FMS-TQ: combining smartphone and iBeacon technologies in a transit quality survey

Corinna Li[1, 3]
(347) 863-5765; clli@mit.edu
(Corresponding author)

P. Christopher Zegras[1]
(617) 452-2433; czgras@mit.edu

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

Francisco Pereira[3, 4]
(617) 253-9729; czgras@mit.edu

Kalan Vishwanath Nawarathne[2]
(65) 6516 8603; kalan@smart.mit.edu

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

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

Jinhua Zhao[1]
(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
2800 Kongens-Lyngby, Denmark

6,586 words + 5 figures + 4 tables= 8,836 words

Submitted on August 1, 2015

ABSTRACT

The Internet of Things (IoT) will offer transit agencies an opportunity to transform ways to measure, monitor, and manage performance. We demonstrate the potential value of two combined technologies, smartphones and iBeacons, for actively engaging customers in measuring satisfaction and co-monitoring bus service quality. Specifically, we adapt our smartphone-based survey system, Future Mobility Sensing (FMS), to connect with iBeacons for an event-driven approach to measure user-reported satisfaction before (i.e. at the stop), during (i.e., while traveling), and after (reflectively) transit trips. The system collects a combination of sensor (GPS, WiFi, GSM and accelerometer) data to track transit trips, while soliciting users' feedback on trip experience with in-app pop-up surveys. Both bus trip data and passenger feedback are collected and uploaded onto the server at the end of each day. These data are not intended to replace traditional monitoring channels and processes, but, rather, they complement official performance monitoring through a more customer-centric perspective in relative real-time. The paper presents the theoretical foundations, describes a pilot implementation of the platform in Singapore, and discusses preliminary results that demonstrate technical feasibility.

INTRODUCTION

As the digital age brings shifts in customer preferences and emerging business models to urban mobility, these changes pose both challenges and opportunities for public transit operators and regulators. Today's transit riders, growingly more accustomed to the on-demand economy and customer-centric user experience, continue to hold transit services to ever higher standards. In attempts to remain sensitive to consumers' needs, service providers need to pay more attention to proactively engaging customers to gather feedback and build more personal relationships (1). These evolving new norms in the service sector come increasingly at odds with the traditional methods and metrics for monitoring and evaluating transit service quality. Customer satisfaction surveys, though common in the industry, are administered too infrequently and by-and-large ask subjects to provide only general, overall ratings. This abstractness 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.

Mobile technology and digital platforms bear the potential to make intelligence gathering more dynamic, effective, and engaging. The Internet of Things (IoT) brings the promise of a widespread network of uniquely identifiable sensing and computing devices of different kinds, able to communicate with each other. Such a vision brings opportunities and challenges for transport applications, particularly in the realm of behavior sensing. One such device is the iBeacon (2). In its simplest form, an iBeacon broadcasts continuously its unique ID to the surrounding area, allowing devices such as smartphones to detect their co-location. Thus, an iBeacon can be attached to a stop or a bus to serve as a "context identifier."

In this pilot study, we combine the iBeacon technology with the Future Mobility Sensing (FMS) (3, 4), a smartphone-based travel survey system developed by the Singapore-MIT Alliance for Research and Technology (SMART). The iBeacons provide FMS with the additional capability to detect users' arrival at and departure from bus stops, thus allowing us to conduct surveys on the transit experience as users are waiting, boarding, or completing their bus trips. For each trip, users receive a single survey randomly selected from the set of three stages, followed by a retrospective survey about the same trip at the end of day. These real-time and retrospective survey responses enable comparisons between real-time and recalled levels of satisfaction, potentially offering exploratory insights into the complexities of passenger satisfaction and decision towards using transit.

This pilot is a proof-of-concept, aiming to test the technical feasibility to crowdsource bus service monitoring and passenger satisfaction sensing. The pilot is conducted in collaboration with the Land Transport Authority of Singapore (LTA). The key capabilities of interest are 1) accurate and reliable detection of a passenger's stage in the trip that requires minimum user input from the user; and 2) gathering of meaningful, high-resolution feedback down to the level of a given trip by an individual passenger. This paper is organized as follows: we review related literature in the next section and then describe our technology. We subsequently introduce the survey design and present and discuss the pilot results.

LITERATURE REVIEW

In the public transit world, the concepts of service quality and customer satisfaction are closely connected. Service quality represents transit performance from passengers' perspectives (5, 6, 7), and should reflect quantitative as well as qualitative measures (8, 9). Inherently customer-centric, assessing quality of service thus entails gauging customer satisfaction, often through surveys (8).

The most common method for measuring service quality and customer satisfaction are intercept surveys on board and at stops/stations. Of the 27 U.S. transit agencies and Metropolitan Planning Organizations surveyed in 2006, three-quarters reported of using intercept surveys, with two-third of which also supplementing with telephone interviews (10). This is unsurprising, given the advantages of onboard and intercept surveys in gaining direct access to transit customers and obtaining relatively representative samples (11). Questionnaires commonly ask respondents to rate their satisfaction with overall transit service and its individual attributes: wait time, reliability comfort, etc. (12, 13, 14, 15, 16). Agencies may also solicit passengers' reported importance of various service attributes (17). Evaluations are collected periodically, and ratings are averaged to generate scores for historical comparisons.

The primary limitation of this *report card* approach is the incapability for ubiquitous, detailed assessment and feedback (18, 19). In Adler et al's (10) examination of survey practices, only 30% of the agencies conducted customer satisfaction assessments more than once a year; another 30% carried out such surveys less than once a year, and 20% had never done one. As a result of the low frequency, questionnaires typically ask respondents for general ratings on various aspects of the transit service, based on prior experiences (20). This puts transit regulators and system administrators at a disadvantage in understanding how performance varies by driver, route, and time of day, as well as in identifying the precise area for commendation and targeted improvement (19). Of course, in-person questionnaires, administered during or after the ride, could solicit trip-specific assessment, but their scopes would be quite limited. Transit experiences can well vary from trip to trip, but it would not be realistic to deploy survey teams to every bus, stop, and station every day.

The snapshot nature of traditional satisfaction surveys also calls into question the accuracy and meaningfulness of data. People's actual and recalled experiences often differ due to psychological heuristics (21). Pedersen et al (22) recorded 62 volunteers' predicted, experienced, and remembered satisfaction of transit trips for a month, revealing that their recalled satisfaction was significantly lower than experienced satisfaction. Abou-Zeid et al (23) observed a similar bias, noting that subjects report lower satisfaction with transit after experiencing a commute by automobile. This phenomenon implies two things. First, if we want to better capture riders' transit experience, surveying should be done as in real-time as possible, potentially even segmenting by the waiting, on board, and post-alighting portions of the trip. Second, we have been ill-equipped at understanding the relative importance of actual versus recalled customer experiences for the health of the transit system. Should transit operators and regulators care more about the joys and pains on a certain trip, or people's overall, longer-lasting quality assessment? How do they influence each other?

Given smartphone's ubiquity and apps' capacity to prompt data collection and transmit information in real-time, smartphone-based survey systems hold great promise. Smartphone-based surveys can be carried out without the need for mobilizing field surveyors, reducing the time and cost of survey administration. They can also be deployed over extended periods of time and space, enabling intra-day, inter-day and inter-seasonal assessments for numerous routes, stops, segments, etc. Carrel et al (20, 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 were then matched with the automatic vehicle location data to infer transit trips. Using a BlackBerry OS-based app (TOES), Dunlop et al (18) implemented a series of surveys to measure riders' emotional state before, during, and after each bus trip. Participants needed to manually signal the beginning of their trip and, subsequently, each survey in the sequence; otherwise, they would be prompted by the app to complete the next stage of the trip every six minutes. Undoubtedly, these two pioneering efforts generated invaluable multi-day, real-time data for understanding transit experiences from the customers' perspectives. Given the extensive manual input required, however, these apps are unsuited for a sustainable quality monitoring platform through rider feedback.

The existing literature on transit service quality and customer satisfaction reveals little consensus on what to measure and how to measure it. And on the differences in measuring perception and experience in-the-moment versus recalled. Studies have measured these concepts with as few as six attributes (6) and as many as 31 components (25). Most studies fall between 8-22 factors (5, 7, 8, 9, 26, 27, 28, 29, 30, 31, 32). Of course, this diversity comes partially from varying levels of specificity and ways to categorize the attributes; overlaps are common. Roughly, most of the above literature encompass the following indicators: system coverage/accessibility, wait time, travel time, transfers, onboard comfort and crowding, customer service, information availability, safety, and stop and station facilities.

The general concept of service quality offers a framework to help contextualize and define transit service metrics. The most widely applied approaches in market research are SERVQUAL (33) and SERVPERF (34). Though the two frameworks differ in the theory underlying customers' satisfaction, they both measure perceived quality of service with the same 22 specific indicators along five dimensions: Reliability, Assurance, Tangibility, Empathy, and Responsiveness (RATER). Many transit studies have been inspired by this five-dimension framework, adapting the indicators to better fit special characteristics of transit services and of the local rider communities (19, 35, 36). The "empathy" dimension is the trickiest, since the nature of mass transit services juxtaposes intimate, one-to-one relationship-based services. Some scholars hence propose excluding the "Empathy" category (37, 38) when assessing transit quality. We believe, however, that empathy can still manifest through personal relationship-building: understanding the customers, seeking feedback on their experiences and input for improvement.

FMS' real-time surveying capability presents an opportunity to test another innovation – measuring riders' subjective well-being and happiness from transit experiences. Transit services aim to improve people's quality of life by providing access to work, education, and recreation, as well as through improving the urban environment. Thus, surveys that effectively gauge riders' well-being could better reflect service quality than satisfaction reports. A number of studies have sought to measure happiness and well-being from travelling (39, 40, 41, 42). In particular, Ettema et al. (43) develop and test a measure consisting of a self-reported satisfaction with travel scale, an affective mood scale, and a satisfaction with day scale. In modelling commute satisfaction, Abou-Zeid (44) employs questionnaires that include questions on travelers' affect (such as com- mute enjoyment and stress), personality, and well-being. Ory and Mokhtarian (45) devise a set of questionnaires to predict people's liking for travel based on attitude, personality, and lifestyle preferences. These studies provide direct precedent to the design of mood, personality, and well-being-related questions in our pilot.

In summary, considerable room exists to innovate in measuring and monitoring transit service quality. With the aim of providing transit agencies and regulators with an innovative, yet practical tool, one which is more real-time, granular, and actionable, we adopt a mobile-phone based approach, using Future Mobility Sensing (FMS).

FUTURE MOBILITY SENSING - TRANSIT QUALITY (FMS-TQ)

With the vision of a practical tool in mind, we designed FMS-TQ to combine three objectives: (1) to gather operational intelligence in a more real-time and granular fashion; (2) to more accurately capture riders' transit experience by enriching self-reported satisfaction with emotions and happiness measures; and, (3) to pair real-time and retrospective surveys on any given trip. The latter aims to shed further light on the study of perception and experience, improving our understanding of the differences in measuring happiness as an in-the-moment versus a recalled experience.

FMS technology

The Future Mobility Sensing platform (FMS) (3, 4) was developed for high-resolution longitudinal travel surveys, primarily for activity-based modeling. Its original system architecture contains three components: smartphone app; backend; web interface.

The **smartphone app** (currently in iOS or Android) has the role of collecting location (GPS, WiFi, GSM), accelerometer data and some other information (e.g. battery level) on a continuous basis. The goal is to capture all trip data, be it through motorized (car, bus, train, taxi, motorbike) or non-motorized modes (walking, bicycle). The app is allowed to rest idle when the user is not moving. A state machine algorithm in the app leverages transition events data from the accelerometer and OS API events (e.g. Significant Location Change in iOS) to push notifications to users. Besides collecting location data, the app is also responsible for uploading data to the backend in batches, when the phone is charging. The **backend server** receives and stores the data and turns them, in real time, into trip information (stops, trips, modes) using machine learning algorithms (46). These data then supports a **web interface** that presents the user with a sketch of her day. A map, together with an editable timeline with sequences of stops and trips, are available for users to "validate" their data. For further details about FMS, please refer to earlier literature (3, 4).

For the FMS-TQ survey, we used only the smartphone and backend components of this technology. Figure 1 depicts FMS-TQ architecture.

iBeacon extension

The iBeacon (2) is a small device that leverages low power Bluetooth transmission to frequently broadcast its unique ID for a very long battery life. Current Android and iOS phones are able to recognize such signals and identify the unique ID. When attached to a bus stop, an iBeacon becomes a very simple yet high resolution method to enable smartphone apps to determine users' arrival to, waiting at and leaving from a bus stop. In our study, the weatherproof iBeacons were mounted in a strapped pouch near the top of the bus stops (Figure 2); this set-up is to maximize signal transmission and facilitate installation and removal without modifications to the stop infrastructure. The iBeacons cost approximately USD 25 each, and have a reported battery life of two years.

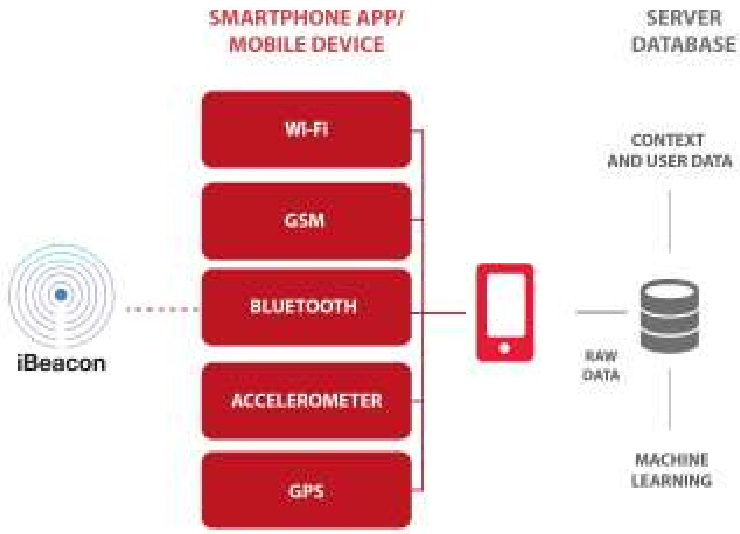

FIGURE 1 FMS-TQ architecture

FIGURE 2 Left: Estimote iBeacon used in the study; Middle and right: iBeacon installation at two bus stops.

We extended FMS' current state machine algorithm to consider two new events: iBeacon entry event (detecting a known iBeacon) and iBeacon exit event (no longer detecting a known iBeacon). These events are important for determining the present stage of a user's bus trip, and, correspondingly, the correct questionnaire to display in the app. We ran initial tests on the iBeacon technology to measure the proximity accuracy (distance from the beacon) and the received signal strength (rssi). Weaker signal strengths generally resulted in larger proximity accuracy values. We also found that the rssi and proximity accuracy were affected by the number of people in the surrounding environment. This relates to a well- known problem that human body interferes with high frequency signals. This shortcoming should not affect our study much, given that the FMS-TQ app only needs to detect the iBeacon as opposed to accurately determining its location. We suspect, however, that the adverse effect would be more pronounced in overcrowding conditions.

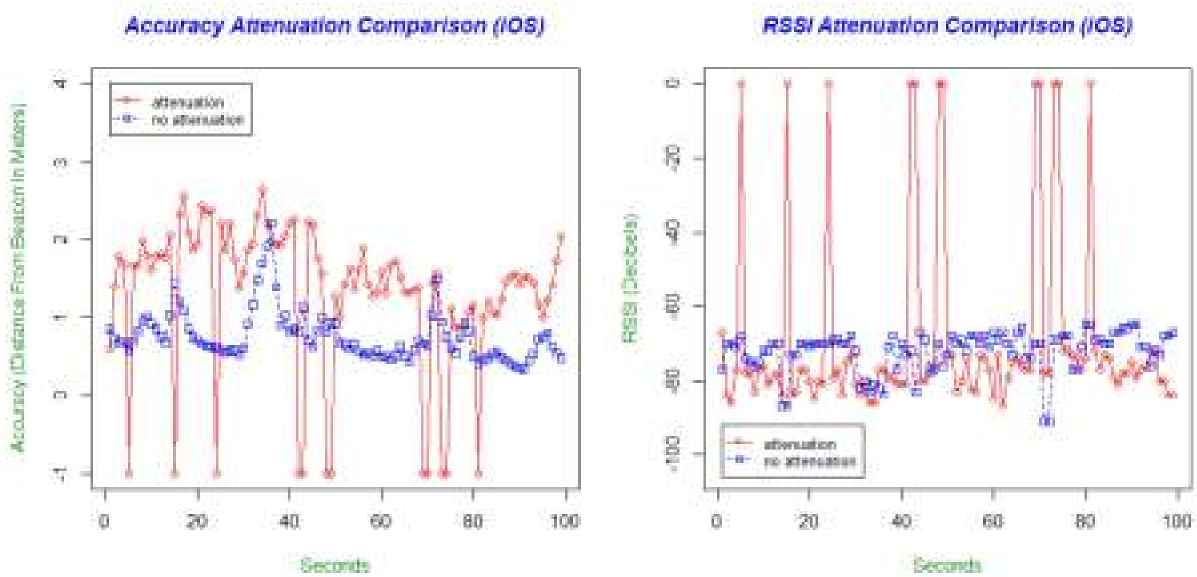

FIGURE 3 Effect of human body on accuracy (left) and on signal strength (right). "Attenuation" denotes having one person between the iBeacon and the smartphone; on the other hand, "no attenuation state" signify a clear line-of-sight.

From these tests, we concluded that line-of-sight transmission would maximize detection accuracy. Hence, we decided to place iBeacons at high positions at bus stops. For this pilot, the FMS-TQ uses iBeacons exclusively for real-time detection of bus stops, although the FMS backend also possess similar capabilities non-real-time using users' location data submitted by the app.

FMS-TQ pilot survey

Survey design

The service attributes measured by FMS-TQ aim to be comprehensive and relevant to riders' bus experiences at each of the trip stages. Thus, we modified the RATER framework to accommodate characteristics of bus services. We arrived at 17 factors, coalesced from the European (EN 13816) and American standards (TCRP Reports 100 and 47) for transit service quality, and targeted at the most relevant bus trip stage (Table 1). For example, passengers' satisfaction with wait time should be measured after riders board the bus, whereas judgment on directness of route/convenience can be salient during all three stages of a bus trip. Ultimately, the questionnaires encompass 14 factors. "Safety and security" is excluded due to Singapore's local characteristics; achieving "understanding of customer" and "personalized customer relationships" require much more effort and thus not yet ready to be measured.

TABLE 1 Adapted RATER Table

SERVQUAL/ SERVPERF Categories	Factors of bus service quality	Trip Stage of Measurement		
		At stop	On-board	After alighting
TANGIBLES	Accessibility to bus stop	X		
	Stop facilities	X		(X)
	Wait time		X	
	Travel speed/time		X	X
	Seating & personal space	X	X	
	On-board comfort		X	
	Directness of route/ need for transfers	X	X	X
RELIABILITY	Reliability of wait time		X	
	Reliability of being able to board bus		X	
	Reliability of travel time			X
ASSURANCE	Availability & accuracy of information	X		
	Bus driver's skills		X	X
	Attitude and quality of customer service		X	X
	Sense of safety & security	N/A	N/A	N/A
RESPONSIVENESS	Assistance to customers when needed		X	X
EMPATHY	Understanding of customer experiences	Not measured by trip.		
	Personalizing customer relationships			

FMS-TQ comprises a set of five questionnaires: an entrance survey, three event-based surveys for the three bus trip stages, and an end-of-day survey. To mitigate user burden, each questionnaire consists of only a few multiple choice questions, and each user receives only two questionnaires per day: one randomly-selected event-based questionnaire and the end of day questionnaire. Figure 4 depicts the method used to detect the various trip stages and their relationship with the questionnaire generation process. Notice that there are two general processes, each one runs only once per day: one in real-time, triggered by iBeacon events; the other runs in the end of the day, if at least one trip was previously detected for that day.

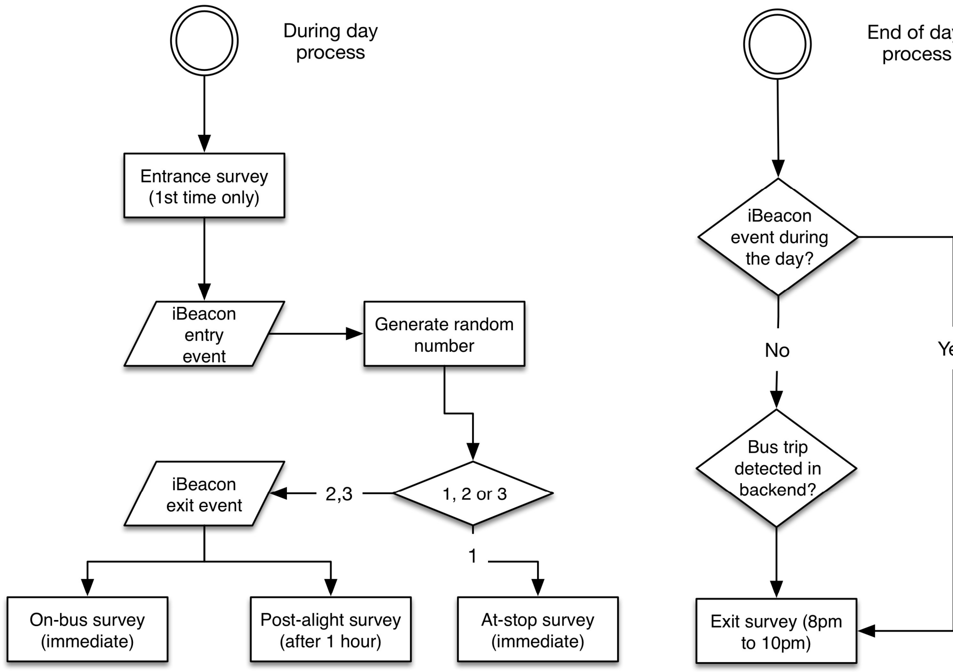

FIGURE 4 Trip stage detection and survey solicitation logic

While the "during the day" questionnaires are designed to be answered in real-time, FMS- TQ can accommodate late responses as well. In the case where the user fails to respond to any of the event-based questionnaires within the first hour, the question phrasing changes from the present to past tense. The system also monitors and analyzes people's response time and rates, which will lend insights to future refinement of the app and survey design. We now explain the content of each of the five surveys:

Entrance survey: After installing the app on her phone, the user receives an initial intake survey to collect baseline information on the user, including travel habits and attitudes, general satisfaction with transit, satisfaction with life, and demographic information.

At-stop survey: When a user accesses an iBeacon-equipped bus stop with FMS-TQ installed and running on her smartphone, the phone identifies the iBeacon transmission, which initiates the trip segment questionnaire generator, randomly assigning to the user the at-stop, on-board, or alighting questionnaire. The iBeacon trigger, combined with subsequent FMS-based sensing, also allows us to post-process estimate the user's wait time at the bus stop. If the app has randomly selected the at-stop questionnaire for the user, this is implemented 60 seconds after the iBeacon- based detection at the stop. The questionnaire begins by asking for confirmation that the user is indeed waiting at the specific bus stop. Upon confirmation, the user would be asked about purpose of the trip, whether any transfers are involved, satisfaction with the bus stop condition, and his/her feelings while waiting.

On-board survey: The FMS app detects the user boarding the bus from two signals: the phone leaving the iBeacon signal area, followed by acceleration at vehicular travel speed. Similar to the at-stop procedure, if the app has randomly selected to survey the user on-board, the ques- tionnaire will first verify that the user is indeed travelling on the bus. If so, then the app poses seven questions aimed at collecting three types of information: 1) subjective customer satisfaction of the wait and onboard experiences, 2) reasons for any dissatisfaction, and 3) observations on onboard crowding level. The latter two types of questions intend to provide feedback on service quality from a relatively objective perspective, covering bus stop condition, service information availability, crowding, comfort, and the driver's service.

Post-alight survey: If selected, the post-alighting questionnaire begins one hour after the user exits the iBeacon area (bus stop). The questionnaire targets the overall bus travel experience, gauging people's perception of their travel times, convenience, and overall satisfaction with the service. Since a rider's perceived bus experience can be much influenced by his/her activities during the trip, questions also focus on the user's on-board activities.

End-of-day At the end of each day, all users who took a bus trip that day are given an end-of-day questionnaire. We determine if a user has taken a bus trip in one of two ways: (1) she completed one of the bus trip stage questionnaires; or (2) we infer they made a bus trip based on FMS back-end analysis of their sensor data. The end of day questionnaire is then implemented in the 8pm to 10pm window. This questionnaire aims to collect respondents' reflections on their transit experience and their evaluation of transit's impacts on their lives and happiness that day. It also asks users the degree to which their bus experience has met their expectations. Comparing to the previously-mentioned questionnaires, which focus more on specific service attributes, this retrospective survey targets riders' broader well-being and travel choices. The pairing of responses to the trip-segment and end-of-day questionnaires intends to provide grounds for comparing real-time and retrospective passenger satisfaction.

Figure 5 presents a few snapshots of the several different surveys.

In addition to passengers' responses, the app also automatically collects the following trip-based data:

- Boarding stop, estimated arrival to stop, estimated arrival and departure time of trip;
- GPS location readings (time and place stamps) and speed estimates.

Based on these readings, we are able to estimate, on the FMS back-end: waiting time, in-vehicle travel time and average speed, alighting stop, and characteristics of bus stop access and egress (likely mode, total distance and time).

Survey administration

To pilot the approach, we partnered with Singapore's Land Transport Authority (LTA). The LTA disseminated an email invitation to employees at one of their offices, encouraging them to "help innovate Singapore's bus service" by "transforming the way we sense transit service quality and customer satisfaction" and offering a prize draw of SGD 100 gift card to be awarded via lottery (with chances of winning directly related to the number of questionnaires completed during the course of the pilot). Android and iPhone users were invited to download the FMS-TQ app from Google Play and the iOS App Store, respectively, and then prompted to register for an account. We installed iBeacons at four bus stops near the LTA office, identifying with LTA administratively feasible stops (as some bus stops are managed via private concessionaires) most likely to be used by employees (see Figure 2. The pilot began on June 18, 2015 and ended on July 4, 2015.

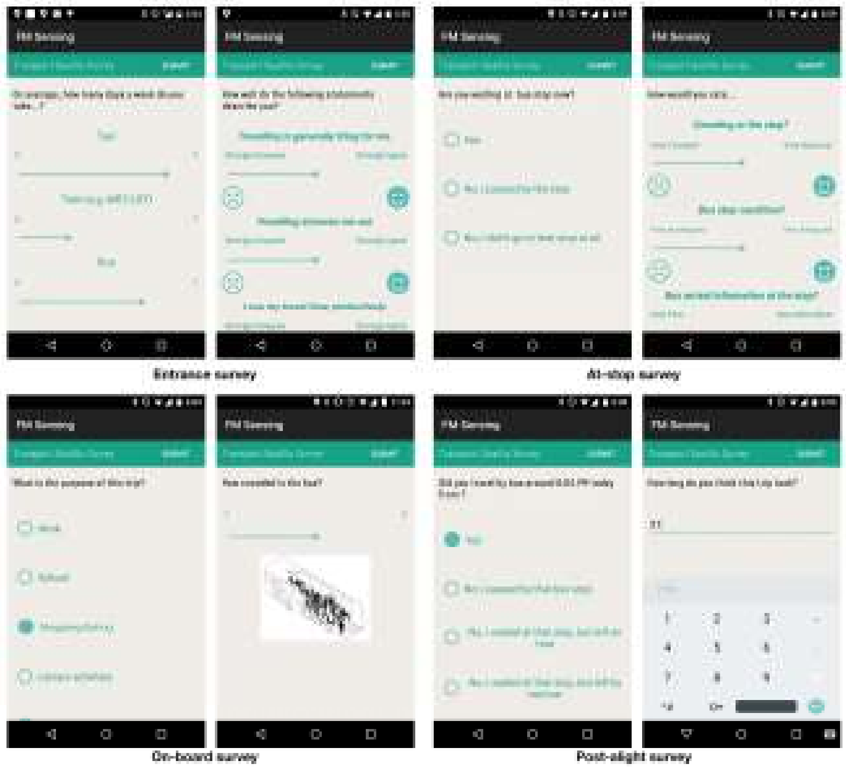

FIGURE 5 Illustrative snapshots of the different surveys.

Results and discussion

In total, 32 users initially registered, 24 completed the entrance questionnaire. Unsurprisingly, the group is dominated by bus riders - 15 respondents report taking the bus every day, and five report bus usage six days a week. The most common reason for choosing bus over other alternatives is convenience. Despite their employment affiliation with the LTA, their responses to the entrance survey reveal no glaring biases in opinion towards transit. Less than half of the surveyed agree with the statement “I like using public transport” In fact, they may be more critical of the local transit services than the general population, with only 46% of respondents report being “satisfied” with their bus experience in Singapore - compared to 88% of those surveyed in the 2014 LTA Public Transport Customer Satisfaction Survey (17). While the majority think that travel is tiring and offer little benefits besides reaching the destination, their opinions on stressfulness, environmental consciousness and ability to use time productively during travel, are roughly evenly distributed across the spectrum.

TABLE 2 Pilot FMS-TQ Respondents and Responses by Questionnaire

Questionnaire	Number responses (false positives)	Number respondents
Entrance	25	25
At-Stop	22 (8)	10
On-board	23 (9)	9
Post-alight	8 (5)	6
End-of-day	51	14

For each real-time survey, we ask the user if s/he had indeed been at the bus stop. In this way, we can understand the accuracy provided by the iBeacons. From a total of 52 iBeacon triggered surveys, 21 were considered false positives. A detailed analysis revealed that the iBeacon range may be set too high (can reach up to 70 meters) and the threshold of 60 seconds continuously receiving iBeacon signal to avoid false positives (e.g. walking near the bus stop) was too low. We also compared the quality of stop detection from backend intelligence (post-processing) with what is provided with the iBeacons. Notice that the end-of-day survey would also take into account backend analysis. In total, there were 14 cases where iBeacon could not detect a backend detected stop. It is likely (but not verifiable in current data) that these were backend false positives or that iBeacon signal was attenuated on crowded situations. There were 26 cases of stops detected through iBeacons but not caught by the backend. This may happen if FMS location data is noisy, absent, or the wait for the bus was too short to be considered a stop. Interestingly, whenever the two technologies agreed, the detection time difference was negligible.

Over the course of the pilot, a total of 129 questionnaires were completed, in addition to the 25 entrance responses: 22 at-stop, 23 on-board, 8 post-alighting, and 51 end-of-day (Table 2). The low number of post-alighting responses was due to a bug in the software discovered after the pilot had begun (a roughly equal share of post-alighting, at-stop, and on-board questionnaires should have been implemented). Furthermore, only after the pilot began did we notice that we needed to remind Apple users to keep Bluetooth on (in Android, we could set this programmatically). At the end of the pilot, only 16 users responded to a questionnaire other than the entrance questionnaire.Over the course of the pilot, more respondents (14) completed the end-of-day questionnaire at least once than any of the trip-stage-based questionnaires (Table 2).

In order to enable real-time surveying, the system must be able to detect trips as they happen. Results show that FMS-TQ can detect trip stages with essentially no user input, though its accuracy needs further improvement. On 22 instances, respondents reported of not having waited for or boarded a bus at the stop detected by the app (false positives). Questionnaire timestamps and pilot exit survey responses suggest that, for many of these cases, the app may have misinterpreted people alighting at an iBeacon-equipped stop as waiting or boarding at that stop. While we had set a minimum threshold of 1 minute (of continuously receiving the iBeacon signal) to prevent this, this points to the need of adjusting this setting. The imperfectness of iBeacon-based trip detection also affirms, for now, the importance of placing a trip-verifying question in the beginning of surveys that employ automatic trip detection.

TABLE 3 At-Stop Survey Response Summary

User ID	Date & time	Trip purpose	Transfers on this trip?	Passenger's affective state		Bus stop crowding*	Rating on bus stop condition **	Rating on bus arrival information at the stop ***	Rating on accessibility to the bus stop ****
				Primary descriptor	Second descriptor				
442	6/25/2015 18:23	Shopping / Eating	no	Anxious		5	4	3	3
453	6/19/2015 8:53	Work	no	Neutral		3	3	3	3
453	6/23/2015 19:07	Home	yes	Neutral		3	3	2	4
453	6/25/2015 8:55	Work	no	Neutral		3	3	3	3
453	7/1/2015 9:12	Work	no	Neutral		3	3	3	3
456	6/22/2015 7:47	Work	yes	Neutral		3	3	3	3
456	6/29/2015 7:53	Work	yes	Neutral		3	3	3	3
456	7/8/2015 7:55	Work	yes	Neutral		3	3	3	3
459	7/3/2015 18:21	Shopping / Eating	No	Very calm	Very content	2	2	2	4
465	6/25/2015 8:10	Work	no	Neutral		4	4	4	4
465	6/26/2015 8:10	work	no	Neutral		4	3	4	4
466	6/23/2015 12:50	Shopping /Eating	no	Neutral	Tired	4	3	3	4
470	6/29/2015 18:09	Going home	yes	Very tired	Very hurried	2	3	4	4
470	7/1/2015 18:20	Going home	yes	Very anxious	Very bored	4	4	1	1

*Rate on scale of 1-5: 1 = very crowded, 5 = very spacious **Rate on scale of 1-5: 1 = very inadequate, 5 = very adequate

***Rate on scale of 1-5: 1 = very poor, 5 = very informative ****Rate on scale of 1-5: 1 = very inconvenient, 5 = very convenient

Despite the pilot's limitations, the responses demonstrate FMS-TQ's capability to gather information on transit service quality that may better reflect customers' actual experience. For illustration, user #465, for example, reports two very different bus experiences on the same commute on two consecutive days (Table 4). On June 22, the user reports satisfactory wait and onboard experiences, while on the next day, however, s/he waits longer for the bus. According to the entrance survey, this user usually checks real-time bus arrival information before the trip, which suggests that the bus might have encountered an abnormal delay. The user's responses also reveal on- board crowding and a somewhat rough ride. The differentiating details between these two bus trips would have almost certainly been lost in more traditional customer satisfaction surveys.

For transit operators and regulators, the most useful benefit of this smartphone-based system may be trip-specific feedback. The FMS-TQ system records the phone's geographical coordinates when surveys are solicited and returned, which, when combined with the automatically- collected GPS traces, enable us to infer the bus route taken. For example, at 18:14 on June 22, User #465 receives an onboard survey at (1.3104978, 103.8477377). This matches the in-vehicle trip detected by FMS between 18:12 (1.310086, 103.848) and 18:22 (1.32541, 103.8419). We can thus infer that the user was traveling on one of the five bus routes serving that corridor at that time of day (lines 56, 57, 166, 851 and 980). If given access to automatic vehicle location (AVL) data, we could also know the exact bus trip by matching the user's coordinates in FMS with AVL data. A simpler technological approach, at the cost of additional user input, would be to include a question about the bus number in the onboard and alighting surveys, or even to install iBeacons in front and back of every bus.

Generating trip-specific feedback will likely become a crucial capability for transit industry in the coming years. Compared to the rapidly-growing on-demand mobility services (e.g. Uber, Lyft), transit agencies have fallen far behind in soliciting individualized feedback for quality control and cultivating relationships with riders. Companies such as Uber and Lyft have built meticulous quality control on the mandatory mutual ratings between drivers and passengers at the end of every trip (48). As a result, not only do they have detailed service quality data for analysis and monitoring, these new mobility service providers have also become much more self-aware and self-regulating of their quality standards.

Lastly, with both real-time and retrospective customer opinions, we can begin to explore how in-the-moment experiences may differ from recalled memory. 21 bus trips have both real-time and end-of-day survey responses. While this is not a sufficient pool to statistically detect patterns, incidents of discrepancies are visible to the naked eye. For instance, User #458 expresses dissatisfaction with the wait (rated 3/5) and comfort (rated 2/5) while onboard the bus, citing "wait time too long" and "too many people on the bus" as rea- sons. However, s/he seems rather happy with the trip experience (rated 4/5) when asked again later that day. Such discrepancies should not be surprising - the unpleasantness of being on a crowded bus after a long wait would reasonably be most intense during the trip. Of course, this does not imply superiority of one type of measures over the other. The real-time indicators are helpful at reflecting riders' actual experience, which could aid efforts to mitigate peak-period frustration for customers; on the other hand, recalled reports of satisfaction may be more salient to users' longer-term decision-making, including future mode and route choice.

TABLE 4 Onboard Survey Response Summary

User ID	Date & time	Wait satisfaction*	Why less than satisfied with wait? (if wait rating < 4)	Bus crowdedness**	Comfort satisfaction*	Why less than satisfied with comfort? (if comfort rating < 4)	Driver service satisfaction*
442	6/24/2015 12:42	4	N/A	2	3	Ride is not smooth	3
448	6/26/2015 8:26	3	Other	3	4	N/A	4
453	6/22/2015 9:00	3	Wait time too long	2	3		"Don't know"
453	7/2/2015 9:00	3	Rather not say	4	3	Rather not say	"Don't' know"
453	7/8/2015 8:50	3	Rather not say	4	3	Cannot get a seat on the bus	"Don't know"
456	6/24/2015 7:37	3		2	3	Temperature is too hot/cold	3
456	6/30/2015 17:02	3	Bus too crowded to board	2	3	Too many people on the bus	3
456	7/2/2015 7:52	3		2	4		3
456	7/7/2015 7:54		4	2	4		3
458	6/19/2015 17:42	5	N/A	5	2	Too many people on the bus	5
458	6/24/2015 12:32	3	Wait time too long	4	2	Too many people on the bus	4
465	6/22/2015 18:14	4	N/A	3	4	N/A	"Don't know"
465	6/23/2015 18:15	3	Wait time too long	4	2	Ride is not smooth; Too many people on the bus	"Don't know"
473	6/26/2015 7:48	4	N/A	3	3	Cannot get a seat on the bus	4

* Rate on scale of 1-5: 1 = very dissatisfied, 5 = very satisfied
** Bus crowdedness: user shown a pictorial scale of 5 levels of crowding, with 1 = very empty, 5 = very crowded.

While the results from the pilot have been encouraging, the FMS-TQ platform brings its challenges. For example, iPhone users would need to manually enable Bluetooth on their phones.Another limitation is the dependency on having a wide set of iBeacon-equipped stops. If the ultimate vision is to develop a platform for a city's entire bus network, it may be impractical to install and maintain iBeacons at every bus stop.

Going forward, the approach can be enhanced through improved location and travel mode identification. We also need to learn more about user burden in using the app, perhaps though more agile selective surveying based on pattern detection. For instance, a commuter's morning trip to work tends to be consistent spatially and temporally across workdays, so his experience may be similar across multiple trips. We can thus push the survey to him once every 5 or 10 trips.

Lastly, we likely need a more effective incentives program to sustain participation, such as offering reward points or fare discounts. Our pilot did not focus on incentive design for public participation nor possible adjustments in transit regulatory frameworks; these elements will be examined in future work which aims to develop a mature, effective, and sustainable crowdsourcing system that helps agencies better monitor, improve, and govern transit services.

Despite the tasks ahead, FMS-TQ holds good potential to improve transit service monitoring and regulation. The advances in smartphones and IoT technologies represent a great opportunity (and a necessity) for transit operators, regulators, planners, and citizens alike to leverage new technologies for generating more meaningful information.

CONCLUSION

This research aspires to bring multiple innovations to the field of transit service management and regulation. First, the pilot demonstrates FMS' capability to detect various stages of a bus trip that enables surveying a passenger during travel with an unprecedented high resolution. Information collected by FMS-TQ can be linked to specific bus trips, signifying a step towards allowing transit operators and regulators to identify precise areas for improvement or commendation. This granular, real-time, and customer-centric information collection can help agencies, operators and others to become more in-tune with what passengers actually experience, representing an important step towards building more personal relationships with customers.

ACKNOLWEDGEMENTS

This research was supported by the Singapore MIT Alliance for Research and Technology's FM IRG research program. We thank the Land Transport Authority of Singapore for their support of the pilot. We would also like to thank Paula Garbagnoli, Inês Dias, Ajinkya Ghorpade, Jorge Santos, William Ko, Bruno Santos and Rui Baltazar, who have contributed to making the FMS system a reality.

REFERENCES

[1] Prahalad, C. K. and V. Ramaswamy, Co-creation experiences: The next practice in value creation. *Journal of interactive marketing*, Vol. 18, No. 3, 2004, pp. 5–14.

[2] Aislelabs, *The Hitchhikers Guide to iBeacon Hardware: A Comprehensive Report by Aisle-labs (2015)*. http://www.aislelabs.com/reports/beacon-guide/, 2015.

[3] Cottrill, C., F. Pereira, F. Zhao, I. Dias, H. Lim, M. Ben-Akiva, and P. Zegras, Future mobility survey: Experience in developing a smartphone-based travel survey in singapore. *Transportation Research Record: Journal of the Transportation Research Board*, , No. 2354, 2013, pp. 59–67.

[4] Zhao, F., F. C. Pereira, R. Ball, Y. Kim, Y. Han, C. Zegras, and M. Ben-Akiva, Ex- ploratory Analysis of a Smartphone-Based Travel Survey in Singapore. *Transportation Re- search Record: Journal of the Transportation Research Board*, , No. to appear, 2015.

[5] Hensher, D. A., 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*, Vol. 37, No. 6, 2003, pp. 499–517.

[6] Kittelson & Associates, Transit Cooperative Research Program T. C. R. Program, and Transit

Development Corporation, eds. *Transit capacity and quality of service manual*, Vol. 100. Transportation Research Board, 2003.

[7] Tyrinopoulos, Y. and C. Antoniou, Public transit user satisfaction: Variability and policy implications. *Transport Policy*, Vol. 15, No. 4, 2008, pp. 260–272.

[8] Eboli, L. and G. Mazzulla. How to capture the passenge's point of view on a transit service through rating and choice options. *Transportation reviews, Vol. 30, NO.4, 2010, pp. 435-450*

[9] Eboli, L. and G. Mazzulla, A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. *Transport Policy*, Vol. 18, No. 1, 2011, pp. 172–181.

[10] Spitz, G., Niles F., and Adler, T. Web-Based Survey Techniques. A Synthesis of Transit Practice. *TCRP Synthesis*, 2007.

[11] Schaller, B., *On-board and intercept transit survey techniques*. 63, Transportation Research Board, 2005.

[12] Dunlop, I. N., J. M. Casello, and S. T. Doherty, Tracking Transit Rider Experience: Using Smartphones to Measure Comfort and Well-Being Throughout the Trip. In *Transportation Research Board. Meeting (93rd: 2014 : Washington, D.C.).*, 2012.

[13] Elmore-Yalch, R., *A Handbook: integrating market research into transit management*, Vol. 37. Transportation Research Board, 1998.

[14] Carrel, A., R. Sengupta, and J. L. Walker, The San Francisco Travel Quality Study: Tracking Trials and Tribulations of a Transit Taker. *Transportation*, 2015.

[15] Fredrickson, B. L. and D. Kahneman, Duration neglect in retrospective evaluations of affective episodes. *Journal of personality and social psychology*, Vol. 65, No. 1, 1993, p. 45.

[16] 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*, Vol. 18, No. 5, 2011, pp. 471–475.

[17] 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. *Transport Policy*, Vol. 19, No. 1, 2012, pp. 93–104.

[18] Habib, K. M. N., L. Kattan, M. Islam, et al., Model of personal attitudes towards transit service quality. *Journal of Advanced Transportation*, Vol. 45, No. 4, 2011, pp. 271–285.

[19] de Oña, J., R. de Oña, L. Eboli, and G. Mazzulla, Perceived service quality in bus transit service: a structural equation approach. *Transport Policy*, Vol. 29, 2013, pp. 219–226.

[20] Eboli, L. and G. Mazzulla, Service quality attributes affecting customer satisfaction for bus transit. *Journal of Public Transportation*, Vol. 10, No. 3, 2007, pp. 35–46.

[21] European Committee for Standardization. *EN 13816. Transportation - Logistics and Ser- vices - Public Passenger Transport - Service Quality Definition, Targeting and Measurement*. EN 13816:2002 E, 2002.

[22] Friman, M. and M. Fellesson, Service supply and customer satisfaction in public transportation: the quality paradox. *Journal of Public transportation*, Vol. 12, No. 4, 2009, p. 4.

[23] Parasuraman, A., V. A. Zeithaml, and L. L. Berry, A conceptual model of service quality and its implications for future research. *the Journal of Marketing*, 1985, pp. 41–50.

[24] Cronin Jr, J. J. and S. A. Taylor, Measuring service quality: a reexamination and extension. *The journal of marketing*, 1992, pp. 55–68.

[25] Cavana, R. Y., L. M. Corbett, and Y. Lo, Developing zones of tolerance for managing passenger rail service quality. *International Journal of Quality & Reliability Management*, Vol. 24, No. 1, 2007, pp. 7–31.

[26] Tripathi, V., A. Kumar, and G. Nema, Measuring Service Quality in India Public Transport System: A Comprehensive reviewâA˘ Ű. *Asia-Pacific Marketing Review*, Vol. 1, No. 1, 2012, pp. 113–119.

[27] Barabino, B., E. Deiana, and P. Tilocca, Measuring service quality in urban bus transport: a modified SERVQUAL approach. *International Journal of Quality and Service Sciences*, Vol. 4, No. 3, 2012, pp. 238–252.

[28] Too, L. and G. Earl, Public transport service quality and sustainable development: a community stakeholder perspective. *Sustainable development*, Vol. 18, No. 1, 2010, pp. 51–61.

[29] Bergstad, C. J., A. Gamble, T. Gärling, O. Hagman, M. Polk, D. Ettema, M. Friman, and L. E. Olsson, Subjective well-being related to satisfaction with daily travel. *Transportation*, Vol. 38, No. 1, 2011, pp. 1–15.

[30] Duarte, A., C. Garcia, G. Giannarakis, S. Limão, A. Polydoropoulou, and N. Litinas, New approaches in transportation planning: happiness and transport economics. *NETNOMICS: Economic Research and Electronic Networking*, Vol. 11, No. 1, 2010, pp. 5–32.

[31] Olsson, L. E., T. Gärling, S. Fujii, D. Ettema, H. Lekedal, and M. Friman, Relationship Between Satisfaction with Daily Travel and Subjective Well-Being in Three Urban Areas in Sweden: Description of Survey Questionnaire, Sample Characteristics and Preliminary Results, 2011.

[32] Olsson, L. E., T. Gärling, D. Ettema, M. Friman, and S. Fujii, Happiness and satisfaction with work commute. *Social indicators research*, Vol. 111, No. 1, 2013, pp. 255–263.

[33] Ettema, D., T. Gärling, L. Eriksson, M. Friman, L. E. Olsson, and S. Fujii, Satisfaction with travel and subjective well-being: Development and test of a measurement tool. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 14, No. 3, 2011, pp. 167–175.

[34] Abou-Zeid, M., *Measuring and modeling activity and travel well-being*. Ph.D. thesis, Massachusetts Institute of Technology, 2009.

[35] Ory, D. T. and P. L. Mokhtarian, When is getting there half the fun? Modeling the liking for travel. *Transportation Research Part A: Policy and Practice*, Vol. 39, No. 2, 2005, pp. 97–123.

[36] Ghorpade, A., F. C. Pereira, F. Zhao, C. Zegras, and M. Ben-Akiva, An Integrated Stop-Mode Detection Algorithm for Real World Smartphone-Based Travel Survey. *Transportation Research Board 94th Annual Meeting*, 2015.

[37] Cook, J., Uber's internal charts show how its driver-rating system actually works. *Business Insider*, 2015.