AN AUTOMATED, DATA-DRIVEN PERFORMANCE REGIME FOR OPERATIONS MANAGEMENT, PLANNING, AND CONTROL

Dominick Tribone (Corresponding Author)
At time of research: Department of Civil and Environmental Engineering, Massachusetts Institute of Technology. 77 Massachusetts Avenue, Room 1-235, Cambridge MA 02139.
Present affiliation: Massachusetts Bay Transportation Authority
10 Park Plaza Room 8312, Boston, MA 02116
781-234-5568 | dtribone@mbta.com

David Block-Schachter, Ph.D.
Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.
77 Massachusetts Avenue, Room 1-235, Cambridge MA 02139
dcbs@mit.edu

Frederick P. Salvucci
Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.
77 Massachusetts Avenue, Room 1-270, Cambridge MA 02139
617-253-5378 | salvucci@mit.edu

John Attanucci
Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.
77 Massachusetts Avenue, Room 1-274, Cambridge MA 02139
617-253-7022 | jattan@mit.edu

Nigel H.M. Wilson, Ph.D.
Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.
77 Massachusetts Avenue, Room 1-238, Cambridge MA 02139
617-253-5046 | nhmw@mit.edu

Words: 6,539
Tables & Figures: 1,250 (5 x 250)
Total word count: 7,789

ABSTRACT

As public transit agencies install new technology systems they are gaining increasing amounts of data. This data has the potential to change how they operate by generating better information for decision-making. Deriving value from this data and applying it to improve service requires changing the institutional processes that developed when agencies had little reliable information about their systems and customers. This research uses the Massachusetts Bay Transportation Authority (MBTA) as a case study. It first assesses how the MBTA currently measures performance. It then redesigns and advances the agency's daily performance reports for rapid transit through a collaborative and iterative process with the Operations Control Center staff. These reports are used to identify poor performance, implement pilot projects to address its causes, and evaluate the effects of these pilots. Through this case study, this research finds that service controllers' trust and interpretation of performance information determines its impact on operations. It concludes that new data will be most effective in producing service improvements if measurements accurately reflect human experience and are developed in conjunction with their intended users. It also finds that developing small pilot projects during this collaborative process enables new performance information to result in sustainable service improvements.

1 INTRODUCTION

While public transit agencies are becoming data-rich as they upgrade their technology, many of their institutional processes and behaviors developed when they had little reliable information about their customers or their operating performance. Until recently transit agencies relied on surveys and sampling to determine how many passengers they served, where these people were going, how long vehicles took to run routes, or how often service was on time (*1*). New automated data sources have the potential to change the way agencies operate by providing them with more and better information on which to base decisions. The presence of good information, however, is a necessary but not sufficient condition for improvements in service. Improving an agency's operations also requires understanding how to make this information meaningful to those responsible for service and how to make old institutional processes responsive to new information.

This research uses the Massachusetts Bay Transportation Authority (MBTA) in Boston, Massachusetts as a case study. It assesses the MBTA's current use of both real-time and historical data. Based on this assessment, it redesigns the agency's daily performance reports for rapid transit. By collaborating with MBTA personnel, it examines how MBTA employees interpret information and what they need to impact decisions about service. These reports are used to identify poor performance and develop pilot projects to address its causes. Because both the performance reports and pilot projects are developed within the institutional constraints of service management, these projects have been successfully implemented.

The MBTA provides a case where "big data" is available, but has not yet been harnessed to feed back into service provision. The agency could use its data to better understand trends, learn from them, and make improvements. While the data has been analyzed in the past, the fact that few changes have resulted from such analyses suggests that the problem is not purely analytical. The agency is a bureaucratic organization that relies on human action, human perception, and existing institutional processes, all of which constrain the use of such data. It thus provides an opportunity to explore, develop and test strategies to overcome the other factors influencing the ability of new data to impact public transit operations.

2 LITERATURE

The introduction of Automated Data Collection Systems (ADCS) in the transit industry has been accompanied by a wealth of research on how the data can be used to gain insight into service delivery. Many quantitative methods have been developed to accomplish this. Because people take information as an input into their actions, the effect of better information depends on their interpretation of its value and meaning.

2.1 Transit Performance Measurement Using Automated Data

In providing guidance on developing performance management plans for transit agencies, Transit Cooperative Research Program (TCRP) Report 88 (*2*) provides a comprehensive list of performance measurements for public transit systems and how to calculate them. While these could be calculated with manual data, automated data allows system performance to be measured in much finer detail and at much lower marginal cost (*1*). TCRP Report 88 includes hundreds of possible performance metrics.

Focusing in on implemented passenger-oriented metrics: New York City Transit calculates a "wait assessment" metric that measures service regularity. It is defined as the number of headways that are less than 125% of the scheduled headway (*3*). The London Underground uses travel time and its variability to judge service quality. Its Journey Time Metric (JTM) calculates customers' time between entering and leaving the system (since they must validate on both entry and exit). To capture variability, the JTM is compared to a scheduled value for that trip, based on scheduled headways and running times for the trains plus assumed access, egress, and interchange time (*4*).

2.2 Effecting Change in a Public Sector Bureaucracy

Making use of new automated data in transit agencies has both an analytical dimension and a managerial dimension. While the analytics are important, this research focuses on innovating within the bureaucracy and how automated data can be applied within the existing bureaucratic management structure. Many of the characteristics of innovation may conflict with the highly structured and methodical nature of bureaucracy. Behn argues that this creates inherent dilemmas for those attempting to make changes in government agencies. Innovation is not routine (*5*). Innovation is often experimental. It may result in failure as often as success. In many cases it involves changing procedure. This may disrupt the mechanized bureaucratic process.

The high degree of scrutiny placed on public agencies makes managers risk-averse. They are inclined to prioritize avoiding incidents over trying new things to optimize performance. Altshuler and Zegans note that much innovation originates from the lower ranks of an organization that are closer to service provision. This may conflict with the hierarchical, top-down nature of bureaucratic organizations (*6*). Altshuler and Zegans outline several broad strategies that they have found in successful cases of public sector innovation:

1. Proceeding incrementally;
2. Alleviating problems widely-recognized as urgent and explaining how the innovation addresses the problem;
3. Being close to clients and relying on them to convey positive messages to political authorities that support the innovation;
4. Casting a wide net in search of support and aligning existing institutional resources with the work;
5. Building and sustaining a coalition that supports the innovation and has the power to authorize and implement it;
6. Being open to feedback, which allows continuous learning and adaptation;
7. Being tenacious, dedicated, and optimistic in order to overcome major setbacks (*6*).

These observations suggest that new information is more likely to lead to innovation if it makes a clear case for change and addresses existing problems. This provides an argument for producing information in close collaboration with its intended end users in order to gain a better understanding of what information would help improve current practice. Because individuals in government agencies can be protective of their domains, working closely with them may help produce a sense of ownership and break down territorial barriers to innovation and embracing new information.

2.3 Role of This Research

The existing literature on translating data into information has successfully developed methods of applying ADCS data to measure transit service. It is not a lack of good information that is restricting the application of knowledge to improve transit services. While previous work has created new knowledge from ADCS-based information, there has been little research as to how information is used within an organization, what effect it has, and what influences its effectiveness.

This work seeks to begin filling in the gap in literature between how to measure service and how to make changes in a public organization. To this end, it focuses on how the measurements chosen, the design of the reports, and the process of creating them influence the impact information has on service delivery.

3 BACKGROUND: THE MBTA'S CURRENT PERFORMANCE STANDARDS AND REPORTS

The MBTA's regular use of historical data has been limited to On-Time Performance (OTP) reports: a single percentage for each route every day. "On-time" is based on the MBTA's service standards. These standards are developed and revised through a public process that takes customer input into account (*7*).

FIGURE 1 Red Line Diagram.

The rail OTP report summarizes OTP by period and direction (the Red Line has two branches, as shown in FIGURE 1, for a total of four directions). OTP for rail is judged solely on the headway departing the terminal. A 93% OTP for the Red Line means that 93% of trips left the terminal within 1.5 times their scheduled headway (the allowable deviation). This report only measures service on the two branches individually. There is no measure of combined service on the trunk portion, though 67% of all passenger trips are strictly within the trunk.[1] This means that the scheduled headway that is the basis of being on-time is the headway between two trains of the same branch. Branch headways are 9 minutes in the peaks, so trains are on-time if they leave within 13.5 minutes of the prior trip on that branch. Two northbound trains that reach the merge point at JFK/UMass 10 minutes apart can be on-time, even though this separation is more than double the expected joint headway of 4.5 minutes. Short headways (bunches) are also reported as on-time.

Moreover, the existing report is the first of 13 pages that present OTP for each trip run that day. The rail report contains only highly aggregate and highly disaggregate information. This limits the ability to understand trends and patterns. Moreover, there is no context for the OTP numbers other than time and direction, which inhibits the viewer from understanding potential problem areas. Furthermore, the laxness of the standard means that all three MBTA rail lines are usually above 90% on-time, even when service is delayed or disrupted.

The limitations of the current performance information restrict the amount of knowledge that can be generated. Internally, the amount of work required to relate different pieces of information and generate useful knowledge is time-prohibitive. The existing performance reports do not provide enough detail to show the impacts of dispatchers' reactions to real-time information. Headway adjustments mid-route to avoid bunching are not reflected in the current OTP numbers. Not measuring a joint headway (time between vehicles regardless of route) means that actions to even out service between routes do not factor into OTP.

The MBTA's reporting system could be improved by modifying the historical performance information to eliminate some of the barriers to its use. Namely (1) changing the way service is measured to reflect how customers experience service; (2) eliminating the need to search for detailed information; (3) showing and relating multiple dimensions of service.

Having assessed limitations in the MBTA's current use of its historical data, both for internal and external audiences, this research attempts to address the issues that limit its usefulness, particularly for operations personnel. In doing so, it rethinks both the metrics themselves, their presentation, and the process used to create them.

4 METHODS: DEVELOPING NEW PERFORMANCE METRICS AND REPORTS

The work was originally intended to provide more frequent and detailed information to customers about service quality that complements the performance "snapshots" produced by the MBTA's recently introduced -countdown signs. It started with reconceptualizing metrics, and engaged the MBTA's Operations Control Center (OCC) staff early in the process. The rationale behind this was that if a quantitative assessment of their work is to be made public, service controllers should first be given input

[1] Based on the Origin-Destination calculations from ongoing MIT research

into the measurements. Moreover, they should be given the chance to see and address issues that become evident with new measurement techniques. The initial discussions with the OCC staff revealed that they were also interested in revised performance metrics, which shifted the focus of this research to creating performance reports that contribute to dispatchers' knowledge. This process has been critical to the project's acceptance by the OCC and its ability to propose and implement service changes (described later). It included visits to the OCC to meet with dispatchers and managers and observe their work. Though the intention of releasing information publicly has been retained, this objective was not achieved in the research to date.

4.1 Designing New Performance Reports

Based on the shortcomings listed in the previous section about the existing metrics, this research identified multiple objectives for revised performance reports. These include being (1) reflective of the customer experience, capturing the operating characteristics of transit service that are salient to riders such as speed, frequency, and reliability; (2) sensitive to variations in service that passengers are likely to perceive, such as long headways or dropped trips; (3) limited to one page (either physical or virtual) so that information is less likely to be overlooked or ignored; (4) easily understood by operations control staff, managers, other MBTA personnel, and passengers alike; (5) detailed enough for operations staff to identify problems underlying poor performance and take corrective action; and (6) based on existing automatically collected data so that calculation can be automated and done in real-time or for the previous day.

This work hypothesizes that these qualities enable performance information to impact service. The reactions and feedback from OCC managers and staff were the primary means of determining how well objectives four and five were being met. Their feedback provides important lessons on how one of the reports' audiences understands them. This feedback was gathered informally in several meetings with the OCC director and Chief Operating Officer (COO) and in spending hours sitting with dispatchers during pilot projects.

This is not the first time passenger-weighted performance metrics have been used in transit, as noted in the literature review. Rather than re-invent the wheel, this research took effective performance measurement practices from elsewhere and focused on how to maximize its impact on service quality. This was accomplished both through focusing on the design of the performance information and through pilot projects that engaged staff with the numbers.

In the new reports (FIGURE 2; page 7), top-level numbers that summarize overall performance in all directions and on all branches for the whole day are given prominence. The focus is on the proportion of passengers whose waits were below a given threshold or whose trips were within a specified threshold above the normal travel time. The percentage is intended to mimic the traditional OTP number but replace the denominator with passengers rather than vehicle trips. These metrics are calculated using historical average origin-destination information that is translated into passenger arrival rates for different times of day, based on the assumption of random arrivals. The rates permit the estimation of the number of passengers arriving during a long headway and how many of these passengers waited longer than a given threshold. They also permit the estimation of train loads, so when a train is delayed at any time during its trip, the number of passengers affected can be estimated. Unlike the previous metrics, which only measure headways at the terminals and running times from end to end, these new metrics measure headways at every station along the line and running times on every link. Moreover, they take into account the number of passengers experiencing this service.

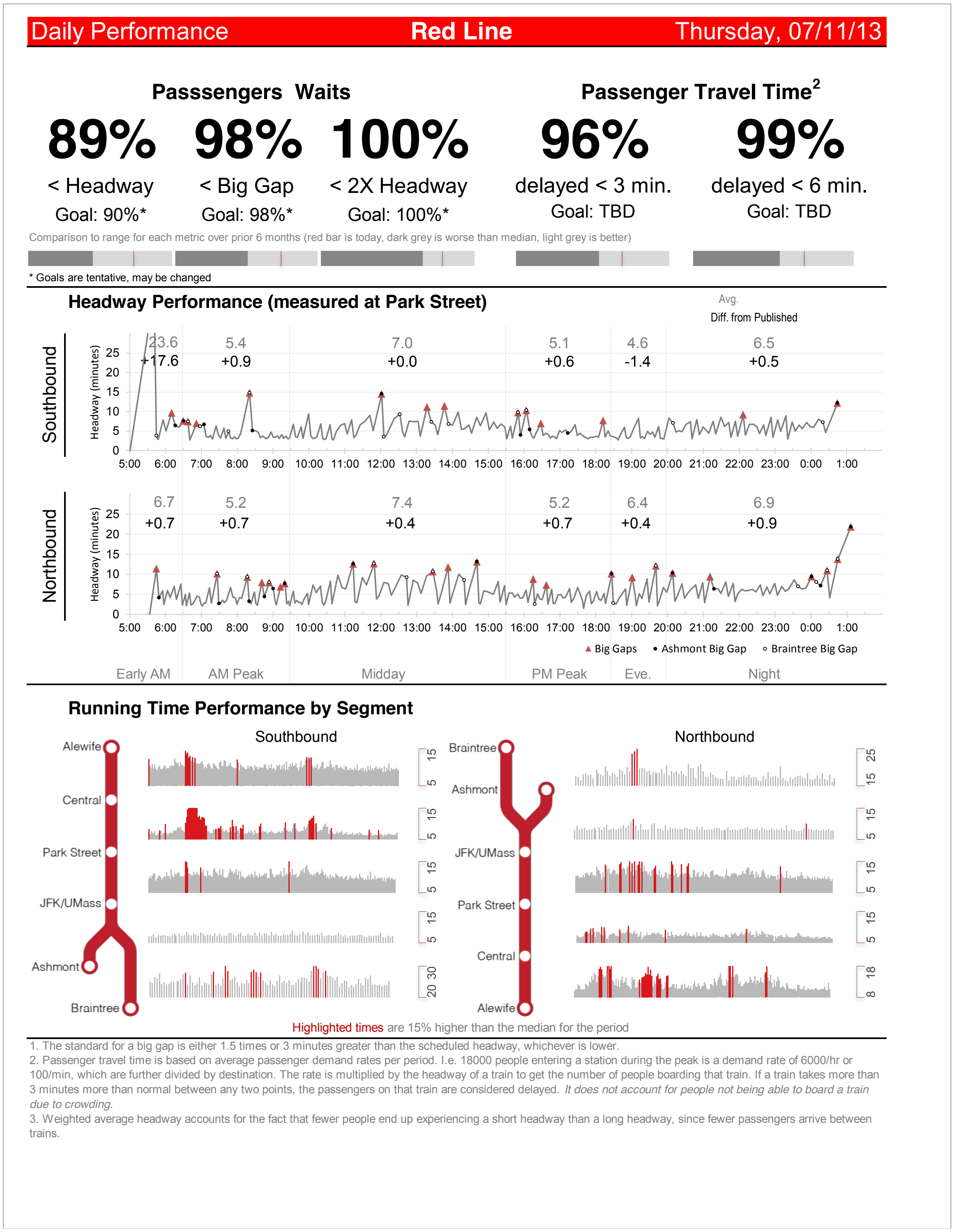

FIGURE 2 Performance Report Incorporating Passenger-Weighted Metrics.

The number in the upper left (89% < Headway in this example) represents the percent of passengers who waited less than the published headway, meaning 89% of passengers experienced an “on-time” train arrival. Likewise, the two numbers next to it indicate the percent of passengers with waits shorter than a big gap (the lesser of 1.5 times or 3 minutes greater than the published headway) and twice the published headway. The number of people waiting longer than these thresholds can be calculated by multiplying the passenger arrival rate by the difference between the actual headway and the threshold, as shown in EQUATION 1. This is the expected number of people arriving during that interval who wait longer than the published headway. Passengers arriving after this interval do not actually experience a long wait. Likewise, the number of passengers waiting longer than a big gap or twice the headway is calculated by multiplying the arrival rate by the difference between the scheduled headway and the respective threshold. These calculations can be done separately for passengers waiting for trunk and branch services, using branch-specific arrival rates and headways. Summing over all periods and both trunk and branch services provides an estimate of the total number of passengers who experienced a wait greater than the threshold.

$$\sum_{h_i > h_t} \lambda_p (h_i - h_t)$$

EQUATION 1: Passengers Affected by Big Gaps

Where:

λ_p = passenger arrival rate for the period the headway occurs in
h_i = headway for train *i*
h_t = headway threshold above which passengers are counted (scheduled headway, big gap, etc.)

The numbers in the upper right express the percent of passengers experiencing less than a certain level of delay in in-vehicle travel time. Unlike headways, there are no published travel times to set passenger expectations, so the thresholds are set as the median travel time for each period plus a buffer (arbitrarily set at 3 minutes, a level of delay noticeable to passengers and disruptive to operations). Passengers delayed beyond a certain threshold can be calculated using each train's travel time between every O-D pair and the number of passengers traveling between each O-D pair. For each origin station, a destination-specific arrival rate can be estimated using fare card data (e.g. 0.5 persons/second entering Harvard destined for Park Street). Because the MBTA does not require exit validation, destinations are estimated based on Barry et al. 2002 (*8*). For every O-D pair where a train's travel time exceeds the threshold, the number of people boarding at the origin station for that destination (the headway times the destination-specific arrival rate) are counted as delayed. The overall calculation is summarized in EQUATION 2.

$$\sum_i \sum_o \sum_d (\lambda_{od}^p)(h_o^i) \ \forall\, od^i \text{ where } RT_{od}^i > \widehat{RT}_{od}^i$$

EQUATION 2 Total Passenger Travel Time

Where:

λ_{od}^p = passenger arrival rate at station *o* for station *d* in period *p* (*p* determined by time at terminal station for train *i*)
h_o^i = headway for train *i* at origin station *o*
od^i = origin-destination pair traveled by train *i*
RT_{od}^i = running time for train *i* between stations *o* and *d*
$\widehat{RT}_{od}^i$ = acceptable running time for train *i* between stations *o* and *d*

This equation assumes that all passengers who arrive during a headway are able to board the first train that arrives, which is not always the case for long headways during peak periods when crowds build up and vehicle capacity can be exceeded. This is an important limitation that should be addressed in future performance metrics. None of the metrics in this research quantified crowding, though this is a significant factor in transit service quality.

This report also compares each metric to past performance. The objective is to quantitatively express the managers' impressions of good days and bad days. The bars below each metric place the value for that day relative to the range and median for that metric in the preceding six-month period (i.e. days in the first half of 2013 are compared with days in the last half of 2012)[2]. The light gray represents values above the median, while dark gray is below median. This additional information helps put the performance numbers in context.

In addition to the top-level summary numbers, this report uses graphical techniques to provide detail about headways and running times for every train over the course of the day. Graphs representing all headways over the day provide a disaggregate view of service, with big gaps marked by red triangles. Plotting the headway values over the course of the day on a line chart emphasizes the *change* in headway from one train to the next. Customers in theory would prefer as little variability in headways as possible, since this makes their wait time more predictable. These graphs highlight headway variations in addition to big gaps, as both negatively impact the customer experience. These graphs include an effective headway calculation[3] for the period that summarizes the variation in the chart in an intuitive unit (minutes). The graph serves as an explanation for the top-level passenger wait-time metric.

At the bottom of the page, a series of charts display running times for each major segment of a line. The standard for slow trains is 15% longer than the median, and the bars for slow trains on a segment are highlighted for emphasis. These charts provide the detail of what is driving the top-level passenger travel time metric.

4.2 Implementing Performance Reports

The development of new performance reports addresses two problems with the MBTA's existing reports. The first is the inaccuracy in representing the customers' experiences of service. The second, which stems from the first problem, is an institutional distrust of the reports and performance measurement more generally.

Measuring service at every station and weighting using passenger volumes addresses the accuracy problem. The distrust of performance reporting has also been assuaged by designing and implementing pilot projects that addressed issues and improved the numbers. The pilot projects – discussed in the next section – were intended to demonstrate the value of accurate performance reports, engage and familiarize OCC staff with the new numbers, and address service issues.

4.3 Results: Using Reports to Identify Problems and Test Solutions

Developing the new performance reports in collaboration with operations control managers provided an opportunity for the OCC director and line managers to see performance information regularly. This has revealed line segments and times with consistently poor performance on all lines. The Red Line was selected for further investigation because it carries substantially more passengers than the other rapid transit lines – 317,000 on the Red versus 197,000 on the Orange and 68,000 on the Blue[4] – and thus receives more institutional attention. It is also a two-branch line, running from either Ashmont or Braintree in the southern part of the metropolitan area through Boston and Cambridge to Alewife in the near northwestern suburbs.

In collaboration with the OCC dispatchers and managers we identified northbound coordination at the merge point and turnarounds at the northern terminus (Alewife) as the primary operational

---

[2] In the future this may be changed to the same month or quarter in the preceding year, but at the time of this research, there was not a full year of data.

[3] The effective headway is twice the average wait time, as passengers wait, on average, half the headway assuming a random arrival process. If headways are uneven, more passengers arrive during a long headway, increasing the average wait time, and thus the effective headway.

[4] These numbers are based on the analysis that generated the origin-destination information used in the reports, and thus differ from the MBTA's published figures. They include passengers who enter on another line and transfer. This measures the total number of people experiencing the service of a line. Passengers who transfer are counted on all lines they take.

challenges. Conversations with OCC staff led to ideas for pilot projects that might address these issues. The first pilot project has delayed departures from Braintree in an attempt to reduce northbound bunching on the Red Line, and the second provided additional drivers at Alewife to help turn trains more quickly in the PM peak. Refer to earlier FIGURE 1 for context.

4.4 Braintree/Ashmont Coordination

The first pilot project stems from the observation that the headway graphs in the reports were showing significant variation on the northbound trunk segment (after the merger of the Ashmont and Braintree branches). Headways frequently alternated between big gaps and bunches, as shown in the example in FIGURE 3. An analysis of historical and scheduled running times indicated that too much time had been scheduled between Braintree and the merge point at JFK/UMass. The scheduled running time between Braintree and JFK/UMass was several minutes longer than the observed median. The pilot project initially delayed departures from Braintree by two minutes, which is the approximate difference between the median and scheduled run times. It then evaluated the change in northbound headway regularity.

In discussing adjustments to the Braintree departure times, multiple dispatchers said that they knew the schedule was inaccurate because they frequently slowed down Braintree trains to stagger arrivals at the junction, while Ashmont trains often did not have enough time to turn around. Slowing trains down for headway has a negative impact on travel time performance and often results in customer frustration. While OCC managers said they had complained to the scheduling department before, there was no quantitative evidence because the previous performance reports showed high OTP. Moreover, the OCC was supportive of this project in part because the pilot required only passive input on their part – just modifying departure times in the dispatch software.

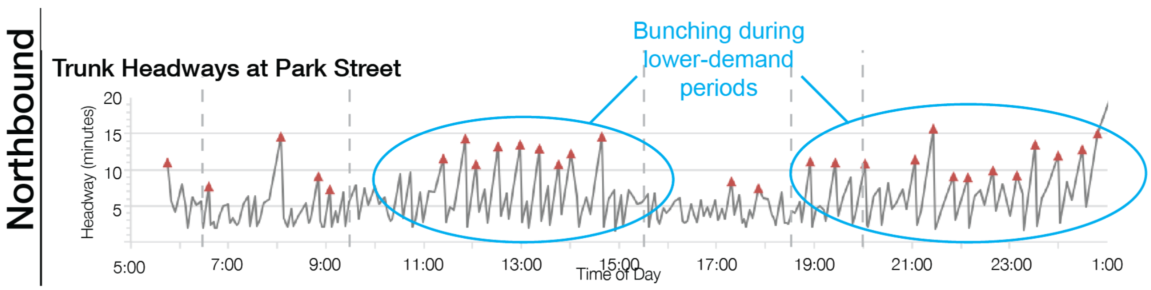

FIGURE 3 Headway chart showing long headways followed by short headways during off-peak periods, when it is unlikely that customer demand is influencing train running times. Seeing this pattern frequently on the performance reports initiated further analysis into its causes.

The results of this initial pilot were positive, and its success prompted the director of operations to request a meeting with the scheduling department to discuss the results of running time analyses. In addition to the Braintree-JFK/UMass merge timing issue, the OCC requested an analysis of train running and cycle times[5] on the Ashmont branch. Since the introduction of single-person train operation (SPTO), trains were having trouble making their next trips at Ashmont. This made coordinating the merger of the two branches even more difficult and likely exacerbated the bunching seen in the performance reports. The analysis showed that Ashmont trains indeed did not have enough recovery time at Ashmont.

The scheduling department was receptive to the running time analyses. They lack sufficient staff to analyze rail running times regularly, so these pilot project analyses have filled an acknowledged hole in their work. The scheduling department agreed to incorporate the revised branch running times into the upcoming schedule, which went into effect on January 2, 2013. The results of the schedule change have been significant improvements in headway regularity, particularly in the off-peak and at weekends. TABLE 1 shows the average effective headway (twice the average passenger wait time), which is higher with uneven headways because more passengers arrive during the longer headways. It is important to note

[5] Cycle time is the start-to-end running time plus an allowance for variability and reversing direction at the end

that increasing the turn time at Ashmont required an increased headway from 13 to 14 minutes for each branch (thus in theory a 7 minute joint headway on the trunk). The average headway *experienced by passengers* in this period, however, did not change despite this scheduled increase. Overall, the schedule changes resulted in slightly lower effective headways and slightly better reliability during most periods of the day without adding additional service.

TABLE 1 Change in Distribution of Average Effective Headways

	Median (minutes)			90th Percentile (minutes)		
Period	Pre	Post	Change	Pre	Post	Change
AM Peak	5.1	5.1	0.0	5.8	7.6	1.8
Midday	7.3	7.3	0.0	8.0	8.1	0.1
PM Peak	5.5	5.3	-0.2	6.5	6.1	-0.4
Evening	6.4	6.1	-0.3	7.8	7.1	-0.7
Night	7.7	7.8	0.1	9.0	8.7	-0.3
Saturday	9.0	7.9	-1.1	11.7	8.8	-2.9
Sunday	9.5	8.9	-0.6	11.4	10.2	-1.2

4.5 Alewife Terminal Project

The second pilot project originated from dispatcher complaints that northbound trains queue outside Alewife in the peaks. Red Line trains are scheduled to arrive and depart with 4-5 minute headways in these periods. Turning the train at Alewife also takes about four minutes: the driver has to close the doors, walk to the other end, and then reopen the doors and let passengers board. This has been exacerbated with the introduction of SPTO on March 25th, 2012, since there was no longer a conductor to help turn the train. A train arriving late blocks one of the platforms for at least one headway and backs up service outside of Alewife, aggravating customers and potentially making them miss bus connections. The pilot project arranged for additional operators[6] at Alewife who take control of the train and drive it southbound when there is insufficient recovery time. The original operator stays on the train and retakes control at Davis, while the spare operator returns to Alewife. This achieved some of the benefit of using drop-back operators while avoiding the need to rebuild the crew schedule.

The results of the initial pilot project were positive. The success of this pilot led the director of the OCC to have reserve operators stationed at Alewife for the PM peak every weekday, starting on January 2, 2013. The results of this extended pilot have been positive, particularly in relieving pressure on the Davis-to-Alewife segment, which was the primary goal. Through March 2013, the median travel time between Davis and Alewife has fallen by 15%, or about 40 seconds. The 90th percentile of running times for this segment also dropped by 40 seconds, indicating that the worst delays have improved.

TABLE 2 Running Times Between Davis and Alewife (Minutes)

Period	Median	90th	Median Savings	90th Savings	Median Savings %	90th Savings %
Pre-pilot	4.7	7.6	–	–	–	–
Pilot	3.6	6.1	1.1	1.5	24%	20%
2013	4.1	7.0	0.7	0.6	14%	8%

The average number of trips taking longer than five minutes between Davis and Alewife has dropped by 30% from 17 to 12 per day. Likewise, the average number of trips taking longer than 7 minutes dropped 30% from 5.5 to 3.8. This means fewer passengers are experiencing long holds entering Alewife. Before the pilot, 43% of PM Peak trips took longer than 5 minutes and now only 32% of PM

[6] It is unusual that an agency has additional operators during the PM peak, but because the MBTA had just moved to SPTO on the Red Line, it had a number of operators in non-driving positions around the system.

trips exceed 5 minutes. From June-October 2012, 19% of Alewife-bound customers were delayed more than 3 minutes. For the same period in 2013, that number dropped to 13%.

5 DISCUSSION

The institutional adoption of the performance reports, and the positive results from the pilot programs to improve service support the main hypothesis of this research that performance information can improve service, but that its interpretation and use is influenced by (1) how data is translated into performance metrics and (2) the process of choosing the metrics. Through its case study of the MBTA, this work concludes that data will be most effective if the measurements developed from it reflect human experience and are developed in conjunction with their eventual users.

These results imply several suggestions about how to take ADCS data, translate it into performance information that is useful for the *people* controlling service, and apply this information to change service. These are summarized below and are discussed in the remainder of this paper.

- Perspective matters: measuring from a customer perspective is a strong basis for evaluating performance
- Process matters: developing measures and reports in collaboration with service controllers makes them more likely to understand, trust, and ultimately use the information
- Design matters: performance reports should be comprehensive, concise and clear in order to provide users with as much information as they need to understand service while requiring little time and effort to digest
- Pilot projects and performance reports reinforce one another
    - Pilots show how information can be applied to improve service
    - Performance information shows the value of changes in service
    - Performance information serves as a communications tool to address problems that require coordination between departments

5.1 Perspective Matters

How an agency measures service determines what it knows and is therefore able to improve. Evaluating MBTA performance under the former on-time performance standards paint a very different picture of service quality compared with the new measures such customers experiencing big gaps and long travel times. While perhaps an obvious point, it behooves transit agencies to consider the perspective they are taking when designing or updating performance measurements. Weighting performance on each segment by the number of customers experiencing that quality of service creates a different view of service that may imply alternative management strategies or areas of focus. The current practice of measuring the performance with the train as the unit of analysis is a poor proxy for the primary purpose of a transit system, which is to move people.

To further incorporate the customer perspective into its performance measurement, the MBTA is in the process of establishing a survey panel of customers. Panelists are asked to provide their fare card number in addition to demographic information, and will subsequently be surveyed regarding their perception of and satisfaction with specific trips. This will allow the MBTA to assess whether changes in actual performance of the types detailed above are reflected in customers' perceived performance.

5.2 Process Matters

A collaborative, iterative design process that engages the eventual audience of the performance information can guide the design to more effectively generate knowledge. Knowledge is a property of people, not contained on a sheet of paper (*8*). If the intention of the performance information is to generate knowledge – an understanding of relationships that can be applied to affect service – then soliciting feedback from the users of the information provides insight into what they are interpreting from it and what they can do with it. Collaboration reveals variations in the way that different people interpret the same information and allows it to be tailored to improve clarity. Genuine collaboration respects and incorporates the opinions and experience of service managers. In so doing, it improves their acceptance of the metrics.

5.3 Design Matters

Performance reports should be concise (one page, digital or physical was the standard in this work), easy to read, and provide enough detailed information on which to base decisions. They should be as simple as possible, and no simpler. This provides additional justification for a collaborative process to determine what elements are most communicative. Operations personnel demonstrably do not have the time to seek out performance information from a reporting system. Managing performance is competing for their attention with other aspects of service provision, such as addressing equipment failures, labor and vehicle availability, and passenger incidents.

While performance reports can certainly be information-rich, this also demands that close attention is paid to information design. Too many numbers or repeated graphs can result in information fatigue, as this research learned through the intermediate drafts of its reports. Graphics can be useful for condensing information.

5.4 Pilot Projects Build Support for Performance Information, and These Reinforce One Another

Pilot projects based on analyses of performance information build support for and acceptance of performance information in general. The close monitoring of pilot projects using the performance reports enabled a quick evaluation of their effectiveness. It also helped to familiarize staff with the performance information in an operational context because they were seeing the reports on a regular basis. This quantitative evidence of how things are working has been important for outside researchers to gain credibility with the OCC staff and build support for future work. The positive outcomes of the Braintree pilot built confidence in the accuracy of performance information for operations personnel, which made the Alewife pilot possible. While this research did not generate any pilot projects judged as failures, such cases may still have the benefit of acquainting operations staff with performance information.

Rescheduling the Red Line based on revised running times has made the dispatchers' work easier. Moreover, it validates their experience of service. Both of these facts may help to justify the reports in their eyes. They can use them to make the case for how the work of other departments impacts service and customers. OCC staff may have accepted new reports despite them showing poorer overall performance in part because they confirmed their intuition that the trains were poorly scheduled. This is an important role for performance information that was not considered explicitly in the initial designs of the performance reports.

The ability of pilot projects to disrupt the status quo and make people think critically about service is an important part of their ability to influence service quality. Turning trains around quickly at Alewife required dispatchers and station managers to actively engage in train departures, and this added to the benefit of the pilot procedure. As the procedure has become institutionalized, its performance benefits have slightly diminished, in part due to less active management of Alewife.

This suggests that a mechanism to maintain a high degree of attention to service quality after the novelty of a pilot procedure wears off is critical to maintaining the benefit of service changes. Weekly performance reporting may be a part of this, providing managers with regular insight into how their efforts are working while smoothing out day-to-day variations in service. Making reports available to customers and to the general public also increases the pressure for consistently high-quality service, though this work did not succeed at taking performance information to this point.

After the pilot projects were extended beyond January, the OCC director asked if the reports could be produced daily, indicating that he saw management value in them. The COO and OCC director started getting daily performance reports for the Red, Blue, and Orange Lines in March 2013. This request represents a second-order impact of the pilot project strategy. The first is the actual improvements in service, while the second is an increased institutional appetite for information and innovation.

6 CONCLUSION

The reports and projects described in this paper are instances of how simple spatial and temporal data for vehicles and customers can result in both ongoing and targeted improvements in institutional processes and service using a collaborative process and focusing on design. The opportunities for expanding beyond this initial work are wide ranging. Incorporating additional data such as traffic, incidents, or disruptions adds another dimension to this data. These additional data sources have the potential to increase the amount information and knowledge transit agencies have.

However, as the number of data sources and dimensions of analysis increases, so does the need for collaborative information design. Adding information adds new elements that a viewer must interpret and relate. These tasks need to be considered and facilitated through design, then verified through collaboration. This research suggests that the process of collaboration is at least as important as the design itself in influencing the interpretation and responses of viewers. As long as the eventual application of the information relies on human action, ensuring that end users can easily interpret the information is critical to enabling them to apply it effectively.

ACKNOWLEDGEMENTS

The authors would like to thank the MBTA for both their financial support and cooperation during this research. Providing the opportunity to work closely with those in the operations control center and pilot changes in service has given this work a dimension that would not have otherwise been possible.

REFERENCES

1. Wilson, Nigel H. M. "Roles for Richer Data and Better Analysis in Improving the Effectiveness of Transit Systems." Latsis Symposium 2012. 2012 Sep. http://latsis2012.epfl.ch/files/content/sites/latsis2012/files/Presentations/Wilson.pdf (accessed 2013 July 26).
2. TCRP. TCRP Report 88: A Guidebook for Developing a Transit Performance-Measurement System. Transit Cooperative Research Program, Washington D.C.: Transportation Research Board, 2003.
3. MTA. Key Performance Indicators. Metropolitan Transportation Authority. 2013. http://www.mta.info/persdashboard/agencies/nyctsubway/cp/392289_indicator_narrative.htm (accessed May 5, 2013).
4. Uniman, David, John Attanucci, Rabi Mishalani, and Nigel H. M. Wilson. "Service Reliability Measurement Using Automated Fare Card Data." Transportation Research Record (Transportation Research Board) 2143 (2010): 92-99.
5. Behn, Robert. "The Dilemmas of Innovation in American Government." In Innovation in American government : challenges, opportunities, and dilemmas, edited by Alan Altshuler and Robert Behn, 3-37. Washington D.C.: Brookings Institution Press, 1997.
6. Altshuler, Alan, and Marc Zegans. "Innovation and Public Management: Notes form the State House and City Hall." In Innovation in American government : challenges, opportunities, and dilemmas, edited by Alan Altshuler and Robert Behn, 68-80. Washington D.C.: Brookings Institution Press, 1997.
7. MBTA. "Service Delivery Policy." Massachusetts Bay Transportation Authority, Boston, 2010.
8. Barry, James, Robert Newhouser, Adam Rahbee, and Shermeen Sayeda. "Origin and Destination Estimation in New York City With Automated Fare System Data." In *Transportation Research Record Journal of the Transportation Resaerch Board*, No. 1817, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 183-187.
9. Rowley, Jennifer. "The wisdom hierarchy: representations of the DIKW hierarchy." Journal of Information Science 33, no. 2 (2007): 163-180.