibility to calculate the passenger’s wait time, by subtracting the at-stop timestamp from the bus-boarding timestamp. Since BLE does not require any form of Internet connectivity, this information, together with user responses to surveys, can be recorded on smartphones without Internet access, thereby extending the capability to include participants without active data plans (with data uploaded when Internet access is available via WiFi).

In the Boston pilot, weatherproof beacons were affixed at the top of bus stops and behind a panel near the ceiling of the buses. This set-up created a direct line-of-sight between the beacons and users’ smartphones, maximizing signal transmission. The particular brand of beacons used in this study cost approximately USD 20 each, and have a manufacturer-reported battery life of two years.

Compared to pure smartphone sensor-based methods, pairing smartphones and beacons theoretically brings several advantages for automatic transit trip detection. Relying solely on a phone's sensors (GPS, WiFi, cell tower signals, and accelerometer), would require that the detection app frequently match the user's geolocation with the transit network on the server. This process is computationally and battery-intensive; subsequent battery drain would likely quickly dissuade users from participating. Furthermore, phone-only detection infers speed of travel – and looks for speed patterns resembling bus travel – from changes in detected locations over time. Hence, any disruption in the strength or accuracy of locational signals may compromise phone-only detection. Given the beacon's efficient interaction with smartphones and continuity in transmitting close-range signals, incorporating beacons attempts to overcome the two main challenges associated with phone-only trip sensing.

BOSTON PILOT

The Boston pilot – named QualiT – was implemented in close partnership with the Massachusetts Bay Transportation Authority (MBTA), the public operator of most bus, subway, commuter rail, and ferry routes in greater Boston (Massachusetts). The pilot ran on Boston's Silver Line (SL), a bus rapid transit system with a total of four routes and a daily ridership of over 33,000 on average weekdays (*35*). The SL plays an important role in Boston's urban transportation, serving Boston's historic downtown and financial districts, the predominantly Black and Hispanic Roxbury neighborhood, the new Seaport Innovation District, and Logan International Airport. All four SL branches rank in the top 20 bus routes by typical weekday ridership (*35*).

The SL also provided a close-to-ideal setting in terms of implementation logistics. The best way to associate survey responses to a specific vehicle is to equip all participating buses with a Bluetooth beacon. In contrast to the rest of the MBTA's bus routes, which run on an interlining basis without designated vehicles, the SL operates 56 special buses dedicated exclusively to the service. Furthermore, almost all SL stops and stations have dedicated infrastructures (e.g., shelters), facilitating secure beacon installation. We were hence able to conduct the pilot with only 100 beacons, covering 56 vehicles and 44 stops.

Survey Design

FMS-TQ has an entrance survey and three event-based surveys. To mitigate user burden, each survey questionnaire consists of only a few multiple choice questions, designed for easy display on a smartphone screen (Figure 1) and delivered via the in-app local notification system. A simple sampling algorithm reduced the number of on-board questionnaires a user would receive on a given day if she takes multiple bus trips. A description of survey content and implementation follow.

Entrance questionnaire. After installing the app on her phone, the user receives an intake survey to collect baseline information, including her mobility options, general satisfaction with transit, and demographic information.

End-of-trip questionnaire. When a user enters a bus beacon's signal range and exits from the same range after at least 90 seconds, the app assumes the user has just completed an SL trip. At this time, the app prompts the user, on a single screen, to rate overall service and four attributes: wait experience, travel time, comfort and cleanliness. Also on the screen: a text box allowing for optional comments, and an option to report false positives – i.e., that the user did not actually ride on the SL.

Onboard questionnaire. On select occasions, a user would receive an onboard questionnaire which first verifies that the user is travelling on the bus. If so, it asks about 1) purpose of the trip, 2) satisfaction with each of five service attributes – cleanliness of the stop, wait experience, onboard crowdedness, onboard comfort, and driver's service, and 3) reasons for dissatisfaction, if any.

End-of-day questionnaire. This survey is solely solicited from users who have been detected at an SL stop but never boarded a vehicle. We incorporated this "safety net" to capture the experience of customers who might have waited at a bus stop and left without boarding, and those who actually rode the bus but whose bus trip was missed by the detection algorithm. To generate this survey, the platform combs the backend data at 8pm each day, and pushes a questionnaire to identified users. To minimize user burden, if a user completed an end-of-trip and/or onboard questionnaire that day, she would not receive the end-of-day questionnaire.

FMS-TQ also records the timestamp when the survey was presented and the timestamp when the user loads the survey on her device. The difference between these two timestamps may be large enough that the respondent might be susceptible to memory distortion effects (*24*), so FMS-TQ tries to mitigate these effects by incorporating contextual information in the questionnaires. For example, the End-of-trip and End-of-day questionnaires include the start time of the relevant trip. The app also periodically collects the phone's GPS, WiFi, GSM, and accelerometer readings. The latter are used by the backend server-based machine learning to determine trip characteristics and transport mode. This information can potentially be used to infer onboard comfort levels and study the impact of access and egress modes on travel satisfaction.

Pilot Implementation

The three-month pilot began on April 15, 2016. We marketed the project through a wide range of channels, including MBTA and MIT media, outreach to transportation management associations, employers and prominent organizations in areas served by the SL. Recruitment messages appealed to people's intrinsic motivations to give feedback to help the MBTA improve service and to the fun and innovative nature of the act ("you can now rate your bus experience just like you can with Uber or Lyft rides!"). The outreach messages also highlighted the financial incentive: for every questionnaire completed, the user would be automatically entered into a monthly sweepstake to win one of three available MBTA monthly passes.

Prospective participants were directed to the initiative's website, where they could learn more, sign up for an account, and proceed to install the FMS app from Google Play (Android) or AppStore (iPhone). Once logged into the app, participants received the entrance questionnaire and would otherwise be surveyed automatically according to the protocol described above.

In addition, since FMS-TQ relies heavily on BLE technology for real time bus trip detection, all survey participants received instructions to activate Bluetooth and daily notification reminders to enable Bluetooth functionality if it was disabled.

FINDINGS

Over the three-month pilot period, the recruitment website registered over 1,500 unique visits; 136 people signed up, and 76 participants completed the entrance questionnaire. Half of those, 38 users, actively participated, completing 346 end-of-trip, 135 onboard, and 21 end-of-day questionnaires. Recall that users only receive an end-of-day questionnaire when they are detected at a stop but never received an onboard of end-of-trip questionnaire on that day; thus, the small number of end-of-day surveys generated is a desirable outcome, indicating that most users who waited at a bus stop were subsequently detected on a bus. Based on user action (reporting of “false positives”), the app performed better in correctly detecting a trip had been taken (5% false positive rate reported for end-of-trip questionnaires) than that a user was on-board (13% false positive rate for on-board questionnaires). This could be a result of the user not wanting to respond to a questionnaire and simply choosing to report an error instead. Due to an engineering design decision, we were unable to estimate the response rate (i.e., number of questionnaires delivered to a user but not responded to).

About two-thirds of the active participants completed five or fewer trip-based questionnaires each; a quarter completed 20 or more questionnaires each. The average response time (between questionnaire delivery and user response) was 17 minutes for the onboard survey and 50 minutes for the end-of-trip survey. Relative to those who registered, but never participated, active participants were more frequent SL users, had slightly higher access to a private car, and more were White. Overall, active participants skew towards being young, White, frequent SL users, from higher income households; the majority (60%) report being dissatisfied, in general, with their MBTA bus experience (Table 2).

In aggregate, participants positively rate stop cleanliness, driver’s service, and the overall experience, while they tend towards neutral in reported satisfaction for the wait, onboard crowding, and onboard comfort, with crowding and waiting receiving a notable share of high dissatisfaction (Table 3). To associate these responses to a particular trip and its characteristics we capitalize on the unique data generated and used by FMS-TQ: user ID, bus stop ID, bus ID, times of survey solicitation and response, and the user’s phone’s entries and exits from any FMS-TQ beacon range with these events’ corresponding timestamps. These data allow for reconstructing key characteristics of a user’s trip (Table 4) and matching that trip to the completed questionnaire. With the exception of the alighting stop, we were able to identify most of the associated trip characteristics well – between 87% and 99% for end-of-trip questionnaires, 79% and 99% for onboard questionnaires. The relative difficulty in identifying the alighting stop likely reflects the phone’s inability to detect the stop’s beacon signal when the user quickly departs the area.

Despite the small and certainly biased sample, we use the results in some initial models to explore variables explaining the various dimensions of participants’ reported satisfaction – overall service, waiting at stop, onboard comfort, onboard crowding, and driver’s service. At the time of this writing, we have only been able to incorporate the respondents’ wait time and “observed” in-vehicle travel time (IVTT) with the latter calculated as the difference between the stop destination timestamp and boarding timestamp. IVTT could not be calculated for all responses. We are not yet able to associate the survey responses with additional performance characteristics (derived from the MBTA’s AVL data).

We separately model the five reported satisfaction ratings using a mixed-effect ordered logit model, with random effects applied to individual participants. Table 5 presents the “best” specifications, after exploring various combinations of explanatory variables and with and without “observed” IVTT. The results reveal that only estimated wait time impacts the overall satisfaction level. For waiting satisfaction, wait time is also negatively correlated, as is AM peak, understandable as users may be anxious to get to work. Interestingly, PM peak has a positive relationship with waiting satisfaction, perhaps due to lower relative end-of-day anxiety. Low and middle income corresponds with lower levels of satisfaction with waiting, as does having access to a bike; perhaps bikers, used to door-to-door on-demand mobility, have less tolerance for waiting. For satisfaction with onboard comfort, user-reported onboard crowding has a statistically significant, negative, relationship, a logical result; middle income respondents, interestingly, report higher satisfaction; again, those with bike access report relative dissatisfaction. For satisfaction with onboard crowding, this is worse during AM peak, for low and middle income respondents and, again, for those with bike access; IVTT is negative, but only significant at slightly greater than 80% confidence. Finally, reported satisfaction with the driver's service was higher in both peaks – perhaps users empathize with drivers during this noticeably difficult time; users on the Washington Street SL segment also reported lower levels of satisfaction with the driver's service, perhaps due to more erratic and congested conditions associated with this part of the SL which lacks segregated lanes and meaningful bus priority measures. A similar negative association emerges for those with access to a bike or car, perhaps reflecting the preferred relative control such modes confer. For waiting, onboard crowding and driver's service, satisfaction levels tend to be lower for end-of-trip surveys; ex-post reflection might negatively influence users' opinion.

DISCUSSION

The public pilot demonstrates the viability of FMS-TQ, but also reveals weaknesses of the beacon technology and the general approach. First, signal detection accuracy relies on beacon signal strength and broadcasting frequencies, but attenuation by the physical environment and interference from nearby beacons make it challenging to find a configuration that can account for the many different scenarios within a transit system. For example, the bus beacons were installed behind a fiberglass panel, which, along with human bodies (e.g., crowding), weakens the broadcasted signal. But setting signal strength higher to overcome attenuation can cause beacons from nearby buses or stops to interfere with detection accuracy. Bus-bunching, a frequent SL problem, exacerbates the challenge. Despite much testing, the pilot's beacon configuration remained imperfect. The data collected and participant comments show that the app missed some SL trips due to weak or inconsistent signals. A substantial number of observations lack beacon signals representing boarding and alighting, hindering inference of trip characteristics. Additional data gaps may have been due to user error, such as participants not remembering to turn on their phone's Bluetooth before beginning their trip.

In terms of the survey, we do not know the timeliness of the users' response vis-à-vis the actual trip (time between delivery and response); we also do not know the non-response rate (i.e., questionnaire delivered, but no response). Consider the number of surveys delivered: if a user decides to ignore a questionnaire, the easiest thing to do would be to just press "submit"; the system would register it as a blank completed questionnaire (we received 17 blank onboard responses; 0 blank end-of-trip responses). Some people may have received questionnaires and ignored them. Future system design should incorporate a method to capture the true response rate.

As for testing the idea of customer-driven service co-monitoring, the pilot reveals three challenges. First, people apparently do not feel compelled to sign up for an initiative and download an app solely to give feedback on their bus rides. More meaningful external incentives will likely be crucial at motivating participation. Relatedly, sustaining participation is hard. Precedent studies committing substantial rewards to participants (a free monthly pass, for example) saw participation rates drop off drastically once users satisfy minimum usage requirements (*24*,*30*). Lastly, without targeted outreach and technological enablement (such as providing smartphones with adequate data plans), the sample will likely not represent the general ridership. The overwhelmingly White respondent rate was particularly surprising in this pilot. Smartphone-based civic engagement around public transit remains far from "mainstream."

Numerous short- and longer-term research opportunities exist. The pilot generated unique, real-time customer data, and brought together traditionally segregated data sources for the first time: customer reports of service quality, observed trip characteristics, participants' locations, and AVL-based operational information. These datasets overlap to various degrees, giving rise to redundancy for triangulation and validation. So far, we have only explored a relatively small set of possibilities. Some areas for further exploration include: incorporating a variable representing deviations of observed IVTT vis-à-vis scheduled IVTT (*24*); using accelerometer data collected by FMS to reveal relative smoothness of the bus ride; further analyzing MBTA's data for vehicle headway information; and triangulating origin-destination inference with GPS data. The models can be enhanced by: fusing the AVL service measures; examining users' estimated total daily travel and activity patterns' impacts on public transit satisfaction (FMS generates these data as long as the user had it continuously enabled); and adding relevant environmental conditions, particularly weather, which can be derived from historical records for date/timestamps (we did not use the temperature data from the sensor-equipped beacons due to programming constraints). Overcoming the low participation rate also merits additional research. Finally, questions can be raised as to the value added from this approach relative to "traditional" customer satisfaction measurement approaches. The models estimated show that end-of-trip surveys (most akin to "traditional" approaches) are marginally different for satisfaction measures of waiting, on-board crowding, and driver service satisfaction. These results might further change if additional, trip-specific attributes are added (e.g., from AVL); but future research could examine what matters most in terms of user reported satisfaction and overall opinions on transit service.

CONCLUSIONS

The QualiT pilot deployed a location-aware app that enables real-time engagement with transit customers. The approach successfully surveyed users during or immediately after bus trips and enabled those trips to subsequently be matched with trip performance characteristics. The pilot opens up many possibilities for capturing new insights on users' bus trips. Despite the technological success, the pilot involved few, and non-representative, participants over its three months. Indicative models of user reported satisfaction reveal some interesting relationships; for example, users with access to bikes tend to report lower satisfaction with the various performance dimensions assessed; during AM and PM peak, users report higher satisfaction with drivers' service. The results reveal little difference between on-board versus end-of-trip survey responses.

The pilot will hopefully lead to further innovations towards ultimately enriching the meaning of the word "service" in public transport services and enabling the capacity to fulfill that service. Smartphone-based feedback has practical merits and shortcomings. Ultimately, the success of a user-feedback platform must be measured by its ability to gain traction among customers and generate meaningful data for service operators. Time will tell whether the potential can be realized.

ACKNOWLEDGEMENTS
This research was supported by the Singapore National Research Foundation under the Singapore MIT Alliance for Research and Technology's Future Urban Mobility research group. We thank the MBTA, MassDOT, and Massport for their support; Andre Carrel for insights and advice; and (alphabetically) Rui Baltazar, Inês Dias, Paula Garbagnoli, Ajinkya Ghorpade, William Ko, Kalan Nawarathne, Jorge Santos, and Bruno Santos, for helping to make FMS a reality.

REFERENCES

1. Prahalad C.K., and V. Ramaswamy. Co-creation experiences: The next practice in value creation. *Journal of Interactive Marketing*, 18(3), Jan. 2004, pp. 5–14.
2. Clark B.Y., J.L. Brudney, and S-G Jang. Coproduction of Government Services and the New Information Technology: Investigating the Distributional Biases. *Public Administration Review*. Sept. 2013, 73(5), pp. 687–701.
3. Ostrom E. Crossing the great divide: Coproduction, synergy, and development. *World Development*. June 1996, 24(6), pp. 1073–87.
4. Kaufman S. Co-Monitoring for Transit management -- Using Web-Based Input for Transit Management. NYU Rudin Center, Feb. 2014. https://wagner.nyu.edu/rudincenter/wp-content/uploads/2014/02/CoMonitoringForTransit_web.pdf. Accessed Oct. 7, 2015.
5. SeeClickFix. *SeeClickFix | Distributed 311 for Distributed Governance*. SeeClickFix, 2010. https://gov.seeclickfix.com/2010/01/09/distributed-311-for-distributed-governance. Accessed Mar. 15 2016.
6. Windmiller, S. H, T., and K.E. Watkins. Accessibility of Communication Technology and the Rider Experience Case Study of Saint Louis. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2415, 2014, pp. 118–26.
7. Harman C. *iBeacons and BLE*. Calvium, 2014. https://calvium.com/2014/04/ibeacons-and-ble. Accessed May 19 2016.
8. Hensher DA, P. Stopher, and P. Bullock. Service quality—developing a service quality index in the provision of commercial bus contracts. *Transportation Research Part A: Policy and Practice*. July 2003, 37(6), pp. 499–517.
9. Kittelson & Associates, editor. *TCRP Report 165: Transit capacity and quality of service manual. 3rd ed.* Transportation Research Board, Washington, D.C., 2013.
10. Eboli L, and G. Mazzulla. A New Customer Satisfaction Index for Evaluating Transit Service Quality. *Journal of Public Transportation*, 2009, 12(3), pp. 21 - 37.
11. New York City Transit. *2014 Customer Satisfaction Survey*. 2014. http://web.mta.info/mta/news/books/docs/2014-Subway-CSS-Board-Presentation.pdf. Accessed June 20, 2015.
12. SFMTA. *Muni On-Board Customer Survey*. Mar. 2014. https://www.sfmta.com/sites/default/files/agendaitems/3-14-14%20Item%2011%20Muni%20On-Board%20Customer%20Survey.pdf. Accessed June 20, 2015.
13. NJ Transit. *NJ TRANSIT Customer Satisfaction Survey*. 2015. http://www.njtransit.com/var/var_servlet.srv?hdnPageAction=ScoreCardSurveyTo. Accessed May 25 2015.

14. Habib KM.N., L. Kattan, and M.T. Islam. Model of personal attitudes towards transit service quality. *Journal of Advanced Transportation*, Oct. 2011, 45(4), pp. 271–85.
15. Eboli L., and G. Mazzulla. Service Quality Attributes Affecting Customer Satisfaction for Bus Transit. *Journal of Public Transportation,* 2007, 10(3), pp. 35–46.
16. Eboli L., and G. Mazzulla. How to Capture the Passengers' Point of View on a Transit Service through Rating and Choice Options. *Transportation Reviews,* July 2010, 30(4), pp. 435–50.
17. Eboli L., and G. Mazzulla. A methodology for evaluating transit service quality based on subjective and objective measures from th passenger's point of view. *Transportation Policy,* 2011, 18, pp. 172–81.
18. European Committee for Standardization. *European Standard EN 13816. Transportation - Logistics and services - Public passenger transport - Service quality defintion, targeting and measurement*. 2002.
19. Friman M., and M. Fellesson. Service supply and customer satisfaction in public transportation: The Quality Paradox. *Journal of Public Transportation*, 2009, 12(4), pp. 57–69.
20. Spitz G., F.L. Niles, and T.J. Adler. *TCRP Synthesis 69: Web-based survey techniques: a synthesis of transit practice*. Washington, D.C, Transportation Research Board; 2006.
21. Schaller B, National Research Council. *On-board and intercept transit survey techniques,* Transportation Research Board, Washington D.C., 2005.
22. Chow W. Evaluating Online Surveys for Public Transit Agencies Using a Prompted Recall Approach (thesis). Massachusetts Institute of Technology, 2014. http://hdl.handle.net/1721.1/89853. Accessed June 2 2015.
23. Boroyan N. *Care to Take the MBTA Customer Opinion Survey?* BostInnov, Feb. 23 2014. http://bostinno.streetwise.co/2014/02/23/take-the-mbta-customer-opinion-panel-signup-survey-boston-transit. Accessed April 3, 2015.
24. Carrel A., R.G. Mishalani, R. Sengupta, and J.L. Walker. In Pursuit of the Happy Transit Rider: Dissecting Satisfaction Using Daily Surveys and Tracking Data. *Journal of Intelligent Transportation System*, July 3 2016, 20 (4), pp. 345–62.
25. Elmore-Yalch R. *A handbook: integrating market research into transit management*. National Academy Press, Washington D.C., 1998.
26. Land Transport Authority. *Public Transport Customer Satisfaction Survey 2015 - Annex-A*. http://www.lta.gov.sg/data/apps/news/press/2016/20160309_PTCSSAnnex-A.pdf. Accessed April 15, 2016.
27. Fredrickson B.L., and D. Kahneman. Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social* Psychology, 1993, 65(1), pp. 45–55.
28. Pedersen T., M. Friman, and P. Kristensson. The role of predicted, on-line experienced and remembered satisfaction in current choice to use public transport services. *Journal of Retailing and Consumer Services*, Sept. 2011, 18(5), pp. 471–5.
29. Abou-Zeid M., R. Witter, M. Bierlaire, V. Kaufmann, and M. Ben-Akiva. Happiness and travel mode switching: Findings from a Swiss public transportation experiment. *Transportation Policy*, Jan. 2012, 19(1), pp. 93–104.
30. Dunlop I., J. Casello, and S. Doherty. Tracking Transit Rider Experience: Using Smartphones to Measure Comfort and Well-being Throughout the Trip. Presented at 94th Annual Meeting of the Transportation Research Board, Washington, D.C, 2015.

31. Oña J., R. Oña, L. Eboli, and G. Mazzulla. Perceived service quality in bus transit service: A structural equation approach. *Transportation Policy,* Sept. 2013, 29, pp. 219–26.
32. Eboli L., and G. Mazzulla. A Stated Preference Experiment for Measuring Service Quality in Public Transport. *Transportation Planning and Technology,* Oct. 2008, 31(5), pp. 509–23.
33. Prioni P., and D. Hensher. Measuring service quality in scheduled bus services. *Journal of Public Transportation*, 2000, 3(2), pp. 51–74.
34. Zhao F., F.C. Pereira, R. Ball, Y. Kim, Y. Han, C. Zegras, et al. Exploratory Analysis of a Smartphone-Based Travel Survey in Singapore. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2494, 2015, pp. 45–56.
35. MBTA. *Ridership and Service Statistics, Fourteenth Edition*. 2014. http://www.mbta.com/uploadedfiles/documents/2014%20BLUEBOOK%2014th%20Edition.pdf. Accessed May 13 2015.

LIST OF TABLES AND FIGURES

TABLE 1 Service Quality/Customer Satisfaction Survey Methods Used By Top 10 U.S. Transit Systems

Agency	Most recent survey (publically available)	Intercept	Web-based	Phone	Other
NYCT	2014 Customer Satisfaction Survey			X	
CTA	2013-2014 RTA Customer Satisfaction studies	X	X		
LACMTA	Spring 2015: Metro Rail/Bus/System-wide Customer Satisfaction Survey	X		x	Focus groups
WMATA	Voice of the Customer survey (1st Quarter, FY 2013)			X	Mystery rider
MBTA	Customer Opinion Panel		X		
SEPTA	2012 Customer Satisfaction Survey	X		X	
NJTransit	ScoreCard (2nd Quarter, FY2016)		X		
MUNI	2013 MUNI On-board Customer Survey	X	X		
MARTA	2014 Quality of Service Survey	X			
MTABUS	Unknown				

**Table 2 Summary Statistics from Entrance Survey:
All Respondents versus Active Participants**

Weekly Silver Line usage	All	Active
Less than once	20%	13%
One or two times	16%	11%
Three or more times	64%	76%
Access to private car	**All**	**Active**
Yes, always	39%	47%
Sometimes	20%	18%
Never	41%	34%

Own bike or bikeshare membership	All	Active
Yes	29%	29%
No	71%	71%
Satisfaction with MBTA bus service	**All**	**Active**
Very dissatisfied	7%	5%
Dissatisfied	14%	13%
Somewhat dissatisfied	38%	42%
Neutral	30%	32%
Somewhat satisfied	9%	5%
Satisfied	1%	3%
Very satisfied	0%	0%

Gender	All	Active
Woman	42%	47%
Man	55%	53%
Prefer not to say	3%	0%
Age	**All**	**Active**
< 18	0%	0%
18 to 21	4%	3%
22 to 34	52%	55%
35 to 44	14%	16%
45 to 64	28%	26%
65 +	0%	0%
Prefer not to say	0%	0%
Race	**All**	**Active**
Amer. Indian/Alaska Native	5%	5%
Asian	12%	8%
Black of African American	9%	0%
Native Hawaiian/Pacific Isl.	0%	0%
White	67%	79%
Other	6%	8%
Prefer not to say	7%	0%
Household income	**All**	**Active**
Less than $14,500	6%	3%
$14,500 to $28,999	6%	3%
$29,000 to $43,499	7%	8%
$43,500 to $75,999	14%	13%
$76,000 to $108,499	20%	21%
$108,500 to $151,999	14%	18%
$152,000 or more	10%	11%
Prefer not to say	22%	24%

Table 3 Summary Statistics of Reported Satisfaction (1=very dissatisfied; 4=neutral; 7=very satisfied)

	Stop cleanliness	Wait		Onboard crowding		Onboard comfort		Driver service		Overall
	Onboard	*Onboard*	*Alight*	*Onboard*	*Alight*	*Onboard*	*Alight*	*Onboard*	*Alight*	*Alight*
1	3%	13%	13%	19%	16%	10%	9%	1%	3%	7%
2	3%	3%	9%	7%	12%	9%	11%	2%	2%	6%
3	11%	15%	8%	4%	7%	6%	11%	3%	3%	11%
4	25%	19%	20%	15%	20%	22%	26%	37%	38%	26%
5	11%	10%	15%	15%	18%	15%	15%	14%	18%	14%
6	19%	12%	21%	10%	17%	13%	19%	13%	21%	27%
7	18%	18%	14%	20%	11%	15%	9%	21%	16%	9%
No response	10%	10%	0%	10%	0%	10%	0%	10%	0%	0%

Table 4 Success Rate in Identifying Relevant Trip Characteristics via Beacons

Trip Characteristic	End-of-Trip Survey	Onboard Survey
1. Boarding stop	89%	86%
2. Timestamp for arrival at boarding stop	87%	81%
3. Bus vehicle #	99%	99%
4. Timestamp for boarding bus	99%	96%
5. Wait time (4-2)	97%	79%
6. Alighting stop	69%	62%
7. Time alighting from bus	99%	96%

Table 5 User Reported Satisfaction: Mixed-Effect Ordered Logit Results

Variable	**Overall**		**Overall (IVTT)**		**Waiting**		**Onboard comfort**		**Onboard crowding**		**Onboard crowding (IVTT)**		**Driver's service**	
	Coef.	p-val	Coef.	p-val	Coef.	p-value	Coef.	p-val	Coef.	p-val	Coef.	p-value	Coef.	p-val
Wait time	-0.13	0.005 ***	-0.122	0.007 **	-0.19	<0.000 ***								
IVTT			-0.014	0.55							-0.028	0.184		
AM commute					-0.09	<0.000 ***			-0.89	0.009 ***	-0.86	0.011 *	0.963	0.011 **
PM commute	0.572	0.486	0.448	0.09	0.373	<0.000 ***	0.324	0.184	-0.42	0.233	-0.36	0.311	0.900	0.025 **
Washington St Route													-1.30	0.035 **
Peak direct.														
Reported crowding							5.939	<0.000 ***						
Income-low					-1.1	0.062*	0.752	0.372	-1.76	0.057*	-1.85	0.046*		
Income-medium					-0.75	<0.000 ***	0.763	0.077 *	-0.94	0.070 *	-0.97	0.06*		
Bike access	-0.83	0.127	-0.972	0.101	-0.65	0.02 **	-0.669	0.108	-1.82	0.01 ***	-1.84	0.006**	-2.29	0.01**
Car Access									-0.97	0.136	-0.99	0.125	-2.28	0.01 **
Woman														
End-of-trip survey					-0.41	<0.000 ***			-0.50	0.109	-0.577	0.06*	-0.51	0.121
# of obs.	267		255		313		313		313		299		313	
Log lik. (m)	-438.4		-417.5		-566.2		-444.9		-569.5		-543.02		-443.6	
Log lik. (n)	-470.7		-461.7		-582.5		-587.4		-594.8		-567.26		-430.7	
AIC (model)	896.8		856.94		1160.5		923.8		1166.9		1116.05		913.2	
AIC (null)	953.5		917.08		1178.9		1186.8		1201.6		1146.5		1011.0	
Pr(>Chisq)	0.000 ***		0.000 ***		0.000 ***		0.000 ***		0.000 ***		0.000 ***		0.000 ***	

*** = significant at 0.01; ** significant at 0.05; * = significant at 0.1

Grey cells indicate variables not included in the final model specification for the particular dependent variable.

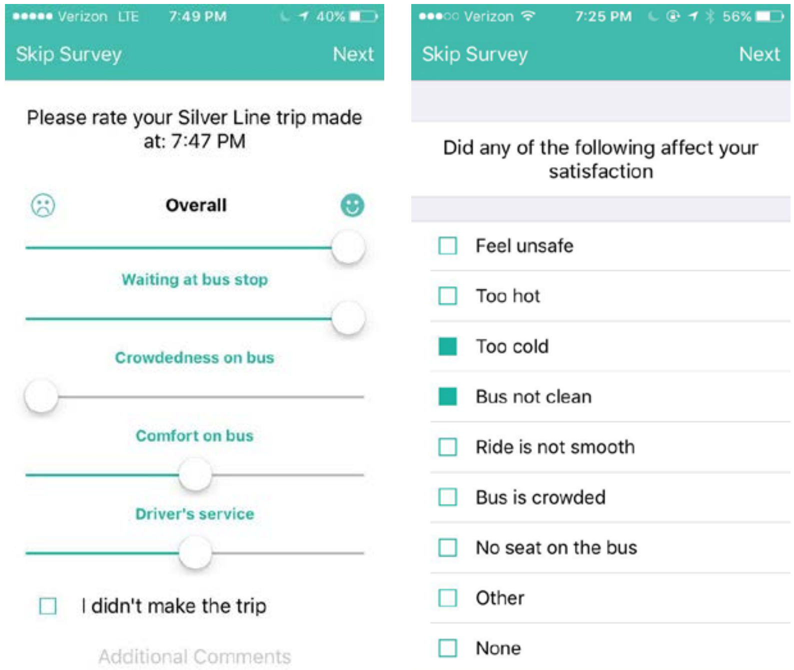

Figure 1 Screenshot example of QualiT survey.