Making Data Matter: The Role of Information Design and Process in Applying Automated Data to Improve Transit Service

by

Dominick Tribone

B.A., Economics and French
New York University, 2007

Submitted to the Department of Urban Studies and Planning
in Partial Fulfillment of the Requirements for the Degree

of

Master in City Planning

Submitted to the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for the Degree

of

Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

Signature of Author ....
Department of Urban Studies and Planning
Department of Civil and Environmental Engineering
May 23, 2013

Certified by ....
Frederick P. Salvucci
Senior Lecturer of Civil and Environmental Engineering
Thesis Supervisor

Accepted by ....
Professor P. Christopher Zegras
Chair, MCP Committee
Department of Urban Studies and Planning

Accepted by ....
Heidi M. Nepf
Chair, Departmental Committee for Graduate Studies
Department of Civil and Environmental Engineering

Making Data Matter: The Role of Information Design and Process in Applying Automated Data to Improve Transit Service

by

Dominick Tribone

Submitted to the Department of Civil and Environmental Engineering on May 23, 2013 in Partial Fulfillment of the Requirements for the Degree of Master of Science in Transportation

Submitted to the Department of Urban Studies and Planning on May 23, 2013 in Partial Fulfillment of the Requirements for the Degree of Master in City Planning

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. With automated systems producing large quantities of high quality data, it becomes the impetus for, rather than simply the input to, measurement. Capturing more value from automated data thus involves rethinking what agencies can know about service.

This research uses the Massachusetts Bay Transportation Authority (MBTA) as a case study. It first assesses how the MBTA currently uses real-time and historical data. Based on this assessment, it redesigns and advances the agency's daily performance reports for rapid transit through a collaborative and iterative process with the Operations Control Center. These reports are then 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 eventual users. It also finds that developing pilot projects during this collaborative process enables new performance information to result in service improvements. Based on these findings, this work produces a set of recommendations for generating useful performance information from transit data, as well as a specific set of recommendations for expanding the use of data at the MBTA.

Thesis Supervisor: Frederick P. Salvucci
Title: Senior Lecturer in Civil and Environmental Engineering

Thesis Reader: David Block-Schachter
Title: Research Associate in Civil and Environmental Engineering

Acknowledgements

Firstly, thank you David for your countless hours of assistance and guidance from the beginning though to the completion of this work. Your conceptual and technical input has been invaluable. Without your direction, this work would not have come close to accomplishing what it has, and you deserve as much credit as I do for the final product.

Josh, thank you for getting this work started and keeping it going at the MBTA. You made us realize that it is possible to make things change and pushed us how to make it happen.

Fred, thank you for convincing me to change my thesis topic to focusing on improving performance information, and then for your supervision and direction of this work. Your insight and perspective on transportation, politics, public sector management, and dozens of other topics form some of my most valuable knowledge from my time at MIT.

Todd, Sean, and the rest of the MBTA OCC, thank you for making this work possible. Without your cooperation and interest, these reports would be confined to a library shelf rather than in your hands every morning.

Nigel, John, Haris, and Mikel, thank you for your input and direction on developing the performance metrics and pilot projects. Your collective wisdom has been of great value.

Sandy, Kirsten, Kris, Ginny, and the rest of the DUSP and CEE administrative staff: thank you for always being there through all of my questions and confusion and for helping me successfully navigate the MIT administration (and always having candy).

To my fellow students in the Transit Lab, thanks for keeping the office fun and for volunteering to spend your precious hours monitoring train stations during the pilot projects.

Yonni, Laura, Cecilia, and Jamie (The Glamazon), thank you for making the MST program and transit lab fun and always being there to laugh with me.

Stephen, thank you for teaching me just about everything I know about design, for stimulating my interest in how people relate to information, and for making my time at MIT some of the most fun I've ever had.

To my friends who have kept me sane and laughing throughout my three years of grad school, thank you for always providing motivation to finish my work so I could go have fun.

Thank you to my parents, Mary Michael and Rich, and my sister Allegra for supporting me in everything I do. You have given me the roots and wings to succeed.

And finally, to Ian, who has put up with three years of me being busy, stressed, tired, and far away: Thank you for your understanding, for your support, your hugs, and for listening to me rant about transit service any time we try to go anywhere. You're the best.

Contents

Tables

Figures

Acronyms and Abbreviations

ADCS – Automated Data Collection Systems
AFC – Automated Fare Collection
ATD – Automatic Train Departure
ATO – Automatic Train Operation
AVI – Automatic Vehicle Identification
AVL – Automatic Vehicle Location
BRT – Bus Rapid Transit
CAD – Computer Aided Dispatch
COO – Chief Operating Officer
COV – Coefficient of Variation
DIKW – Data-Information-Knowledge-Wisdom
DTC – Downtown Crossing station
EJT – Excess Journey Time
GM – General Manager (of the MBTA)
GPA – Grade Point Average
GPS – Global Positioning System
ICT – Information and Communications Technology
JTM – Journey Time Metric
LOS – Level of Service
MBTA – Massachusetts Bay Transportation Authority
O-D – Origin-Destination pair
OCC – Operations Control Center
OTP – On-time Performance
RBT – Reliability Buffer Time
RFID – Radio Frequency Identification
SD – Standard Deviation
SPTO – Single Person Train Operation
SQL – Structure Query Language
STM – Société de Transport de Montréal
TCRP – Transit Cooperative Research Program

1 Introduction

The world is being flooded with data. According to IBM, 2.5 trillion gigabytes of data are generated each day, from weather forecasts to credit card transactions to social media posts (IBM 2013). These records, often referred to as big data, permit an understanding of the world that is both more detailed and more accurate than previously possible. While public agencies are becoming data-rich as they upgrade their technological systems, many of their institutional processes and behaviors developed when they had little reliable information about their customers or their performance. This is particularly true of public transit agencies, who until recently relied on surveys and manual 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. Big data has the potential to change the way public transit agencies operate by providing them with better information on which to base decisions. The presence of good information, however, is a necessary but insufficient condition for physical improvements to service. Improving an agency's operations also requires understanding how to make this information meaningful to those in control of service and how to make old institutional processes responsive to new information.

This research focuses on Boston's Massachusetts Bay Transportation Authority (MBTA) as a case study of a data-rich agency that has not fully integrated new information into its operations. In the past decade the MBTA has installed new systems that produce detailed data about where vehicles are (Automatic Vehicle Location, or AVL) and where customers enter the system (Automated Fare Collection, or AFC). The primary use of AVL data has been to facilitate real-time service management, while AFC has been aimed at improving revenue management. More recently, vehicle locations and arrival times have also been released publicly (NECN 2010). The general customer satisfaction with this information has generated enthusiasm from the state Secretary of Transportation. His desire to do more with the MBTA's data was the genesis of this work. The MBTA knows more about performance in the moment than performance in the past; its use of logged data has been limited. The agency could use its data to better understand trends, learn from them, and make improvements. But the agency is a bureaucratic organization that relies on human action, human perception, and existing institutional processes, which constrains the use of such data. The MBTA provides an opportunity to explore how to make data useful within the existing constraints faced by a U.S. public transit agency.

This research 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 attempts to determine 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. Their positive impact on service has led them to be extended beyond their initial phase.

This research shows that when a system is run by humans, the interpretation and use performance information is influenced by (1) how data is translated into performance metrics and (2) the process of choosing the metrics. This in turn affects how the information is incorporated into the management of the system and thus how it can ultimately impact an agency's operations. Through its case study of the MBTA, this work concludes that big transit data will be most effective if the measurements developed from it accurately reflect human experience and are developed in conjunction with their eventual users. Based on these findings, it produces a general set of recommendations for creating useful performance information from big transit data, as well as a specific set of recommendations for expanding the use of data at the MBTA.

1.1 The Age of Big Data and the Public Sector

Over the past several decades the introduction of information and communications technology (ICT) into many parts of society has exponentially increased the amount of data collected about the world. These technologies are logging information that has the power to change how human beings understand the systems, processes, and events that impact their existence. The increasingly common presence of sensors and electronic transactions is creating a frequent record of the systems that people use in their daily lives. This in turn is making it easier to learn more about the world we live in and make more informed decisions about how to influence it. Urban planners (Evans-Cowley 2011) and ICT experts (Falconer and Mitchell 2012) have posited that this wealth of new information has the potential to transform how cities are managed and how their denizens interact with them.

Analyzing, interpreting, and applying knowledge from big data has been a key to success for many different organizations. Hedge funds and other new investment entities analyze market data along with other trends to predict and take advantage of market fluctuations. The Internet giant Google frequently tests new strategies through randomized trials where different users see slightly different content. The company then analyzes the results for patterns, trends, and correlations, which informs the final design or product (Christian 2012). The Obama campaigns in both 2008 and 2012 analyzed voter data in great detail, which allowed more targeted and effective campaigning (Issenberg 2012). In all of these examples, the ability to analyze and draw conclusions from big data produces a competitive advantage that contributes to the success of the organization.

Public agencies have not been left out of this trend. They are also getting more data about customers and their behavior, particularly in the transportation sector. However, the nature of the data varies by mode. The auto system, which is dominated by local roads, has a limited – though growing – amount of information. Traffic monitoring data from loop detectors, satellites, and roadside sensors provide detailed information about road use and congestion in real-time to both managers and drivers. However, these systems only provide aggregate information; they do not track individual behavior. Electronic tolling, by contrast, produces detailed information about where individual vehicles enter and exit toll facilities, whereas they previously only knew aggregate entries and exits at each interchange. Most information on how people are traveling in cars, however, is based on household travel surveys, which are costly and disaggregate analysis often limited by a small sample size. Real-time location data from GPS both in vehicles and smartphones has the potential to provide more detailed information on individual travel behavior. Transportation agencies have begun to obtain detailed location data through GPS-based household travel survey devices and research is advancing in using GPS signals for other devices to observe travel behavior (Chen, et al. 2010). However, these are still samples that require participant consent. The New York Police Department is implementing a project to record the license plate number of every vehicle entering and exiting Manhattan (Sledge 2013), a technology that has the potential to provide public agencies with more detailed vehicle travel patterns. However, comprehensive data from license plate data on auto origins and destinations would require a more expansive installation, which may face privacy concerns.

Bike share systems, which came into existence in the digital age, are the opposite case. Their operations are dependent on the provision and analysis of big data. Customers are uniquely identified so individual behavior can be tracked. Bike availability and station capacity are electronically monitored and provided to customers via mobile applications. Real-time data on station capacity is combined with historical information about demand at different times of day to determine when and where bikes need to be moved by rebalancing trucks.

Public transit agencies are between these two extremes in terms of what they know about their customers. Electronic fare collection technologies record the boarding station (off-board fare collection) or vehicle (on-board fare collection) for each customer. Systems that require exit validation (like London's Underground or Washington D.C.'s Metro Rail) also record data on where customers exit. New dispatching technologies display vehicle positions at all times, and also log and archive them. This data can be used to inform management decisions in real time, such as holding or re-routing service due to delays. Real-time information can be also provided to customers to give them more information before and during their trips (Wilson 2012).

Public agencies have not been as thorough in analyzing historical data and applying it to improve their operations as private companies. They do not know as much about their customers, and their customers do not know much about them. This may be because they are not subject to competition like private sector companies or politicians. Many transit agencies existed prior to the availability of comprehensive data. They developed planning procedures and operating behavior in a context without good information about the service they were providing or the customers they were serving. Knowing the distribution of trip times to schedule a route required analyzing manual records of terminal departures and arrivals. These were expensive and time-consuming to collect, and subject to human error. With automatic vehicle location, running times are calculated automatically for each trip and can be easily analyzed. Knowing how many passengers were on a bus or train required manual sampling with ride checkers. Now automated systems count boardings on every trip, providing census rather than sample data and allowing for more detailed and reliable analyses. Agencies can thus substitute automated data into their existing analysis processes (Wilson 2012). Despite vast increases in the quantity and quality of data, however, they may not go beyond this to use data any differently than when it was limited. Customers, on the other hand, are getting more information about many other goods and services they purchase, and thus may expect it from transit as well.

Getting more information out of this data and applying this information to impact service requires changing the institutional processes of data analysis and use. Data was previously a limitation on analysis. It was often time consuming and costly to collect. With automated systems and their large quantities of high quality data, it can be an impetus for rather than simply an input to analysis. In addition to "What data do I need to answer this question?" automated data allows agencies to ask, "What can I do with the data that I already have?" Capturing more value from automated data goes beyond replacing the inputs to existing analyses. It involves rethinking what can be analyzed and where data can be applied to improve operations.

Big transit data has the potential to improve agencies' service provision and customer satisfaction. The data advantage that transit has over the auto system could be leveraged to streamline operations and tailor service to attract more riders, potentially bolstering transit's share of the market.

1.2 The MBTA: A Case Study of Big Data in Public Transit

The MBTA is one of the fortunate transit agencies for which the age of big data has arrived. Substantial investments in new technological systems such as AVL, Automatic Train Operation (ATO) provide information about the system to dispatchers in real-time. Despite these "automated" and "automatic" systems, the service is still run by humans. This means that to influence physical outcomes, information must be interpreted and applied by people.

The MBTA's current use of the data from these systems has focused on real-time management, but they are also continuously archiving detailed information about the transit network. Providing bus and rail dispatchers with vehicle locations[1] in real-time gives MBTA personnel an understanding of their network at a point in time. This enables more informed decisions about operations control – holding a train to space out service or advancing a departure from the terminal to free up a platform for an incoming train. More recently, the agency has also begun providing real-time information to customers, who previously had little information beyond the published schedule and what they can see or hear at the stop or station. In 2010, the MBTA opened a real-time feed of bus and train locations to developers (NECN 2010), who have created dozens of bus and train arrival apps. In 2012, the agency began providing real-time train arrival predictions in many of its heavy rail transit stations (on what are commonly called countdown signs). It also makes commuter rail vehicle locations and predicted arrival times available online and to developers. In this way, the MBTA's data systems have substantially increased the amount of information available to both service controllers and customers. These inform both sets of users' decisions and change the way they interact with the transit system.

The use of historic data to learn about trends and issues over longer periods of time, however, has been limited. The agency currently creates on-time performance (OTP) reports for bus and rail on a daily basis for internal use, and publishes a monthly performance report for the public. It does not, however, regularly use this data to further assess the causes or potential remedies for poor performance, despite having internal reporting systems capable of doing so. A notable exception is the recent detailed analysis of vehicle running times to revise vehicle schedules based on more accurate information. University students and consultants have analyzed the MBTA's data and provided constructive recommendations about service in past research. This work has produced some operational changes, though fewer than what have been proposed.

The MBTA provides a case where big data is available, but has not yet been harnessed to feed back into service provision. While the data has been analyzed, the fact that few changes have resulted from such analyses suggests that the problem is not solely analytical. It thus

[1] All heavy rail and bus vehicles. The Green Line and Mattapan line light rail do not have real-time vehicle locations at the time of this writing, though a project to implement this is underway.

provides an opportunity to explore the other factors influencing the ability of big data to impact public transit operations, and how to overcome current limitations.

1.3 Role of This Research

Working with the MBTA as a case study, this research explores how to make data analysis more influential in transit operations. It rethinks not only the analytical methods but also where and how information is created and applied within the organization. This work turns to the Data-Information-Knowledge-Wisdom (DIKW) continuum for thinking about how transit data can become knowledge for management. Described in more detail in Chapter 2, the DIKW continuum is a conceptual framework for understanding how data – unorganized observations or facts about the world – become meaningful and useful knowledge (and eventually wisdom) that humans can apply to make decisions and influence their environment. It also looks at past work on innovation in the public sector to understand the constraints and opportunities for introducing change in a public agency like the MBTA.

Translating data into information that is understandable to people has been one focus of past research, and 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. After assessing the MBTA's current use of big data, this research finds that real-time information has changed the way service is managed. Historical reporting based on this data, however does not have a significant impact on operations. In evaluating the MBTA's current OTP reports, this research finds that the reports for rail are ineffective because they do not accurately measure service. The reports for buses measure service more accurately, but are too numerous and lengthy. This hinders interpretation and limits the ability of staff to identify problems and opportunities. Past attempts to produce service changes by analyzing data have not been implemented due in part to insufficient attention to institutional constraints and processes.

One implicit assumption in past research has been that conducting an analysis and presenting the results to those in control of service provides sufficient motivation to change service. This work takes a different approach: performing data analysis in conjunction with service controllers through a collaborative and iterative process. It solicits their response to information and incorporates their input. Through this process, this work revises the measurements and their presentation. It reorients performance metrics for rapid transit around passengers by combining data from two sources: vehicle location and fare collection. Close attention is paid to making information clear and legible, while still retaining appropriate detail to underlie management decisions. This process reveals the institutional limitations to applying information, which include poor interdepartmental communication, a lack of time and staff to do analysis and look for solutions, and a distrust of information with unclear origins. It does not, however, bring this data "full-circle" by releasing it to customers so they can see a quantification of their experience on the MBTA.

The original intent of this work was to produce information for passengers that provides more insight into MBTA service than their everyday experience. New performance reports were developed with public viewers in mind. These reports were more detailed than the MBTA's internal reporting tools, and were shown to operations staff so that they would understand and have input into what the public sees. Providing better information internally then became the priority for this research, so that operations staff could manage service to the measures being made public. As the detailed reports evolved, some areas of poor service became apparent. The research then expanded to from performance reporting to using the reports to produce service improvements.

Having established a relationship with the operations control center while developing new performance reports, this research proposed two service improvement projects that were successfully piloted and eventually implemented. These include (1) rescheduling the MBTA's busiest line, the Red Line to better coordinate northbound service and (2) staffing addition personnel at its northern terminus to speed turnarounds and reduce delays in the PM peak. This research hypothesizes that three factors enabled the data to be translated into service improvements: (1) changing the way service was measured, (2) changing how these measurements were presented, and (3) developing them in close coordination with the their eventual users.

1.4 Implications

This research looks at the human and institutional dimensions of making big data matter for transit agencies. It has successfully redesigned heavy rail performance reports for the MBTA and has piloted two service improvements that were initially successful and have been extended (and may become permanent). It provides the MBTA and other transit agencies with suggestions about how to turn their big data into information and how to apply it to change service. It finds that competent data analysis is not sufficient to lead to operational changes, and that the human interpretation of and reaction to information must be considered.

This work concludes that performance information should be developed not only based on the input of upper management but also those actually in charge of service. To have an impact, information needs to be meaningful to and trusted by those in direct control of service. Incorporating their feedback helps to ensure this. Design and presentation also play key roles in enabling service controllers to draw useful conclusions from performance information.

This research also finds that developing pilot projects during the collaborative information design process can be a successful strategy to produce changes in service. Pilot projects and performance reports reinforce one another: the reports make an initial case for a pilot project, and implementing the pilot shows how performance information can be used to impact and improve operations.

1.5 Organization of This Research

Chapter 2 will discuss the DIKW framework along with examples of how past work on transit performance measurement fit into it. It will also discuss past research on successful innovation in public sector bureaucracies, moving beyond knowledge to action.

Chapter 3 introduces the MBTA and assesses its current applications of automated data. It identifies the successes and shortcomings of the existing practice, which form a basis for redesigning the MBTA's performance reports for heavy rail.

Chapter 4 describes the process of developing new performance reports for the MBTA's heavy rail services. It focuses not only on changing how service is measured, but also how these measurements are presented. It emphasizes the benefits of collaborating with the MBTA's Operations Control Center (OCC) and how their input has improved the end product.

Chapter 5 details how this performance information has been applied to modify service through two pilot projects conceived and implemented in coordination with the OCC. It discusses both the institutional process of designing and implementing the pilots – how institutional resistance to change has been overcome – and the resulting impact on service.

Finally, Chapter 6 draws lessons from the experiences related in the previous chapters and provides a set of recommendations about applying the findings of this research to additional operations within the MBTA and at other transit agencies.

2 Theoretical Framework and Previous Research

Automated Data Collection Systems (ADCS) accumulate millions of records every day, but these alone do not provide much value to transit operators. This research employs the Data-Information-Knowledge-Wisdom (DIKW) hierarchy as a framework for thinking about the use of data in public transit agencies. This chapter explains the DIKW concepts and then reviews past work on ADCS and transit performance within this framework. The literature on transit performance has developed a variety of tools and methods for extracting meaning from ADCS data. In some cases, the application of these methods to analyze ADCS data has resulted in changes at transit agencies, while in others it has not.

To gain insight into what contributes to some performance information being successful in generating change and some not, this research turns to work on performance management and innovation in the public sector. This literature discusses how public agencies have been able to modify their operations despite institutional resistance to change. The literature also proposes several characteristics of successful innovation that help explain why this research was successful in making changes at the MBTA. These include: alleviating widely-recognized problems, finding support at multiple levels of the institution, being close to those in charge of service, and being open to feedback.

There is little research linking these two bodies of literature, exploring how to leverage data to make institutional progress. None of the literature reviewed examines how the process, design, and institutional context of performance measurements influence the capacity for and effectiveness of performance management. This research to begins to address this gap.

2.1 The Data-Information-Knowledge-Wisdom Hierarchy

Though the terms data and information or knowledge and wisdom are synonyms for one another in common parlance, in information science and knowledge management each of these four words represents a distinct concept. These concepts are often arranged in a hierarchy intended to represent how humans come to understand the world. A study by (Rowley 2007) reviews the information science and knowledge management literature and summarizes definitions of the four concepts in the DIKW hierarchy. This chapter draws on Rowley's review to define the concepts data, information, knowledge, and wisdom in the public transit context.

Data is the base of the hierarchy, the foundation on which information, knowledge, and wisdom are built. Data are defined as events, observations, or other facts that are discerned and/or recorded either by people or machines (Rowley 2007). Data are usually described as unorganized and unprocessed, having little meaning because they lack context and relation to one another. Examples of data in the transit context are records of a vehicle locations in the time and other identifying information from the AVL system. This data tells an agency where a vehicle was at a given point in time. Without organizing the records and relating them to one another, there is no further detail about a vehicle's path, how long it took to get between two points (running time), or the spacing of vehicle arrivals at a stop (headway).

These latter concepts are *information* that can be created from transit data. Information is generally described as data that have been formatted, organized, processed, aggregated, calculated, and otherwise manipulated, and which then take on meaning, value, or usefulness (Rowley 2007). The fundamental concepts defining information are structure and meaning, which the raw factual signals or observations lack. Human action is required to manipulate data so it describes something beyond what the initial observations and signals show. Continuing with the AVL example above, the path, travel time, and headway are all information that results from relating AVL data points. This information is useful for describing the characteristics of a bus trip, for example. Combining multiple pieces of information produces information, such as the average running time for a route or the distribution of headways. Both a single headway and the distribution of headways are information. They both relate data and have value, but describe different aspects of service. What the information describes influences what knowledge can be derived from it.

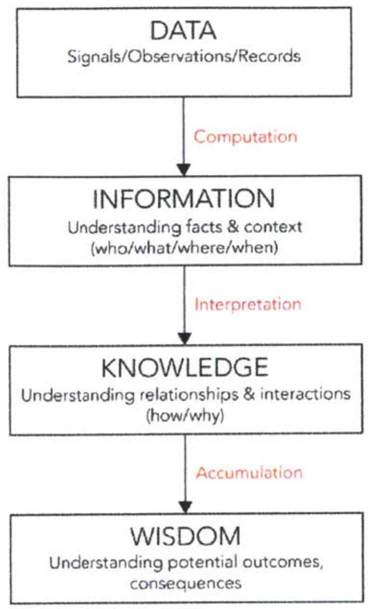


Figure 1: Summary of the DIKW Framework

The distinction between information and *knowledge* is more subjective than that between data and information. Rowley found that definitions of knowledge are more complex and various than those of either data or information. Many sources portray knowledge as personal and subjective. One of the texts she reviewed notes that "While data is a property of things, knowledge is a property of people that predisposes them to act in a particular way" (Boddy, Boonstra and Kennedy 2005). Rowley's review suggests that information is transformed into new knowledge through understanding its relation to other information and existing knowledge (Rowley 2007). A synthesis of the various definitions of knowledge is understanding relationships and interactions among different pieces of information in a way that permits one to take action. It is understanding what the problem is, what can be done about it, and how an action will address it. This action-oriented definition of knowledge will be employed in this research. For example, knowing the headway or the distribution of headways for a particular route and relating it to the scheduled headway provides information about whether the route is running well. Combining this with information about on-time departures, traffic, and incidents creates knowledge about what may be causing unscheduled variation in headways, allowing one to propose potential solutions.

While moving from data to information is computational, moving from information to knowledge is interpretational. The knowledge that one can gain depends on what information is available and how it is presented. This transition is more of an art than a science –meaning hinges on the response of the viewer. While turning data into information relies primarily on mathematics and programming, creating knowledge from information relies on human perception. Two people can interpret the same information differently, which is why knowledge generation is described as subjective. In an institutional setting, this process may be circumscribed both by the decisions about what information to supply and instructions on how to interpret it. Developing new knowledge may thus require changing institutional norms around information.

In many of the texts reviewed in Rowley's study, knowledge was the pinnacle of the hierarchy, the highest level of understanding. Only three of the 16 textbooks in Rowley's review included *wisdom* in their hierarchy. These three definitions all focus on the generalized nature of wisdom, which allows one to react and apply knowledge to new situations. The texts do not provide much insight into the generation of wisdom, except that it is accumulated knowledge. While knowledge is understanding a specific situation and being able to influence it, wisdom is being able to apply knowledge generated in one context to a new situation. In the transit context, wisdom is what enables dispatchers to manage service. From their experience they derive knowledge of the causes of poor performance and the effects of their actions. The accumulation of this knowledge constitutes wisdom about the performance of the system and their ability to influence it. This wisdom allows dispatchers to react to new situations as they arise.

Figure 2, from (Rowley 2007), is a representation of the DIKW hierarchy. Rowley adds two

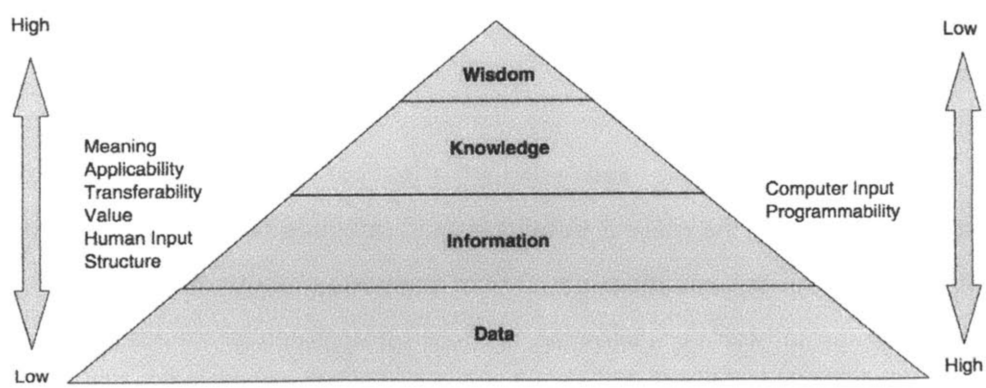


Figure 2: Pyramid representation of the DIKW hierarchy, from Rowley (2007)

continua to the pyramid: the continuum on the left shows characteristics that increase in the transition from data to wisdom while the continuum on the right show characteristics that decrease. This conceptualization reinforces the idea that data are raw representations of the world around us, which through structure and interpretation become intelligence that allows humans to understand and influence the world. Additionally, as human input, value, and applicability increase, programmability and computer input decrease. It is fairly easy to create automated processes for creating information from data; it is more difficult to automate the creation of knowledge.[2]

2.2 Past Work on ADCS in the Transit Context

The introduction of 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, a small fraction of which will be reviewed in this chapter. Wilson (2012) provides a general overview of the information that can be generated from AVL and AFC data:

- Detailed characterizations of route segments and running times;
- Detailed characterizations of stop activity;
- Detailed characterizations of passenger activity.

In providing guidance on developing performance management plans for transit agencies, Transit Cooperate Research Program (TCRP) Report 88 (2003) 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 (Wilson 2012). TCRP Report 88 includes hundreds of possible performance metrics. The most pertinent that can be calculated readily from AVL[3] and AFC data are listed in Table 1 below.

[2] This, however, is the objective of artificial intelligence

[3] In this table, AVL is used to describe any system showing vehicle location, not just GPS-based bus tracking.

Measurement	Definition/Calculation	Data System
Frequency / Headway	Number of vehicles or time between vehicles	AVL
On-time Performance / Schedule Adherence	Departure/arrival of a vehicle relative to its schedule	AVL
Service Regularity	Percentage of trips that operate within a specified range of the scheduled headway	AVL
Missed Trips	Scheduled trips not run	AVL
Running time	Time for vehicle to move between two points	AVL
Run-Time Ratio	Ratio of observed to scheduled run time	AVL
Passenger Load	Number of people on a vehicle	AFC
Travel Time (passenger)	Time for a passenger to go from origin to destination	AVL & AFC
Travel Time variability	Variability in travel time, measured as standard deviation, coefficient of variation, or other distribution statistics	AVL & AFC
Reliability factor	Percent of trips that are within a specified percentage of the average travel time	AVL & AFC
Delay	Actual run-time minus scheduled run time	AVL
Excess wait time	Number of passenger-minutes of wait time greater than expected wait time	AVL & AFC
Big Gaps	Headways over a specified threshold	AVL

Table 1: Performance Measurements That Can Be Calculated From ADCS

2.2.1 Use of ADCS for Performance Metrics in Transit Agencies

New York City Transit calculates a "wait assessment" metric that is a measure of service regularity. It is defined as the number of headways that are less than 125% of the scheduled headway (MTA 2013), which is the inverse of a big gap metric. The agency sets targets for wait assessment in addition to terminal on-time performance (OTP) to manage service. The London Underground uses travel time and its variability to judge service quality. Its Journey Time Metric (JTM) calculates customers' time between entering the system to leaving (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 (Uniman, et al. 2010). The difference is the Excess Journey Time (EJT), which the Underground managers use to evaluate service.

2.2.2 Work Focusing on Translating ADCS Data into Performance Information

Automated data is still relatively new, so research is still developing new analytical and computational methods for drawing useful information out of it. Barry et al. (2002) use AFC data for the New York City Subway, where passengers are only recorded on entry, and infer destinations based on the sequence of entries over the course of a day. Building on this, Gordon combines AVL and AFC data to infer origins, destinations, and transfers for passengers in London's entire public transport network (Gordon 2012). This provides Transport for London with much more detailed demand information, which enables them to improve service planning, market research, and other functions.

In addition to developing analytical methods to get more value out of automated data, other research has built on the standard measurements to create more complex metrics that capture multiple dimensions of service. Uniman (2010) uses Transport for London's AFC data to create a reliability buffer time (RBT) metric. Uniman defines RBT as the "amount of extra time that passengers must budget above the typical journey time in order to arrive on time at their destination with a specified level of certainty." It is calculated as the 95th percentile minus the median running time for a segment or O-D pair. Schil (2012) looks at excess RBT by comparing the RBT for a typical day to the RBT for the disrupted day. He uses this to measure the severity of service disruptions.

Generally, the transition from data to information is conceptually straightforward, involving computations that can be done by any spreadsheet, statistics, or database software. A substantial number of methods for translating ADCS data into information have been developed that effectively characterize many dimensions of public transport service. While new information may not result in service changes, the use of ADCS is not limited by a lack of understanding of how to translate data into information.

2.2.3 Information, Knowledge, and Wisdom

Though past work with transit ADCS has not been discussed within the DIKW framework, many past studies have analyzed large datasets and then applied this information to answer specific questions. In doing so, this research has generated new knowledge about transit service based on more detailed information.

Shireman (2011) uses MBTA AVL data to explore opportunities for more productive vehicle scheduling. Shireman's analysis first generates more detailed information about bus running times. It then explores how changing certain operating assumptions and constraints in the MBTA's scheduling software could produce a more efficient schedule. Shireman codifies this knowledge in his thesis, but his specific findings have not yet been applied by the MBTA to modify its vehicle schedules. The MBTA has, however, begun to use the software and approach from Shireman's work to reschedule its routes.

Other work has attempted to identify and resolve issues on the MBTA's Green Line, a light rail line with a downtown subway and four surface branches. Malikova (2012) uses vehicle location records for the MBTA to assess the impact of introducing three-car trains on the line. Her analysis produces information on running time and headway performance before and after three-car trains began running. From this information, Malikova shows that current implementation of three-car trains had increased headways and bunching in the downtown subway. Based on this knowledge, she proposes alternate implementation schemes that could avoid this issue.

Automated data from other agencies has also been analyzed and applied to improve service. Frumin (2010) analyzes Transport for London's AFC records to characterize both passenger behavior and service quality on the London Overground. This new information generates knowledge of how uneven scheduling on the North and West London lines influences passenger behavior and travel experience. London Overground has applied this knowledge to create a new vehicle schedule that provides more regular service. Frumin uses the metrics developed in his work to evaluate the change and concludes that the new schedule has a positive impact on customers and service quality.

San Francisco's Municipal Transportation Agency (Muni) analyzes ADCS data to develop knowledge about problems and propose changes that address poor performance. Analyzing train turn times has led to a revised turning procedure that reduced turn times. Evaluating bus schedule adherence and supervisor placement led to relocating some supervisors, which has improved departure adherence (Pangilinan 2013).

In Montréal, Tétreault and El-Geneidy (2010) use AVL and AFC data for a route (67 Saint-Michel) to evaluate proposals for new limited-stop services along the same corridor. Their work quantifies the change in travel time for customers on both the limited-stop and existing services and finds savings for both groups. They report their findings to Société de Transport de Montréal (STM), the public transport operator sponsoring the work, who then implemented the service. In an ex-post analysis, the researchers evaluate running times after the implementation of the new route and determine that their estimates were acceptably close to the implemented reality. In this case, two types of knowledge were generate from analyzing travel time information: (1) the most effective stopping pattern for the new service, and (2) the accuracy of the model. This validated the model for future use. This work also implies that the operating agency trusted the researchers and their work, since they implemented their suggestions. The work does not describe how this trust was gained, however.

All of this past work has focused on addressing specific issues, where an analyst interprets the information and their knowledge is then codified and communicated in a report, along with recommended actions. This represents a centralized knowledge generation paradigm. This is in contrast to a distributed paradigm where information is presented and viewers create their own knowledge. Under this paradigm, information design and visualization play an important role in aiding viewers in interpreting the information.

Kennedy (2012) explores and evaluates different techniques for visualizing transportation information for a variety of audiences. He concludes that dynamic information visualizations that allow users to interact with the data and change what information is presented provide the best opportunity for creating knowledge among diversified groups of stakeholders.

Many transit agencies have begun distributing real-time information to customers either via the Internet, mobile apps, or signs at stops and stations. This is a distributed knowledge-generation platform that enables customers to combine the real-time information with information about other routes, traffic, and other factors. Based on their prior experience this may allow a rider to know when to leave, how fast to walk, what route to take, or whether to take a taxi. Over time passengers may develop wisdom such as what path to take in certain situations. If customers had access to detailed, quantitative information about the

MBTA that allowed them to see performance as it relates to their trip, this would provide additional information beyond what they gather from their experiences. Such information is not currently available publicly.

The Toyota Production Model, which is often lauded by business scholars, provides an example of a distributed knowledge generation model that influences organizational practices. The Toyota Model empowers those in most direct contact with the manufacturing process to address problems at that level. All parts of the production process are specified to a minute degree. If an employee is not meeting goals, she or he works with a supervisor to discuss a remedy. In some cases, this involves changing the way the employee is approaching the task. In others, it is changing the specification of the process (Spear and Bowen 1999). The point is that Toyota's performance management incorporates a distributed, bottom-up process to generate knowledge. Employees are taking in data about their adherence to standards and generating information about their performance and the circumstances influencing it. Knowledge about how to improve a failing process is generated from those involved in it, rather than requiring an analyst to gather information and find a solution.

Both the centralized and distributed paradigms have had success in generating knowledge and making changes in organizational practice. This research generally follows a centralized model, with MIT researchers performing the data analysis and leading the development of the performance reports and pilot projects. However, its intention is to create performance information that enables a distributed knowledge generation platform, allowing the MBTA to continue to make service improvements after the conclusion of this work. For this reason, it solicited the input of MBTA operations personnel as to what information was meaningful to them and would enable them to better manage service.

2.3 Innovation and Change in the Public Sector Context

Innovation generally aims to change the way things are done. In the DIKW framework, this implies improving knowledge and wisdom because these underlie action. It may also involve generating new information to create an improved understanding of the situation, which allows for innovation. This section reviews literature on innovation in the public sector to understand how information becomes knowledge in this context.

In general, public sector organizations are characterized as bureaucracy. While bureaucracy often has been associated with inefficiency and frustration, Max Weber argues that it developed due to its technical advantages of mechanizing and routinizing the process of administration, just as industrial processes had done to production. He notes that bureaucracy's strict hierarchical form removes ambiguity and enables tasks to be completed more quickly because they are fully prescribed by the superior to the subordinate. Weber also argues that bureaucracy in its purest form eschews nepotism and uses a meritocratic process for advancing within its hierarchy, providing an incentive to perform well. Individuals functioning in a bureaucracy develop specialized knowledge of their tasks, and thus perform them more efficiently over time. While bureaucracy is sometimes misconstrued as a government phenomenon, Weber observes that it is fully aligned with the ideals of capitalism: efficiency, specialization, and competition. It is the organizational structure of most mature corporations and government agencies (Weber 1946).

Many of the characteristics of innovation may conflict with the highly structured and methodical nature of bureaucracy. Innovation is often experimental. It may result in failure as often as success. Robert Behn argues that this creates inherent dilemmas for those attempting to make changes in government agencies. Innovation is not routine. In many cases it involves changing procedure (Behn 1997). This may disrupt the mechanized bureaucratic process. Alan Altshuler writes that 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 also notes that much innovation originates from the lower ranks of an organization that are closer to service provision (Altshuler and Zegans 1997). This may conflict with the hierarchical, top-down nature of bureaucratic organizations.

In a case study of two government agencies, however, Peter Blau shows that employees in government agencies do welcome changes to procedures. He finds this to be true particularly when the changes address existing problems or make their jobs easier. His study also finds that agencies will welcome change that increase their workload if they see it as enabling the agency to better accomplish its core mission (Blau 1963).

Altshuler and Zegans outline several broad strategies that they have found to be common 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 (Altshuler and Zegans 1997, 78).

Altshuler and Zegans' observations also suggest that new information is more likely to produce innovative knowledge if it makes a clear case for change and addresses existing problems. This provides an argument for producing information in close collaboration with its eventual 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.

These characteristics of bureaucracy provide an important framework for researching how information can be disseminated within a government institution to produce knowledge and wisdom. Because bureaucracies are hierarchical and employees have specialized knowledge of their tasks, the same information presented to different people will likely result in different knowledge, and potentially different applications. This suggests that a critical aspect of innovation and performance management in bureaucracy is identifying employees whose knowledge impacts performance. If the intent is to change a process, those with the power to affect that process must obtain new knowledge about it. The ideal candidates will be those whose actions impact what is being measured. Because their actions are represented by the information, this establishes a feedback loop between action and performance quality.

2.4 Purpose and Need for 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 that it has, and what influences its effectiveness.

Past work has also been successful in translating information into knowledge under a centralized paradigm where analysts take data, developing knowledge, and communicating this knowledge in a report or memorandum to the agency. There has been little research into the effectiveness of this strategy. There is also a lack of research into the effectiveness of the current performance reporting regime, which follows a distributed knowledge generation paradigm. Performance reports are made available to managers, who develop their own knowledge about managing the system and strategies to address issues.

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 The MBTA Context and Its Use of Automated Data

The MBTA is one of the largest and oldest transit systems in the country. It has installed ADCS on many of its modes and is currently using the information for real-time service control and some performance reporting. This chapter describes the MBTA context in more detail. It discusses its current automated systems and how the data is used. It also evaluates the influence of these data applications on service delivery.

3.1 The MBTA

The MBTA is the fifth-largest transit system in the U.S. by total ridership, serving 356 million unlinked passenger trips in 2012 (APTA 2012). It operates all major modes of transit, including three rapid transit lines (Red, Orange, and Blue Lines), two light rail lines (Green Line and Mattapan Trolley), two BRT lines (Silver Line Waterfront and Washington Street), 200 bus & trolleybus routes, 12 Commuter Rail lines, and four ferries.

According to the MBTA's most recent service statistics from 2010, the Red Line has the highest average weekday boardings with over 190,000, followed by the Green Line with over 180,000 (MBTA 2010). Table 2 displays average weekday boardings for most MBTA modes (excluding commuter rail and ferry).

Service	Boardings
Red	192,513
Green	181,434
Orange	141,052
Blue	44,233
Silver	29,649
Mattapan	4,586
Bus	357,482
Total	**950,949**

Table 2: MBTA boardings by Service

Figure 3: MBTA System Diagram showing Rapid Transit and Key Bus Routes

The MBTA is headed by a General Manager (GM), who directs the overall policy and strategy of the organization. The Massachusetts Secretary of Transportation sits on the MBTA board and also influences policy. The Chief Operation Officer (COO) is primarily responsible for the day-to-day operation of the system. The Operations Control Center (OCC) is in charge of many departments that currently produce and use data from the MBTA's automated systems. Dispatchers see train and bus positions from the ATO and AVL systems in real-time. Plans and Schedules uses running time data to plan service. Operations Technology maintains these systems. See Figure 4 for an organizational chart of MBTA staff and departments relevant to this work. Additionally, during the course of this

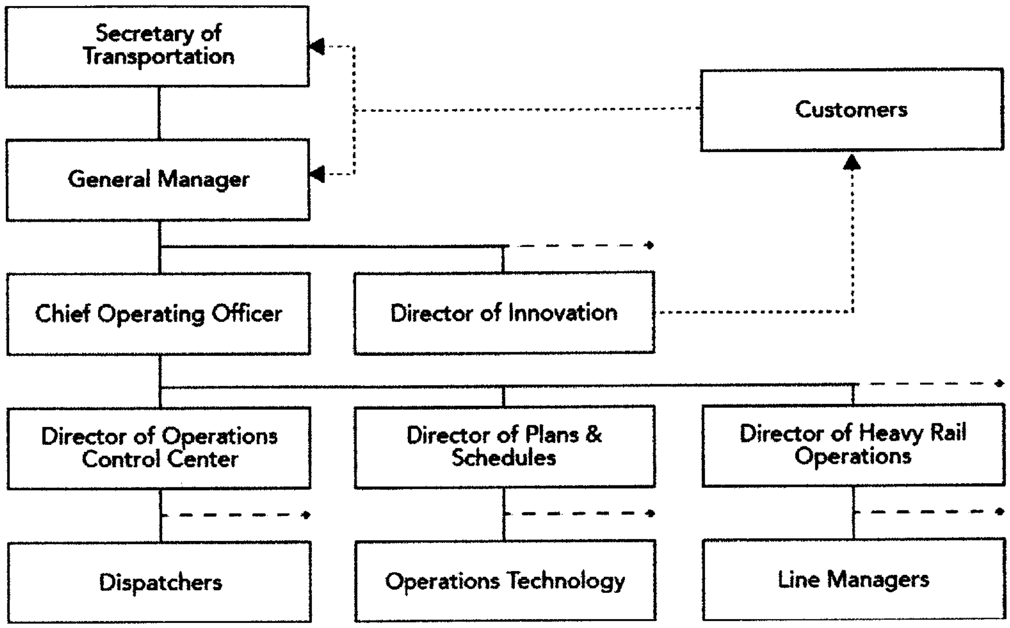


Figure 4: Organization Chart of Relevant MBTA Employees

research the MBTA had a Director of Innovation who reported directly to the GM. He was primarily responsible for creating visible changes that improved the customer experience, particularly through the application of new technologies. His projects were thus reliant upon ADCS data. They included releasing a real-time bus and train arrival feed for mobile applications, displaying real-time arrivals in rail stations, and introducing mobile ticketing on the Commuter Rail.

The MBTA currently has the following ADCS for rail rapid transit, light rail, BRT, and bus.

- Automated Fare Collection (AFC): transaction records for magnetic stripe CharlieTickets and RFID CharlieCards, including time of transaction, rapid transit station or bus route, and fare type, among other pieces of data.
- Automated Vehicle Location (AVL): records of bus position based on GPS and bus odometers, including time of arrivals and departures from key points along the route.
- Automatic Train Operation (ATO): records of heavy rail train positions based on the track circuit from which the train is currently drawing power. Records include time, train direction, destination, and other information.
- Automated Vehicle Identification (AVI): records of light rail (Green Line only) vehicles passing key points along their routes, usually junctions. Records include time, train route, and direction.

While the MBTA's heavy rail, light rail[4], and buses all currently generate data on both vehicle locations and passengers, there are important differences between the modes. For heavy, ATO gives precise train locations and AFC gives precise passenger boarding stations. However, because the fare gate is separate from the train, the time a person enters the station is not the time they board the train. For light rail (Green Line), AVI provides imprecise train locations because AVI points are several stations apart. On the surface branches of the Green Line, there may be as little as one AVI point for the entire surface segment, so location is effectively unknown until the train reaches the end of the route. Because the Green Line runs partly in a subway with gated stations and partly on the surface with open stations, AFC provides two kinds of data. In the subway, the data is similar to heavy rail: precise location but imprecise time. On the surface passengers pay on the vehicle, so the AFC transaction records the precise time they board but only contains the line, not the stop. This is similar to the passenger information available for buses. The AVL system, however, provides precise bus positions. The Silver Line bus rapid transit has the same characteristics as other buses for its vehicle information, but the passenger information characteristics of the Green Line. Table 3 summarizes these differences in the characteristics of ADCS across modes. Transfers between services where passengers board on the vehicle record the subsequent boarding. Transfers within interchange stations are not recorded, but can be inferred, as shown in Barry et al. (2002).

Mode	**Vehicle Position and Time**	**Passenger Entrance Location**	**Passenger Entrance Time**
Heavy Rail (Red, Blue Orange)	Precise	Precise	Approximate
Light Rail (Green)	Imprecise	Precise (subway) Line only (surface)	Approximate (subway) Precise (surface)
Light Rail (Mattapan)	None	Line only	Precise
BRT (Silver)	Precise	Precise (subway) Line only (surface)	Approximate (subway) Precise (surface)
Bus	Precise	Line only	Precise

Table 3: ADCS Characteristics for Different Modes of the MBTA

[4] With the exception of the Mattapan High Speed Line, which does not currently have vehicle location data

3.2 The MBTA's Use of Automated Data

In the past decade the MBTA has begun utilizing the data collected from ADCS. The current applications of automated data can be categorized along two dimensions: scope (internal versus external) and timeframe (real-time versus historical). This implies four broad categories of applications, which are depicted in Figure 5 along with the MBTA's current data uses.

The upper-left quadrant, internal real-time uses, is why many agencies install ADCS. They include displaying vehicle locations, estimating vehicle arrivals at terminals and if they will make their next trips, and displaying schedule or headway adherence. Such information enables more precise operations control because dispatchers have more accurate and detailed information. The lower-left quadrant, external real-time uses, has followed internal real-time uses at the MBTA. ADCS may not have been designed for customer information, but once the data exists it can be disseminated to inform riders about current service. The same data that underlies real-time information for dispatchers can be adapted to estimate bus and train

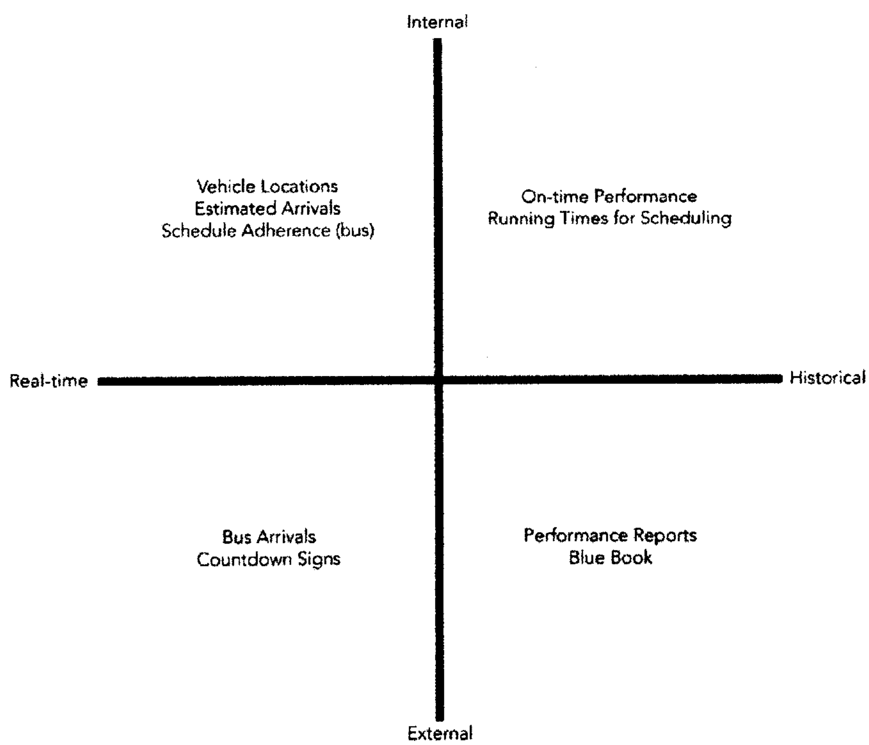


Figure 5: Categorization of ADCS Uses at the MBTA

arrivals, which the MBTA releases to customers via the Internet, mobile apps, and station signs.

Real-time ADCS applications represent new data applications that were driven by ADCS, since there was previously no real-time information. Once the data is logged by the ADCS, this also enables applications of historical data. These have mostly been replacing manually collected data in existing functions. Internally (upper right quadrant), this includes calculating OTP and running times based on automatic vehicle location rather than manual checks. Externally (lower right quadrant), these statistics are summarized for various periods of time and published in monthly and annual performance reports.

This chapter discusses what information the MBTA currently extracts from its automated data as well as the limitations of its current practices. These findings form the basis of the work discussed in the following chapters to create more valuable information from the agency's data.

3.3 Internal Real-time Applications for Operations

At the MBTA, both heavy rail and bus have real-time information for dispatching, though they differ in the information they display. Both the ATO and AVL systems provide real-time vehicle location data to dispatchers, allowing them to see where vehicles are. The bus Computer Aided Dispatching (CAD) system combines this with schedule and other bus location data to give dispatchers information about schedule adherence and headways (Figure 6). The technology suite is generally referred to CAD-AVL. The ATO system gives heavy rail dispatchers train position in a graphic display (Figure 7), from which they can interpret headways and speed, but it does not relate this information directly. The system also provides estimated arrivals at the terminals along with the next schedule departure so dispatchers can see if a train will be late. These visual displays allow both bus and rail dispatchers to respond to delays or disruptions in near real-time, adjust vehicle and crew schedules accordingly, and generally better manage daily operations.

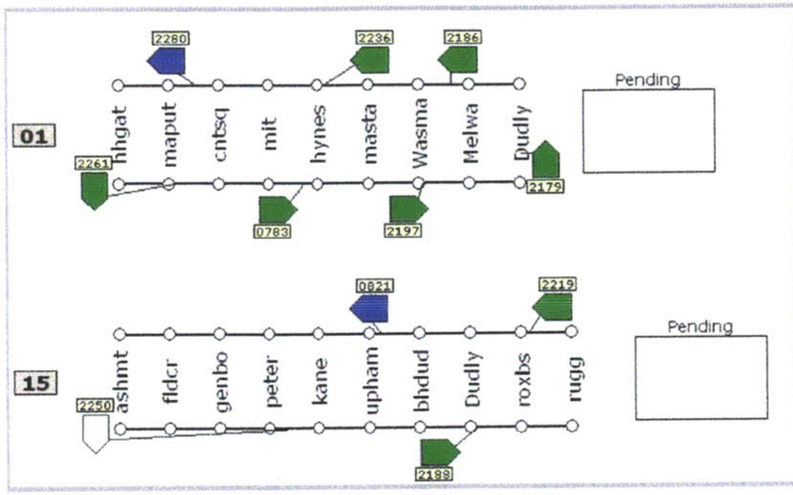


Figure 6: Real-time display of AVL information for bus routes 1 and 15.
Routes are abstracted to a single line for each direction, with timepoint stops marked. The line connecting the pentagonal bus icons to the route indicates where the bus should be based on its schedule. In this example, blue buses are early and green are within the on-time range.

Figure 7: Real-time display of ATO information for the Red Line.
The two parallel lines represent the tracks, and a red section indicates that a train is currently on that section of track.

The ATO system includes Automated Train Dispatching (ATD), which rings a bell at terminal stations at scheduled departure times, prompting drivers to depart. This effectively automates everything but the driving of the Red, Blue, and Orange lines, so dispatchers' use of real-time location data focuses on maintaining good service. This includes ensuring that operators actually leave when the bell rings, adjusting scheduled departures when trains are going to miss their next trip, reassigning vehicles and drivers in the case of disabled trains, and holding trains at intermediate stations to adjust headways or for other reasons. While there is no departure bell for buses, all vehicles are equipped with screens that are linked to the CAD-AVL system. These screens display their next departure time so that drivers know when to leave, as well as their schedule adherence en route. Bus dispatch uses real-time information in much the same way as their rail counterparts: adjusting departures, reassigning vehicles and drivers to avoid missing trips, and expressing or holding buses to break up bunches.

Both the ATO and CAD-AVL systems have their shortcomings. In both cases, there is no immediate feedback about the effects of dispatching action (such as a running calculation of OTP or another metric) other than the visual representation of vehicle locations on the dispatchers' monitors. Additionally, the rail ATO system does not display headways or arrival predictions for stations other than the terminal, which is problematic because headways are fundamental to service quality. On the rail side, there are no indicators or alarms that alert dispatchers to problems; dispatchers must observe them. For buses, the marker on the map changes colors if the bus is early or late. Early and late can be calculated either based on the schedule or on the headway to the previous and next buses, whichever dispatchers select.

3.4 Internal Historical Applications

The ATO and AVL systems have historical reporting tools that produce on-time performance reports for any desired period, based on either terminal departure and arrival times or headway adherence. The AVL system has a more sophisticated reporting tool, Smart Bus Mart, which allows a user to view performance information in different ways. The most commonly used report is the OTP report for the MBTA's 15 most heavily used routes (dubbed Key Routes – see Figure 9). Operations staff can access several other pre-made reports, including on-time performance for any given period, schedule adherence down to the individual trip, and headway adherence. If they desire additional information, they can specify additional reports through a Web interface (Figure 8). The metric of interest, level of aggregation, time period, route(s) are customizable. Performance can be aggregated by driver, garage, route, and other elements.

Current OTP is based on the MBTA's service standards. These standards are developed and revised through a public process that takes customer input into account (MBTA 2010). The current service standard for rail transit is a train departing the terminal within 150% of the scheduled headway. The standard for bus is two-pronged, differentiated based on the frequency of service. Walk-up service – where customers are assumed to show up to a stop or station without looking at a schedule – is defined as service with a frequency of ten minutes or less. Scheduled service is anything with a headway greater than ten minutes. The on-time standard for walk-up service is a vehicle arriving on that route within 150% of the scheduled headway. For scheduled service, on-time is defined as departing a timepoint between one minute early and five minutes late. Bus OTP is measured at multiple points along the route: the origin, several midpoints, and the destination. Overall OTP is calculated

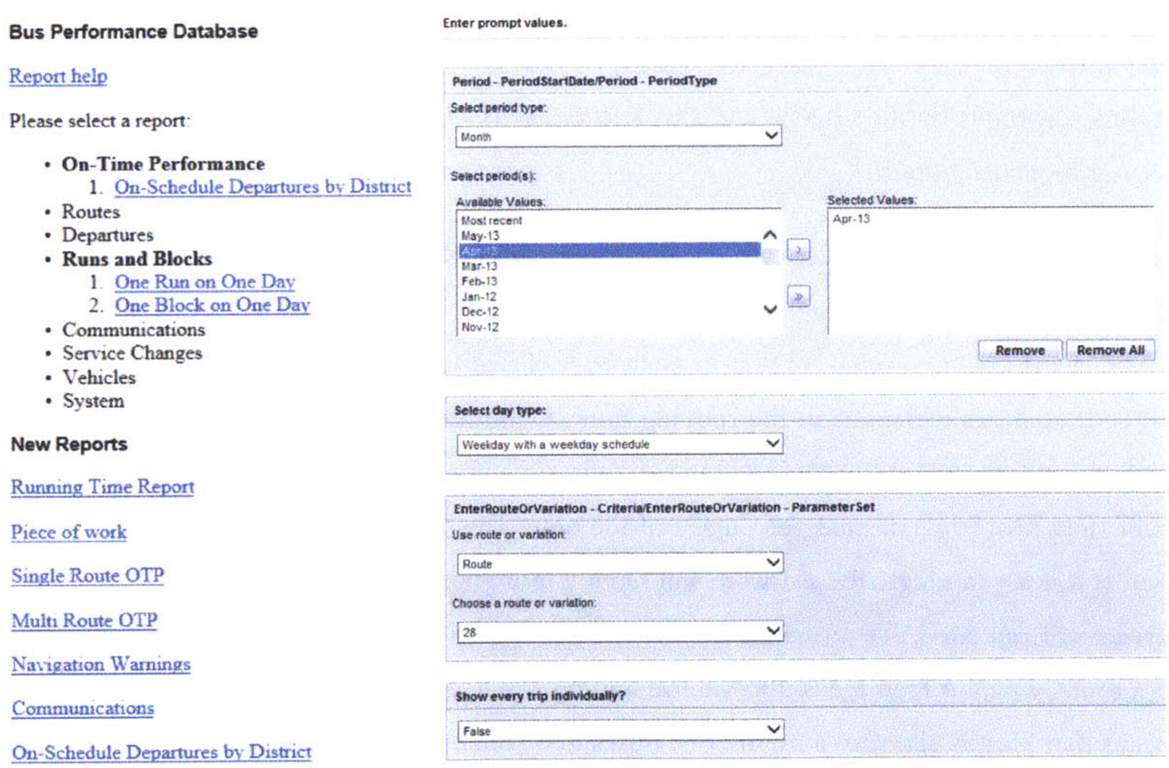


Figure 8: Excerpts from the MBTA's Smart Bus Mart Reporting Tool

as a percentage: total timepoints the bus or route served on-time divided by its total number of timepoints. Schedules for some routes employ different standards throughout the day. The #1 bus, for example, is scheduled to arrive at 7-8 minute intervals in the peaks, so it would be evaluated on the headway standard for those periods. In the midday, it is scheduled at 13-minute intervals, so it is evaluated on the schedule standard. Table 4 summarizes these standards.

Service	On-Time Standard	Measurement Point
Rail	Headway ≤ 1.5 times scheduled	Departure point (1 per trip)
Walk-up Bus	Headway ≤ 1.5 times scheduled	Key timepoints (5-10 per trip)
Scheduled Bus	1 minute early to 5 minutes late	Key timepoints (5-10 per trip)

Table 4: Summary of MBTA Service Standards

In addition to performing its own analyses, the MBTA also provides data to local universities for their research. Usually such research involves analyzing the data beyond what is possible in the reporting system, and the findings are presented to the MBTA in a memo or report, along with recommendations. These reports generate additional information for the MBTA, and the researchers attempt to transfer the knowledge they gain by providing recommendations. In some cases, like Malikova's (2012) suggestions to adjust the headways of three-car trains on the Green Line, this knowledge is applied to improve service. In other cases, it remains unused.

The MBTA has also begun analyzing running times from the AVL system using Hastus ATP (a module of their scheduling software, Hastus). Service planning has begun rescheduling routes based on the results. ATP is an analysis tool that uses AVL data as an input (GIRO 2011). It analyzes variations in the running times within each period of the day, as specified by the user. Its output is a running time for each route in each period that will allow buses to make their next trip a desired percentage of the time. This percentage must be defined by the user. If it is set too high, the software will require more buses to run the service; too low and service will run late. ATP provides a more accurate input that Hastus uses to allocate buses to a route, which it then feeds into vehicle and crew schedules. More accurate running times means that Hastus allocates a number of buses that should enable a route to run on time the desired percent of the time. Underestimating running times means a bus may not make its next trip, while overestimation results in less service than is possible.

This represents a shift in the internal use of ADCS beyond generating information for information's sake. OTP is simply information, and there is not currently a systematic process for applying that information to improve service. Poor performance is acknowledged and dispatchers are sometimes questioned as to why service was poor, but there is no institutional process of determining how to address recurring issues. Analyzing running times with the goal of improving bus scheduling is the creation of information (running times) with the intention of generating knowledge (how to change the schedule).

With the recent exception of Hastus ATP, the MBTA's regular use of historical data has been limited to OTP reports: a single percentage for each route every day. These are individual reports on one dimension of service quality, and different views of OTP such as by route or by timepoint are separated. This limits the amount of knowledge a viewer can obtain. The burden is on the viewer to relate different performance information and identify causes and trends. Showing only one dimension of service at a time, such as OTP or dropped trips, does not provide a comprehensive view of service. Without relating different dimensions of service quality, it is difficult to understand what is causing variations in OTP (management, equipment, passengers, weather, etc.).

Though the MBTA's internal reporting systems are flexible enough to allow staff to gather information on other dimensions of service, they must be willing to take the time to aggregate and analyze the information. Multiple reports can provide information like average speeds, headways, incidents, labor shortages, and other factors that influence of OTP, but seeking out these reports is not part of the daily routine for operators and managers.

The Key Routes On-Time Performance report for buses, shown in Figure 9, is emailed to dispatchers and OCC managers every morning, The report provides summary information about overall OTP on each route and how they compare. However, it provides no detail as to which buses were early or late, how off schedule they were, whether they were judged based on the headway or schedule standard, or how many passengers were affected. Thus there is little knowledge that a viewer can develop from this report. There is nothing to help explain why and how this performance occurred.

The aggregate on-time performance number for a route does not allow operations personnel to know what is the result of factors beyond their control like traffic, and what is due to management. This information requires further investigation using the Smart Bus Mart reporting system, which is time consuming and may still require further analytical work to capture multiple dimensions of performance. Shared segments where passengers can take multiple routes are not judged based on a joint headway across routes, but on the headways or schedule adherence of each route individually. This is particularly problematic when the individual routes are not frequent enough to be judged by the headway standard, but the frequency of the combined service is. This is the case for the #116 and #117, which operate at 20-30-minute frequencies for much of the day. They serve the same termini and share much of their route, so most customers can take either service. Dispatchers can manage service to maintain a combined headway between the two routes, but this may result in many off-schedule departures and low OTP. From a customer perspective, such poorly rated

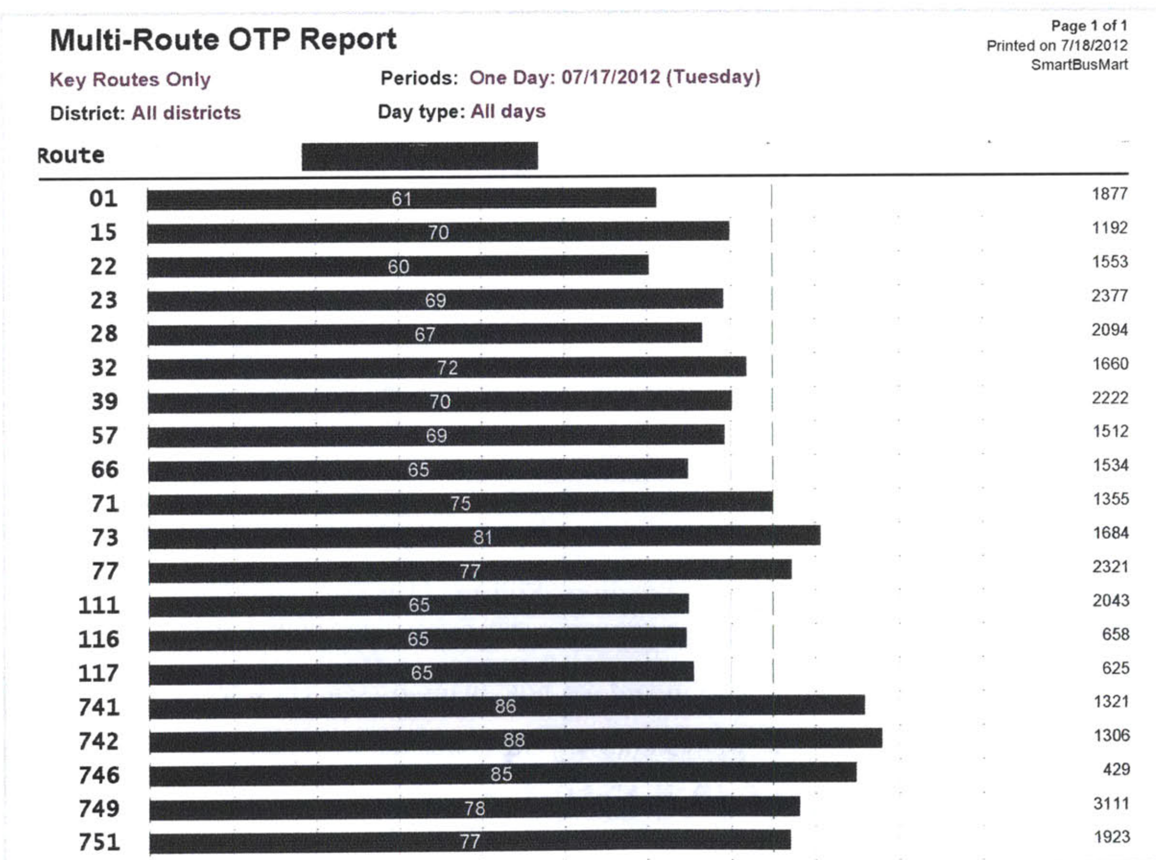


Figure 9: Key Routes Daily On-Time Performance Report

service is good service.

Additional information is available from other reports that must currently be sought out. A line report (available in Appendix A, page 121) provides detail on:

- The breakdown of not-on-time trips by reason (early, late, headway)
- OTP by hour and direction
- OTP at each point along the route
- OTP for each run.

While this report provides details that address many of the shortcomings of the single OTP number, it is four pages long for each route. Additionally, it still examines each dimension separately. OTP by time is separate from OTP by location, so the viewer can understand that there are problems during specific hours (2:00 PM) or at specific places (Hynes Station), but not a specific place and time (such as Central Square at 5:00 PM). This limits the amount of knowledge that can be gleaned from the information.

Rail OTP reporting is similar, though more tabular and less visual, as shown in Figure 10. The report summarizes OTP by period and direction (the Red Line has two branches, for a total of four directions). As described earlier, OTP for rail is judged solely on the headway departing the terminal. The 93% overall OTP for the Red Line on this day means that 93% of trips left the terminal within 150% of their scheduled headway. This report only measures service on the two branches individually. There is no measure of combined service on the trunk portion, though 67% of trips are only on the trunk.[5] This means that the scheduled headway that is the basis of 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 are on-time, even though this separation is more than double the expected joint headway of 4.5 minutes. Short headways (bunches) still count as on-time.

---

[5] Based on the O-D calculations described in Section 4.1.3

Moreover, this report is the first of 13 pages that present OTP for each trip run that day. This is a great deal of information that results in little knowledge. Unlike the bus report, which contains intermediate detail about performance over the course of the day or route, the rail report contains only highly aggregate and highly disaggregate information. This limits the ability to gain knowledge of 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 rail lines are usually above 90% on-time, even when service may be perceived as lacking.

Deval L. Patrick, Governor
Timothy P. Murray, Lt. Governor
Richard A. Davey, MassDOT Secretary & CEO
Jonathan R. Davis, Acting General Manager
and Rail & Transit Administrator

massDOT
Massachusetts Department of Transportation
Rail & Transit Division

Headway On-Time Performance Daily Detail Report

Date: Tue, Oct 02, 2012 **Line: Red Line**

	Trips			Headway			Headway
	Ran	Drop	Add	Short	Normal	Long	Performance
Alewife to Braintree:	109	2	0	2	99	8	93%
AM:	6	0	0	0	6	0	100%
AM Peak:	20	1	0	0	20	0	100%
Mid Day:	30	0	0	1	28	1	97%
PM Peak:	22	0	0	0	17	5	77%
PM:	31	1	0	1	28	2	94%
Braintree to Alewife:	110	1	0	0	105	5	95%
AM:	10	0	0	0	10	0	100%
AM Peak:	21	0	0	0	20	1	95%
Mid Day:	31	0	0	0	31	0	100%
PM Peak:	20	1	0	0	18	2	90%
PM:	28	0	0	0	26	2	93%
Alewife to Ashmont:	108	2	0	1	97	10	91%
AM:	6	0	0	0	6	0	100%
AM Peak:	21	1	0	0	18	3	86%
Mid Day:	29	1	0	1	26	2	93%
PM Peak:	20	0	0	0	17	3	85%
PM:	32	0	0	0	30	2	94%
Ashmont to Alewife:	107	3	0	1	100	6	94%
AM:	9	0	0	0	9	0	100%
AM Peak:	19	2	0	0	17	2	89%
Mid Day:	30	0	0	1	29	0	100%
PM Peak:	19	1	0	0	16	3	84%
PM:	30	0	0	0	29	1	97%

TOTALS

	Ran	Drop	Add	Short	Normal	Long	Performance
Red Line:	***434***	***8***	***0***	***4***	***401***	***29***	***93%***
AM:	31	0	0	0	31	0	100%
AM Peak:	81	4	0	0	75	6	93%
Mid Day:	120	1	0	3	114	3	98%
PM Peak:	81	2	0	0	68	13	84%
PM:	121	1	0	1	113	7	94%

Figure 10: Front Page of the Daily On-Time Performance Report for Rail

The current institutional applications that transform data from ADCS into information are limited in the amount of knowledge they generate. This is a result both of what information is produced (e.g. rail OTP does not reflect service quality) and how it is presented (e.g. bus OTP is not aggregated in ways that can inform management decisions). While the work of Shireman (2011) and Malikova (2012) has successfully generated knowledge from ADCS data, and in the case of Malikova, even led to a change in Green Line headways, this knowledge generation has been based on a single dedicated analyst addressing a specific problem. Their work is not based on the MBTA's standard reports and took weeks or months of analysis. A public agency with limited resources and overburdened staff needs its performance reports to be able to generate similarly useful knowledge that it can apply every day. Broadening the audience to include the public creates an additional set of stakeholders to assess service and suggest improvements.

3.5 External Historical Applications

The MBTA has also developed public-facing information from ADCS data that riders can incorporate into their own understanding of the system. The MBTA currently publishes performance reports on a monthly basis (Figure 11), disaggregated to the individual subway lines, which are complemented by an annual report on service statistics.

The information in the monthly scorecard is aggregated to the line level. The detail pages for each line show the historical performance of each metric over the past 12 months. This is enough information to understand general month-to-month trends and to draw correlations among them. For example, seeing a drop in vehicle maintenance and system maintenance along with poor OTP suggests that maintenance levels influence performance. However, the

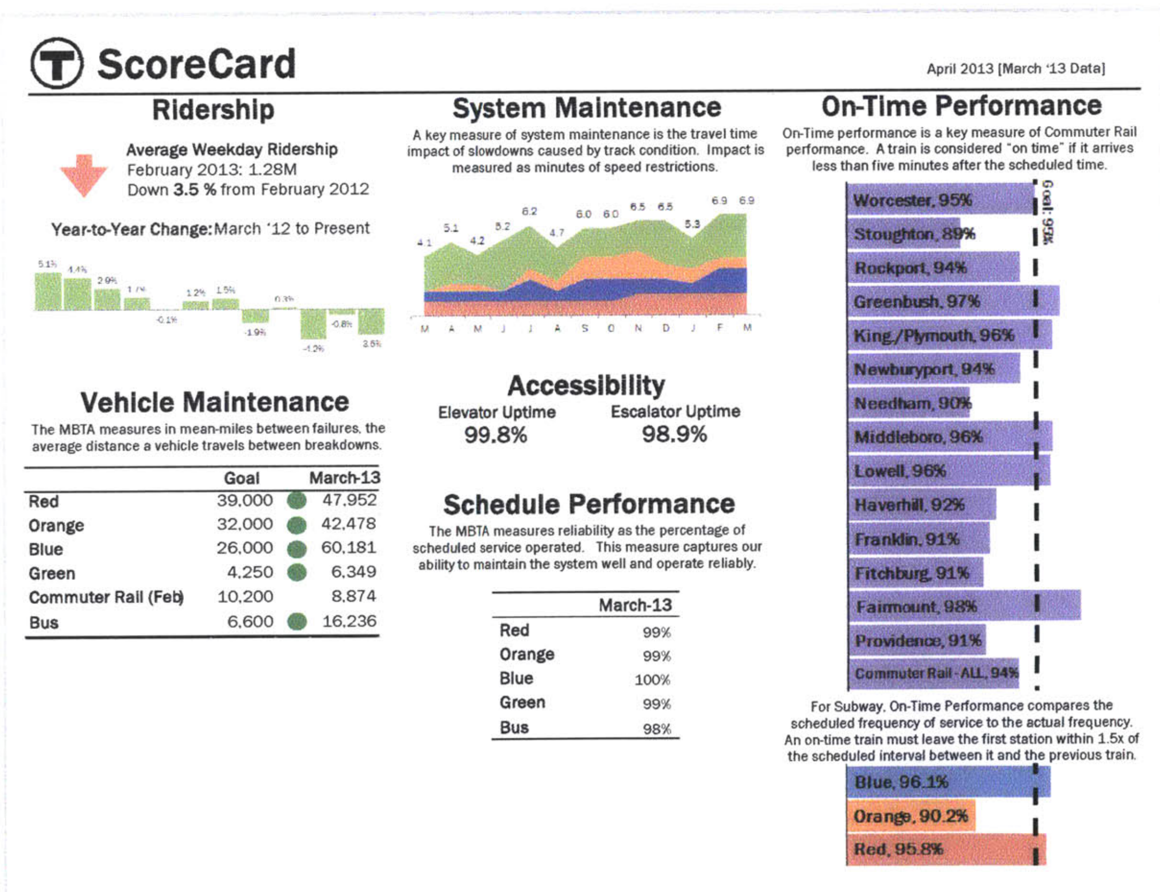


Figure 11: MBTA Monthly Scorecard

historical information is static and non-interactive, limiting the public's ability to view service performance as it relates to their use of the system. The understanding that can emerge from this information is limited to "Is the T performing better or worse than usual?" and "How does my line compare to the rest of the system?" This falls short of knowledge, however, because the understanding of how and why performance is changing is speculative. The interpretation of this information could be that the MBTA is poorly managed, or that it has insufficient resources, that it needs new equipment, among others. This ambiguity reduces the value of this information to both the MBTA and to the public.

The annual report (known as the Bluebook) provides extensive amounts of information on maintenance, ridership, equipment availability, and other characteristics. The information is aggregated along multiple dimensions, among them month, route, and service area. There is a wealth of information to sift through, which can be used by politicians, reporters, advocacy groups, researchers, or riders to understand the state of service at the MBTA. However, it is only produced once per year[6] and is in a relatively inaccessible format (a 100 page document). Generating knowledge from information in the Bluebook thus requires searching through it and relating pieces of information to one another. The knowledge is thus limited to people who are willing to devote time to this research and by how they communicate their findings.

3.6 External Real-time Applications

The MBTA also makes real-time bus and train information available through an open data feed, from which developers have created Internet and smartphone applications providing customers with train and bus arrival times (Figure 12). In 2012, the MBTA also introduced in-station train arrival predictions at most of its heavy rail stations, providing all customers with an estimate of their wait without mobile Internet access (Figure 13). Silver Line BRT stops on Washington Street and at Logan Airport also feature arrival predictions.

[6] Or is supposed to be, though the latest version is from 2010

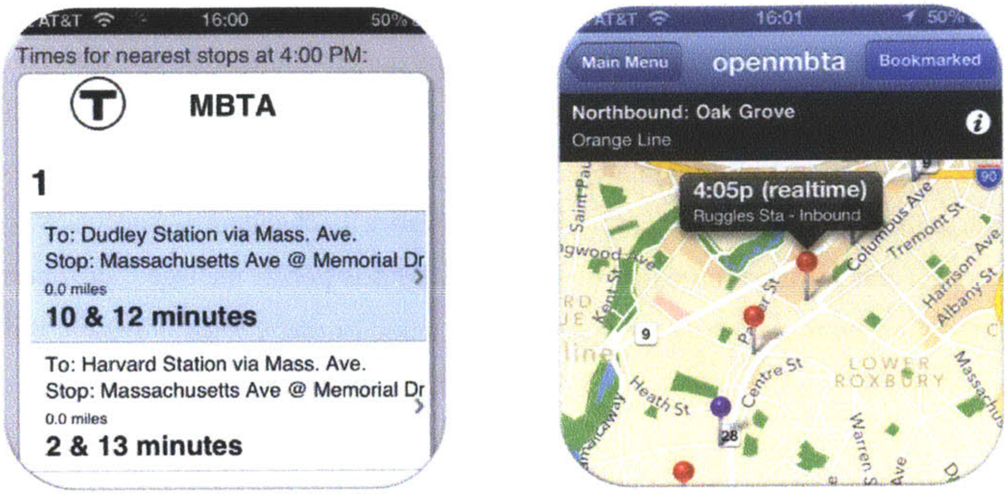


Figure 12: Mobile apps based on the MBTA's NextBus Information and developer feed.
Left: NextBus mobile interface showing the #1 bus. Right: OpenMBTA showing the Orange Line.

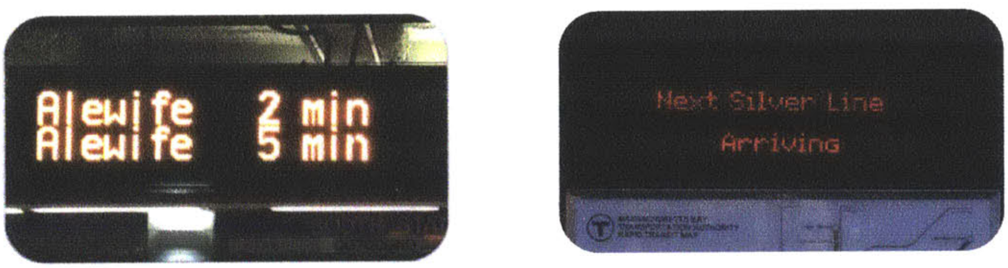


Figure 13: Real-time train and bus arrival displays
Left: a countdown sign on the Red Line. Right: arrival sign for the Washington St. Silver Line.

This has been a successful application of ADCS because it gives the users (customers) the information that they need in order to make a decision. Providing real-time information allows customers to know how long their wait will be and may inform their decision about route or mode choice. A customer can see how many minutes remain until the train or bus arrives, and can choose to take a taxi or walk or bike if it is too long. The real-time information creates an immediate basis for a decision such as "I need to take a cab to make it to the airport on time," or "The 1 bus isn't coming soon so I should take the Red Line." Additionally, seeing the countdown signs every trip establishes knowledge of normal headways on the line, which may ultimately change riders' expectations. This in turn may also create external pressure on management as riders begin to get a quantitative view of service.

The countdown signs and real-time bus arrival apps provide users with a snapshot of service that is relevant to them. While over time they accumulate multiple snapshots, which represent multiple pieces of information, they still do not have a full picture of the system. This means they can see problems with individual trips – a long headway, a slow trip – but do not have information on what is going on in the rest of the system that may be causing these problems. The links between performance on different lines are not always evident. For example, when a rail line has a failure that requires the MBTA to provide shuttles, it pulls the buses from the most frequent routes because these can absorb the loss with less of an impact on headways. Customers, however, do not have this piece of information and thus may never relate a breakdown on the Orange Line with a long wait for the #28 bus.

3.7 Effectiveness of Current Data Usage at the MBTA

Real-time information derived from ADCS provides both dispatchers and passengers with a depiction of current performance that allows them to evaluate the need for and effect of action. It enables dispatchers to see in real-time how unscheduled variations in service impact the system, as well as the effects of actions they take to adjust for these. In theory, this should accumulate over time to form wisdom that allows them to predict the effect of a problem and take action to mitigate it. Similarly, passengers with access to real-time information can make more informed decisions about mode and route choice.

The agency's public information, however, is not comprehensive enough to generate knowledge about causes of problems. The public can either get (1) a granular snapshot of current system performance from the real-time arrival information or (2) an aggregate summary of performance by line over the course of a month or year.

These two extremes do not allow external users like advocacy groups, the press, and elected officials to analyze performance in detail to identify trends, problems, and potential solutions. The historical information is also static and non-interactive, limiting the public's ability to view service performance as it relates to their use of the system. It is thus difficult for the public to provide anything other than anecdotal evidence for complaints about service quality. It is possible to archive the real-time data feed and use this for analysis, which was undertaken by a group of MIT researchers in 2011 and 2012 (Gerstle 2012). Their research successfully analyzes the data to produce useful information about running times, but they note that the data was imprecise due to the relatively infrequent nature of the feed (only refreshed every 60 seconds).

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, thus information is rarely translated into knowledge or action. The existing performance reports do not provide enough detail to show the impacts of dispatchers' reactions to real-time information. On the rail system, headway adjustments mid-route to avoid bunching are not reflected in the current OTP numbers because they only measure terminal departures. The same action would be reflected in the OTP statistic for buses, but the aggregated reporting format makes it difficult to draw direct connections between actions and OTP. Additionally, the current OTP numbers are route-specific, which does not capture joint service for a corridor where customers can take multiple services. The Red Line is the primary example of this, where current OTP evaluates Ashmont and Braintree trains independently. This also occurs on several bus routes such as the #116 and #117, #71 and #73, and #70 and #70A. 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. Table 5 in Chapter 4 explains this limitation in more detail.

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.

The following chapter will discuss the approach this research took to incorporating these changes into a new performance report, and how the process influenced the effectiveness of the performance information. Chapter 6 will propose how this information can be made accessible to other parts of the organization and to the public in the future.

4 Enhancing the Utility of Performance Information

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. This chapter describes the process of developing new performance reports for the MBTA's heavy rail system, as well as the resulting changes to the metrics and reports themselves. It concludes that to be useful, performance information must be both easily comprehensible and trusted by service managers. Reorienting metrics around customers and using graphical techniques to display information may improve comprehensibility. The effect of new communications techniques can be tested via a collaborative process, which also helps to build trust in the reports and a willingness to distribute them beyond the operations team.

4.1 Approach and Objectives

The work was originally conceived to provide more frequent and detailed information to customers about service quality that complements the performance "snapshots" produced by the countdown signs. It started with reconceptualizing metrics, and engaged the MBTA's OCC 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 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 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 specifically. Though the intention of releasing information more broadly within the MBTA and publicly has been retained, this objective was not achieved in the course of this research.

4.1.1 Approach

While not initially planned, engaging the OCC initiated a collaborative and iterative process to incorporate feedback from operations personnel on what types of metrics would impact the way they managed service. This process has been critical to the project's acceptance by the OCC and its ability to propose and implement service changes (described in Chapter 5). It included multiple visits to the OCC to meet with dispatchers and managers and observe their work. Operations staff have domain knowledge of operational problems and the merits of different performance metrics. This has been combined with MIT's technical expertise in manipulating data and ability to review existing practice and literature to produce new performance reports.

A central tenet of this approach is that performance management is not a technical problem to be solved analytically, but a managerial problem to be addressed socially. This applies more to transit systems whose trains are driven and dispatched by humans than to automated rail systems, where the only humans interacting with operations are passengers (most of the time).

4.1.2 Objectives

Based on the concerns listed in Chapter 3 about the existing metrics, this research has 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, like a long headway or a dropped trip;
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;
6. Based on existing automatically collected data so that calculation can be automated and done in real-time or for the past day;

This work hypothesizes that these qualities enable performance information to impact service. This chapter discusses multiple iterations of new performance reports for the MBTA and the rationale behind their evolution as they strove to meet these objectives. The reactions and feedback from OCC managers and staff were the primary means of determining how well the objectives four and five were being met. Their feedback provides important lessons on how one of the intended audiences for the reports understands them.

The performance information and reports that result from this research are for a single agency, based on the data and needs of the MBTA. Other agencies may face different problems with their existing information and service or have management structures that necessitate different solutions. The physical outputs of this research thus may not be applicable to other agencies or even other lines within the MBTA, though the process and principles may still be informative.

4.1.3 Technical methods

One of the initial drivers of this work was the introduction of real-time arrival signs on the subway. This research uses the same data that underlies the prediction software. These are records from the ATO system of a train occupying a specific circuit. The data are archived in a Microsoft SQL Server database, where additional tables are created to calculate headways, running times, and other statistics from them. Passenger information comes from archived AFC transactions, which are stored in a separate SQL database. Because the system records entries only, a process similar to that of Barry, et al. (2002) is used to infer destinations for these transactions. This is part of ongoing MIT research for the MBTA. This results in an origin-destination (O-D) matrix for the rail system for each day of data. An average daily passenger volume is then calculated for each O-D pair. The passenger O-D data used in developing these reports is an average for days in April 2012. In much of this work, the total number of passengers for each O-D is converted into a rate (passengers per second) for each period. This assumes constant arrivals over the period, which is consistent with the theory of random arrivals used in most transit planning (Wilson and Attanucci 2011).

4.2 Initial Performance Report

The ultimate goal of performance measurement is to enable performance improvement. This requires that managers are able to interpret the information and relate it to their knowledge about service. Choosing what information to produce from ACDS data began with hypothesizing what knowledge about service would be most useful and applicable to MBTA operations personnel and to riders. Discussions with OCC staff revealed complaints that the rail reports were meaningless because they were always above 90% on-time, even after significant disruptions. They also criticized the existing reports for not considering joint service, both on the Red Line and on the #116 and #117 bus routes that share a majority of their stops. These concerns suggested that operations staff desired information that more accurately represented service as passengers see it. Knowing that the reports judged these services separately but most customers used them interchangeably invalidated the performance information in their view.

Conversations also revealed that OCC staff did not seek out performance information, but did respond to the Key Bus Routes report that was emailed to them every day. A frequent critique of this report, however, was that it gave no context as to what was driving the OTP numbers. This suggests that OCC staff desired information about specific problems with service that they could influence without having to search for it.

Two elements of the customer experience on public transit can be easily measured by AVL and AFC: waiting time and in-vehicle travel time. Other aspects of the customer experience such as crowding, comfort, and convenience are also important, but less readily measured with these two data sources. To measure travel and wait time, this work began with two basic units of analysis representing these parts of the experience:

1. Headway as a measure of wait time, as expected passenger wait time is half of the headway (Wilson and Attanucci 2011)

2. Station-to-station travel time

The intent of these reports is to measure instances of poor service to provide a customer perspective, instead of the system-oriented OTP. This stems from the assumption that customers expect a certain level of performance, so poor service receives more attention than good service. Put differently, passengers do not give the MBTA credit when it is running well to the same extent that they blame it when service is poor.

Based on this, we have chosen to measure the big gaps for rail services but have excluded counting bunches. The rationale is that for rail services, short headways are not a concern except when they create big gaps behind them; consistently short headways are simply good service. The definition used for big gaps is based on the MBTA's existing headway metric, 1.5 times the scheduled headway, but this is limited to three minutes beyond the scheduled headway in order to account for long headways that occur during off-peak hours. Under the current standards, for example, with a 13 minute scheduled headway, 19.5 minutes is still considered acceptable. From a passenger perspective 13 minutes is already a long headway, so even a few minutes longer is poor service. With a three-minute cap, any headway over 16 minutes is unacceptable for a 13-minute frequency. Additionally, on the Red Line, which has two branches, the current OTP metric measures trains on each branch individually but does not measure the combined service on the shared portion, even though 67% of weekday travel is only on the shared portion.[7] The result is that during the peak periods, where branch headways are scheduled at nine minutes, as long as a train leaves a terminal every 13.5 minutes or less service is on-time, even if service is bunched. Under the proposed metrics, big gaps are counted separately for trunk and branch services. In the example shown in Table 5 (below), where the headways should be 4.5 minutes on the trunk and 9 minutes on the branches (peak hour service levels for the Red Line), the threshold for big gaps is 6.75 and 12 minutes, respectively.

Time	Branch	Trunk (branch) Headway	Proposed Criteria: Trunk Service	Proposed Criteria: Branch Service	Existing OTP Criteria
8:26	Braintree		Good	Good	On-time
8:30	Ashmont	4	Good	Good	On-time
8:37	Braintree	7 (11)	Big Gap	Good	On-time
8:43	Ashmont	6 (13)	Good	Big Gap	On-time
8:50	Braintree	7 (13)	Big Gap	Big Gap	On-time

Table 5: Example of Red Line Headway Performance under Proposed and Existing Criteria

The headway metric is able to be calculated at any station or intermediate point. An analysis of headway variations along the line shows that headways generally remain consistent from terminal to terminal. The median difference in headway for a train's start and mid points is zero or near zero, and 80% of trains' headways vary less than two minutes between different

[7] Based on the O-D calculations described in Section 4.1.3

points along a route. Measuring headways at one point in the system can thus provide an accurate depiction of service along the entire line. This provides justification for measuring headways at the terminals (the current practice), though it provides equal justification for measuring at a midpoint. Managers know that there are more passengers at midpoint stations, therefore measurements at these points are likely to be more salient.

4.2.1 Presentation

As seen in Figure 14, the first draft of the performance report revolves around the absolute number of big gaps. This is similar to the OTP metric as it measures only one thing (in fact, the opposite of OTP), but in more detail. In addition to the top-level total big gaps, the report counts big gaps for service on each of the Red Line's branches. These break down further into subtotals for each period and direction, for both trunk and branch service. The totals and subtotals are color-coded red, orange, and green to good, mediocre, and poor service.[8] The chart in the upper right showing total big gaps over the past five days is meant to give an indication of relative performance and an incentive to perform better than previous days. Finally, graphs representing all headways over the day provide a disaggregate view of service, with big gaps marked by red triangles. This is included because initial analyses showed headway varying from one train to the next. 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.

While this report centers around big gaps, it also includes a count of long dwell times at each station. This metric attempts to capture two things: (1) slow trips due to overcrowding, which increases dwell times because more passengers enter and leave the train, and (2) dispatchers holding for headway adjustments, which substitutes one passenger inconvenience for another. This metric focuses on a different aspect of the customer experience that is not captured in headways or OTP.

[8] The thresholds for these color codes were arbitrary at the time this report was produced, since it was a proof-of-concept. The idea was later abandoned, so no formal methodology was developed.

4.2.2 Reaction and Input of Operations Personnel

This draft of the report was presented to the director of the OCC, who was surprised by the headway graphs. They clearly show irregularity in service, particularly northbound where the two branches merge. He also noted that measuring only big gaps may be inadequate, as dispatchers could hold or express trains to maintain headway, but these also negatively impact the customer experience. This is a fact that is well known in the transportation community, which the counts of long dwell times attempted to capture. In retrospect, the dwell time metric does not provide useful information for dispatchers because it captures two different problems that may require different actions. However, these could result from heavy passenger loads or from dispatcher action.

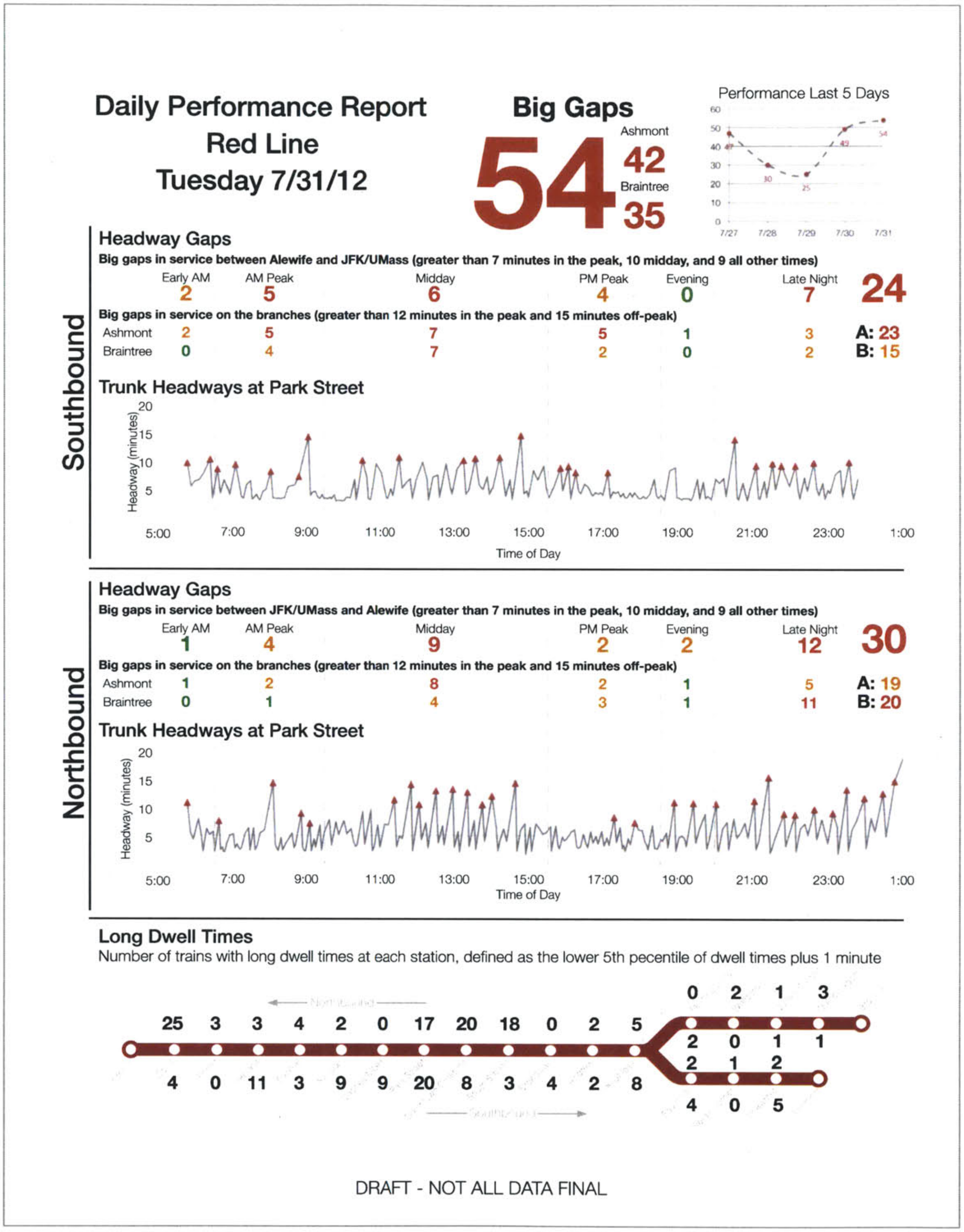


Figure 14: First Draft of Performance Report

Objective	Effectiveness
Capture speed	• Long dwell metric attempts, but does not capture trains slow in between stations • Conflates holding and crowding
Capture frequency	• Big Gaps measure instances of infrequency, but no overall measure • Headway charts provides some visualization of frequency, but do not summarize
Capture reliability	• Big Gaps capture unreliability • Headway charts effectively visualize reliability by highlighting variation in headways
Sensitive to service variations that are perceptible to passengers	• Big Gaps and dwells based on a threshold that represents perceptibly bad headway • Do not distinguish between bad and very bad
Easily understood by OCC staff	• Big Gap numbers are straightforward and understood • Headway charts are powerful visualization
Detailed enough to identify problems and actions	• Headway charts provide detail to see problems, and imply need to manage headways • Long dwell counts do not, since they may be out of dispatch's control

Table 6: Summary of First Iteration of Performance Report

4.3 Modifications and Second Draft

Based on feedback from the OCC director, the report was modified to include additional metrics that complement the big gaps measure better than the counts of long dwells. Adding the number of slow trains attempts to address the concern that dispatchers could hold trains to maintain headway, similar to the previous dwell time metric. Dispatchers are strongly discouraged from expressing trains, so in the MBTA case counting express trains does not add much information. An analysis of the distribution of train running times, shown in Figure 15, revealed that the distributions were fairly tight. Figure 15 shows the median running times for each major segment of the Red Line by period. The error bars that extend to the 10th and 90th percentile values for the running time distribution. The length of the error bar represents the variability in running times. Variability is significantly higher on the trunk than the branches. However, the largest change between the median and the 90th

percentile was 23% (Alewife-JFK, Evening). A threshold for slow trains was set at 15% longer than the median running time for the period. This first iteration of the slow trains metric was based on end-to-end run times (from leaving the first stop to arriving at the terminus), which includes dwell times at all intermediate stations. The slow trains metric replaces the long dwell time metric from the first draft, as it captures both long dwell and running times. It still does not differentiate between slowness due to holding and slowness due to crowding, however. From a passenger perspective, a slow train is inconvenient regardless of its cause, but this may reduce the usefulness of the metric for management.

Since big gaps describe the tails of the headway distribution, the second draft of the report incorporates a measure of service regularity to capture the variation within the full distribution. The objective is to measure the degree of variation in headways, as the graphs on the initial report show headway deviations that do not create big gaps. Variations in headway create uneven train loads, extending dwell time and potentially causing delays. The MBTA's real-time signs displaying the time until the next train add to the importance of consistent service. Customers can now see the time until their train and the train behind it,

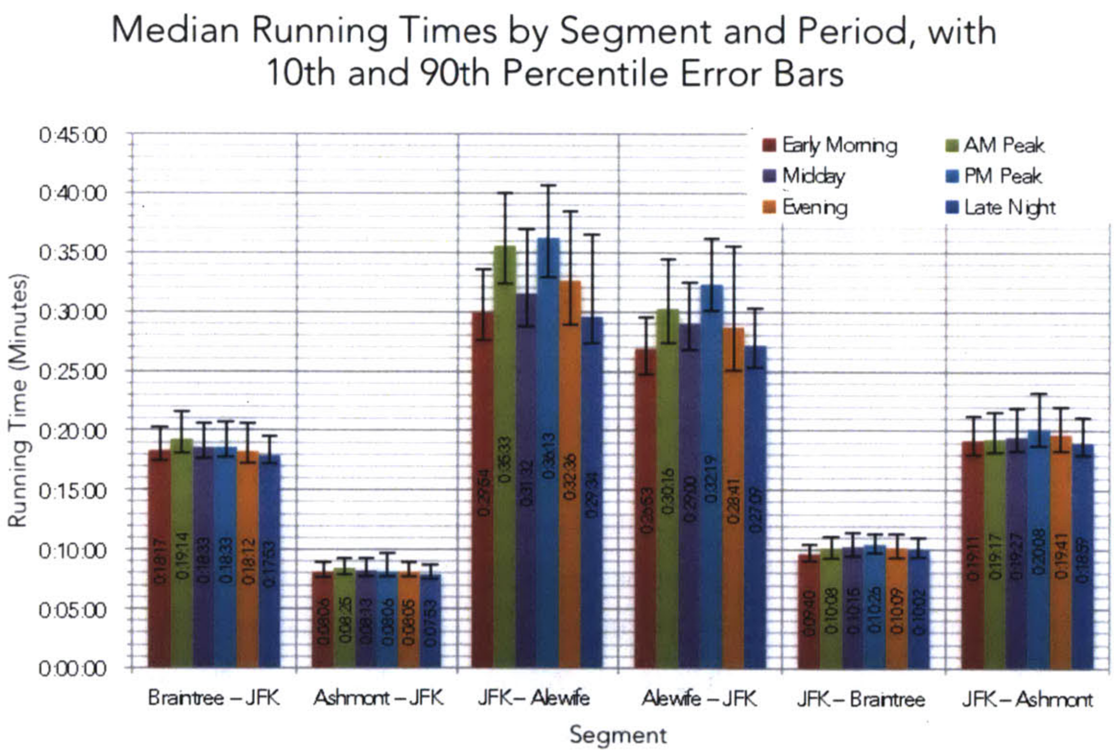


Figure 15: Median, 10th, and 90th Percentile Running Times for the Red Line by Period and Segment

which can cause frustration if they have a long wait and see that the second train is just behind the first.

The Level of Service (LOS) metric from TCRP Report 88 measures the standard deviation of differences from the scheduled headway and maps it to an LOS grade, which mimics the highway LOS grade (TCRP 2003). The calculation, as shown in Equation 1, first calculates the difference between the actual and scheduled headway for all trains in a period (during which the scheduled headway is constant). It then takes the standard deviation of this distribution, and divides this by the scheduled headway. This is effectively a normalized standard deviation, relating variations in the headway to its scheduled value.

$$\frac{SD\{h_1 - h_S, h_2 - h_S, \ldots, h_i - h_S\}}{h_S}$$

Equation 1: Transit LOS

Where:

h_i = headway for train *i*

h_S = scheduled headway during a period of consistent headways

SD{…} denotes the standard deviation of the set of headway deviations

The result is a number usually between 0 and 1, with 0 indicating no deviation from the scheduled headway. TCRP Report 88 maps this metric to letter grades as shown in Table 7.

Grade	Range	Points
A	0.00 - 0.21	4
B	0.21 - 0.30	3
C	0.30 - 0.39	2
D	0.39 - 0.52	1
E	0.52 - 0.74	1/3
F	> 0.74	0

Table 7: LOS Grades

In this research, an aggregate grade for multiple periods with different scheduled headways is calculated using a weighted grade point average (GPA). Each grade is assigned a point value, as with academic grades (shown in Table 7), which are weighted by the duration of the period as a fraction of the service day (i.e. the AM Peak is 3 of the 20 service hours so its weight is .15). The total weighted GPA is the sum of the weighted GPAs for each period, which is then translated back into a grade (i.e. a 2.5 is a C, a 3.5 a B). The advantage of this is that it provides a single grade for the entire day. The disadvantage is that it does not weight by passenger levels, in fact the peak periods with the most passengers are shorter and thus receive less weight. This was intentional, based on a judgment that regularity is more important in the off-peak periods with longer scheduled headways. An alternative would be to weight by passenger volume, to create a combined weight that takes multiple factors into account, or to aggregate peak and off-peak service separately.

4.3.1 Presentation

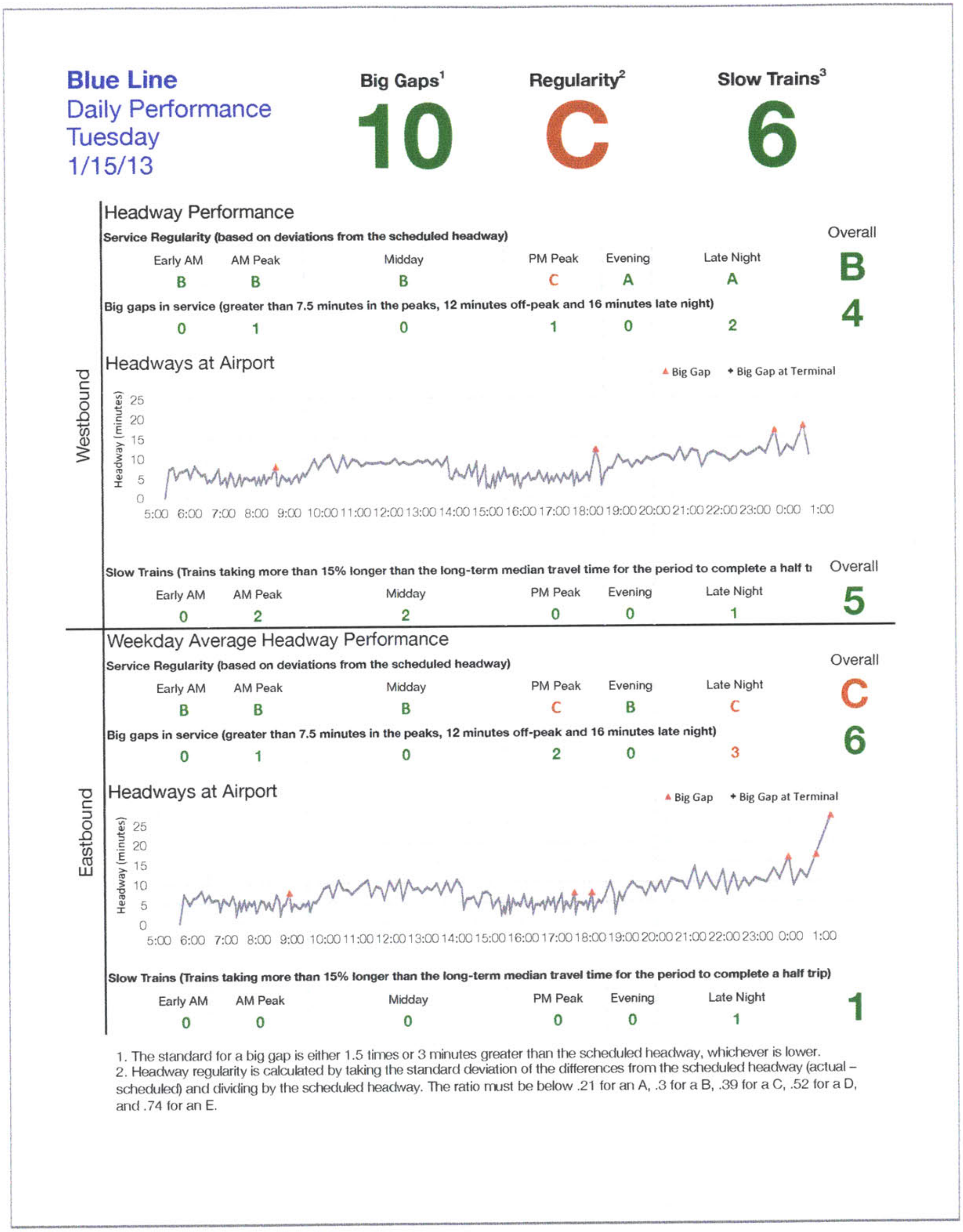
As shown in Figure 16, in the second iteration of the performance report the big gaps measurement is augmented by the LOS and slow trains totals. The report presents information about multiple dimensions of service together. The intention is to emphasize these as equally important and allow correlations to be drawn between them. For example, holding trains at stations to adjust for headways would likely result in a low number of big gaps, but a higher number of slow trains. This research theorized that seeing such values for a day when dispatchers recall holding a lot of trains would underscore both the positive and negative consequences of holding for headway adjustments. The graph charting the performance of the past five days' performance was removed for technical production reasons. At this stage in its development, the reports had moved from proof-of-concept to a preliminary level of production. The algorithms were creating performance metrics on-demand, but not storing them, so the data structure to produce historic comparisons did not exist. The concept of comparing a day to historical performance is reintroduced in subsequent drafts.

On this particularly day on the Blue Line, service was generally consistent, with very few big headway gaps or slow trains. Values for each of the measurements are provided by period and direction below. Providing information about which periods are performing poorly allows dispatchers to focus their management efforts. The headway graph provides detail to substantiate the big gaps and LOS measurements. The legend at the bottom explains the methodology behind each of the calculations. This allows people viewing the report to understand what the numbers are based on and thus how to influence them.

4.3.2 Reaction and Input of Operations Personnel

The incorporation of the additional measurements was received positively, though the letter grade was viewed as somewhat harsh. One operations manager commented that it seemed impossible for service to get above a C, even when everything else looked good. This was discouraging to managers and dispatchers. Their reactions began a conversation about whether to compare to a theoretical ideal (i.e. zero big gaps, no variation in headways, no slow trains) or to an observed achievable level of good service. The next draft of the report attempts to address this dilemma.

These drafts of the reports were used to evaluate the effects of the pilot programs (discussed in Chapter 5). Managers received these reports frequently for several weeks. In meetings with them, we observed that managers paid attention to the top-level numbers and the headway graphs. The breakdowns by period and direction were less important. Another operations manager commented that the report gave him an easy way to investigate customer or employee complaints of long headways because he could simply look at the headway plot. This implies that the intermediate levels of aggregation to the line and direction were not adding useful information for the managers. The combination of summary numbers for the entire day and the detailed graphs showing every train provided enough information to understand how service was that day and what was driving the numbers.



Blue Line
Daily Performance
Tuesday
1/15/13

Big Gaps[1]	Regularity[2]	Slow Trains[3]
10	C	6

Westbound

Headway Performance

Service Regularity (based on deviations from the scheduled headway)

	Early AM	AM Peak	Midday	PM Peak	Evening	Late Night	Overall
Service Regularity	B	B	B	C	A	A	B
Big gaps in service (greater than 7.5 minutes in the peaks, 12 minutes off-peak and 16 minutes late night)	0	1	0	1	0	2	4

Headways at Airport

Slow Trains (Trains taking more than 15% longer than the long-term median travel time for the period to complete a half t

Early AM	AM Peak	Midday	PM Peak	Evening	Late Night	Overall
0	2	2	0	0	1	5

Eastbound

Weekday Average Headway Performance

Service Regularity (based on deviations from the scheduled headway)

	Early AM	AM Peak	Midday	PM Peak	Evening	Late Night	Overall
Service Regularity	B	B	B	C	B	C	C
Big gaps in service (greater than 7.5 minutes in the peaks, 12 minutes off-peak and 16 minutes late night)	0	1	0	2	0	3	6

Headways at Airport

Slow Trains (Trains taking more than 15% longer than the long-term median travel time for the period to complete a half trip)

Early AM	AM Peak	Midday	PM Peak	Evening	Late Night	Overall
0	0	0	0	0	1	1

1. The standard for a big gap is either 1.5 times or 3 minutes greater than the scheduled headway, whichever is lower.
2. Headway regularity is calculated by taking the standard deviation of the differences from the scheduled headway (actual – scheduled) and dividing by the scheduled headway. The ratio must be below .21 for an A, .3 for a B, .39 for a C, .52 for a D, and .74 for an E.

Figure 16: Second Draft of Performance Report

Objective	Effectiveness
Capture speed	• Slow train metric captures trains delayed by the end of their trip • Still conflates holding and crowding, and also bunching
Capture frequency	• Big Gaps measure instances of infrequency, but no overall metric • Headway chart provides some visualization of frequency, but does not summarize
Capture reliability	• Big Gaps capture unreliability • LOS grades provide quantitative measure of overall regularity • Headway charts effectively visualize reliability by highlighting variation in headways
Sensitive to service variations that are perceptible to passengers	• Big Gaps based on a threshold that represents perceptibly bad headway • LOS grades represent noticeable change in regularity from one to next
Easily understood by OCC staff	• Big Gap numbers are straightforward and understood • Headway charts are powerful visualization that managers can actually use • LOS grades are opaque in their calculation; improvements within one grade level are not shown
Detailed enough to identify problems and actions	• Headway charts provide detail to see problems, and imply need to manage headways • LOS grades and big gap counts are too aggregate to identify specific issues • Slow trains are counted but detail is not shown

Table 8: Summary of Second Iteration of Performance Report

4.4 Third Draft: Refocusing the Reports on Passenger Impacts

While conducting pilot projects, passenger volumes were incorporated into estimates of travel time to emphasize how many people experienced service improvements. These numbers resonated with the operations managers, providing motivation to translate the performance metric into units of customers or customer hours rather than trains.

4.4.1 Passenger-Weighted Metrics

O-D data estimated from AFC data, as described by Gordon (2012) produces detailed passenger demand information. These measures include customer boardings at a station and total riders between any two stations during any given period. This enables the performance of each train to be weighted by the expected number of customers experiencing that service. For example, a nine-minute headway at rush hour affects more people than in the late night, though both are big gaps. In the case of branched services, arrival rates for trunk-bound and branch-bound customers can be calculated individually to account for the fact that not all passengers can take every train. This embeds an additional piece of information in the metrics: the impact of performance on passengers. Such information makes explicit the relationship between the performance of trains and the experience of passengers, where it was previously implied. Public transit is a service with the objective of moving people. Measuring aspects of service that matter to customers enables operations personnel to directly understand the impacts of their actions on achieving this objective.

These metrics employ historical passenger demand rather than real-time demand. The MBTA's AFC data on passengers is not processed every day, and is thus not available on the same basis as ATO information on train locations.

Passengers Affected by Headway Variation and Big Gaps

Counting the number of passengers that wait more than the published headway, a big gap, or a very big gap (twice the headway), provides an estimate of how many people likely perceived service as poor because they waited longer than they expected. The number of people waiting longer than the published headway can be calculated by multiplying the passenger arrival rate by the difference between the actual headway and the published headway, as shown in Equation 2. 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 (varying the value of h_t). This gives the subset of those passengers with "extra" wait time who waited the longest. These calculations can be done separately for passengers waiting for trunk and branch services, using branch-specific arrival rates and headways.

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

Equation 2: 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.)

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 what they should expect based on the published schedule. This can also be expressed as a percentage – the proportion of riders who wait too long – which is a salient figure for operations personnel.

Expected Total Passenger Wait Time

This metric is intended to capture the effect of service variability on passenger wait times. Calculations of these metrics assume a constant passenger arrival rate. This rate is used to calculate the number of passengers waiting and the total wait time for each train, assuming all customers board the first train. Because each passenger that arrives waits a different amount of time, longer headways have more passengers who have been waiting for a larger total amount of time. Assuming passengers arrive at a regular rate (i.e. a random arrival process), the average wait time is half the headway, and total wait is the total passengers multiplied by the average wait. This formula is presented in Equation 3:

$$Average\ Wait = \frac{h}{2}$$

$$Total\ Passengers = \lambda * h$$

$$Total\ Wait\ Time = \lambda * h * \left(\frac{h}{2}\right) = \frac{\lambda h^2}{2}$$

Equation 3: Total Passenger Wait Time for a Single Train

Where:

λ = arrival rate of passengers

h = headway

As the final expression shows, total wait time grows geometrically rather than linearly, and thus wait time will be longer with uneven headways than even headways.

With arrival rates for each station in each direction and for each branch the total wait for each train at each station can be calculated. These are then summed to calculate total wait. Equation 4 outlines the calculation of total wait time for non-branched service:

$$\sum_i \sum_o \frac{\lambda_p^o (h_i^o)^2}{2}$$

Equation 4: Total Passenger Wait Time for All Stations

Where:

λ_p^o = passenger arrival rate at origin station *o* for period *p*

h_i^o = headway for train *i* at station *o*

For branched services, the total wait time is then the sum of total wait time for each type of passenger.

$$\sum_i \sum_o \frac{\lambda_p^{oT} (h_i^o)^2}{2} + \sum_B \sum_{i^B} \sum_{o^B} \frac{\lambda_p^{oB} (h_{iB}^o)^2}{2}$$

Equation 5: Total Passenger Wait Time for Branched Service

Where:

λ_p^{oT} = arrival rate at origin station *o* of trunk-bound customers in period *p*

λ_p^{oB} = arrival rate at station *o* of customers bound for branch B in period *p*

h_i^o = headway for train *i* at station *o* (since last train for any branch)

h_{iB}^o = branch headway for train *i* at station *o* serving branch B (since last train for branch B)

Wilson and Attanucci (2011) develop an equivalent formulation of the average wait time.

$$E(w) = \frac{E(h)}{2}[1 + \left(cov(h)\right)^2]$$

Equation 6: Wilson and Attanucci's Formulation of Average Wait Time

Where:

$E(w)$ = expected (average) wait time

$E(h)$ = expected value of the headway distribution

$cov(h)$ = coefficient of variation of the headway distribution

Multiplying this by number of passengers provides an alternate method of estimating the total passenger wait time. From this formulation, it is clear that the total passenger wait time captures variation in the headway distribution, as it explicitly includes its COV.

Total Passenger Travel Time

Like total passenger wait time, total passenger travel time can be calculated based on O-D matrices inferred from AFC data. The running time, including dwell time, for each train between each possible O-D pair along its route can be calculated from the ATO train location data. Multiplying this by the estimated passenger demand for each O-D pair served by a train results in the total number of passenger-hours of travel time for that train. Passenger demand per train can be calculated by multiplying the headway by the O-D-specific arrival rate for the period. This is preferable to assuming even headways and assigning an average passenger load to each train because more people experience the performance of a train arriving after a long headway. Summing the number of passenger hours for each train results in an aggregate total passenger travel time for the period. The overall calculation is summarized in Equation 7.

$$\sum_{t}\sum_{o}\sum_{d}(RT_{od}^{i})(\lambda_{od}^{p})(h_{o}^{i})$$

Equation 7: Total Passenger Travel Time

Where:

RT_{od}^{i} = running time for train *i* between stations *o* and *d*

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*

This equation assumes that passengers that arrive while the train is dwelling in the station board the next train, because the time after the first train's arrival in the station is part of the following train's headway. It also assumes 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.

Effective Headway

The effective headway is defined in this research as the average headway weighted by the number of passengers experiencing each headway. This accounts for the fact that more passengers arrive during a long headway than a short headway, so the average headway experienced by a customer is higher than the average headway of the trains. Under the assumption of random arrivals passengers wait on average half the headway, thus it is calculated as twice the average wait time. The average wait time can be derived from Equation 3 by summing total passenger wait time for a period and dividing by the total number of passengers. Total passengers can be calculated as the arrival rate multiplied by the period length, or the sum of all headways (since headways include dwell time).

$$H_E = 2 * \frac{\sum_i \left(\frac{\lambda h_i{}^2}{2} \right)}{\lambda \sum_i h_i} = \frac{\sum_i \left(h_i{}^2 \right)}{\sum_i h_i}$$

Equation 8: Effective Headway

Where:

H_E = Effective Headway

λ = Passenger arrival rate

h_i = Headway for train i

As seen in Equation 8, the passenger arrival rate cancels out in the effective headway calculation, resulting in the total minutes of waiting divided by the period length. This makes calculating the effective headway possible even without passenger O-D information over a period with a constant arrival rate.

4.4.2 Presentation of Passenger-Weighted Metrics

The structure of the first two iterations of performance reports was generally retained for the third performance report draft (Figure 17), but with significant modifications to incorporate lessons learned.

Top-level numbers that summarize overall performance on all directions and branches for the whole day are given visual prominence. Passengers with long waits are expressed as a percentage of total passengers, while passenger travel and wait time are expressed as a change from the norm (defined by the long-term median). Expressing total passengers who wait too long as both an absolute number and as a percentage is an easily understandable measurement because it is simply a count of people and a proportion of passengers. Total passenger travel and wait time may not be as useful as absolute statistics for management because passenger-hours is a two-dimensional unit and it is not immediately obvious what the total wait time should be. Comparing them to a normal day helps to make them more understandable and useful as management tools. This measures service relative to a level that operators know they can achieve and exceed, providing an incentive to always do better.

This iteration of the reports also introduces a measure ranking 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 to days in the last half of 2012)[9]. The light gray represents values above the median, while dark gray is below median. This additional information helps put the performance numbers in context. This graphic may make the relative change in travel and wait time (described above) redundant. These reports are still evolving and exploring alternative metrics such as the number of customers delayed by a specified number of minutes.

The break down of the top-level measurements by period and direction is left out of this report, as operations managers did not use it. The period-specific statistics were a numeric summary of the information provided on the headway charts. They did not add any new

[9] 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.

information because the graphs show the number of big gaps and the variation in headways. These are replaced by an effective headway calculation for the period which quantifies the graph in a simpler way. Instead of two quantifications for each period, there is one number that that summarizes the variation in the chart in an intuitive unit (minutes). The headway graph is similar to that presented in the previous report, with the addition of the effective headway and markers for branch-specific big gaps (i.e. a big gap in Ashmont service). The graph serves as an explanation for the top-level passenger wait-time metric.

The slow trains metric has been converted into a series of charts that display running times for each major segment of a line, rather than just the end-to-end time. This change came about during the Alewife pilot (see Chapter 5) where an analysis of running times by station segment showed that specific stations and segments accounted for most of the variability in running times. The same standard for slow trains is applied (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.

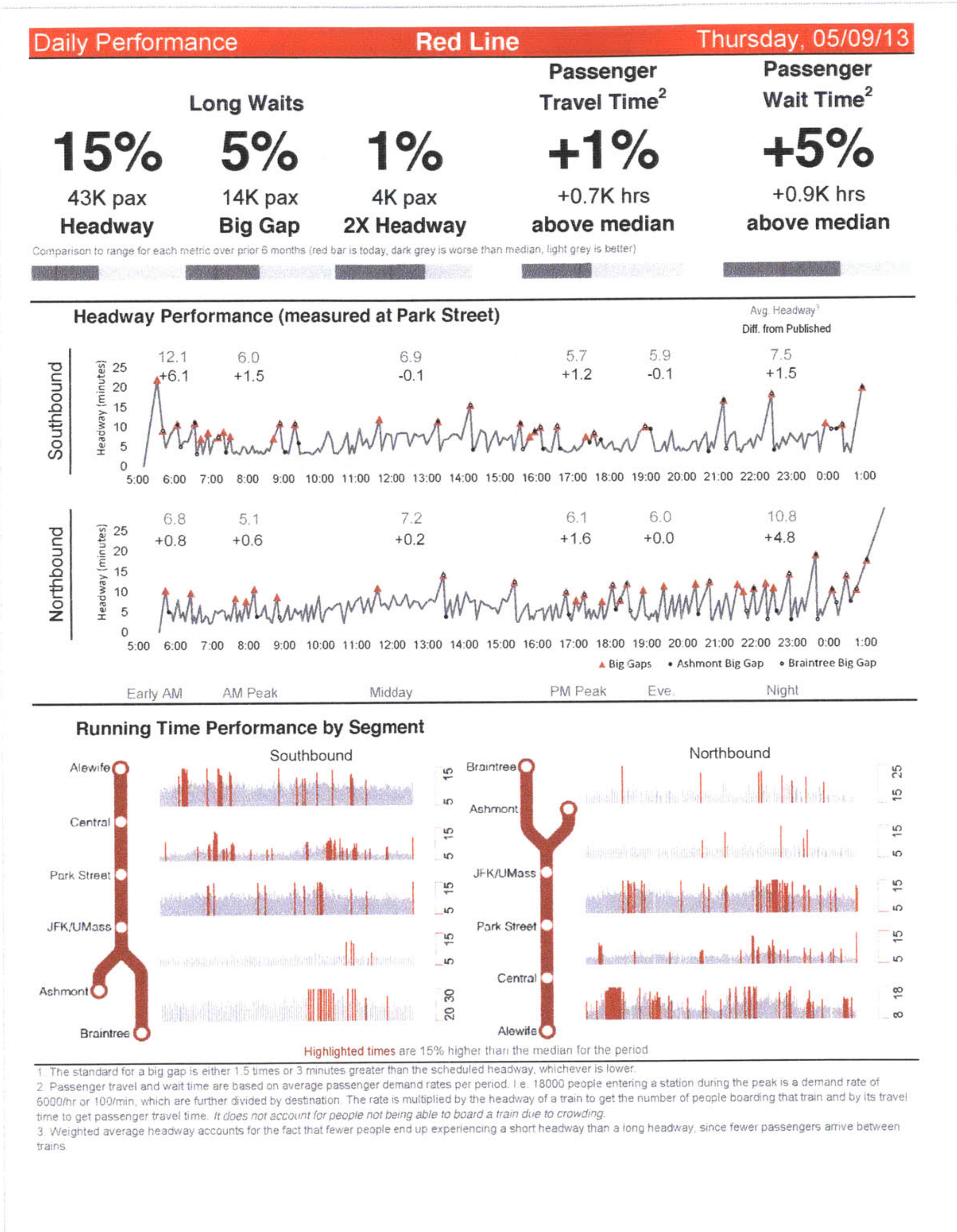


Figure 17: Third Draft of Performance Report Incorporating Passenger-Weighted Metrics

Objective	Effectiveness
Capture speed	• Passenger travel time metric summarizes overall speed • Slow train graphic shows detail of speed by segment for each trip • Still conflates holding, crowding, and bunching
Capture frequency	• Passengers experiencing Big Gaps captures instances of infrequency • Change in total passenger wait time represents change in overall frequency • Effective headways represent average frequency experienced by passengers
Capture reliability	• Passengers experiencing gaps capture experienced unreliability • Relative measures of travel time and wait time represent change from other days, which captures consistency • Headway and run time graphs effectively visualize reliability by making highlighting variation in headways
Sensitive to service variations that are perceptible to passengers	• Multiple thresholds for long passenger waits distinguish between bad and very bad • Metrics capture minute variations in service, but normalizing and limiting significant digits limits variation in numbers to large changes
Easily understood by OCC staff	• Big Gap percentages, slow trains, and headway charts are easily understood • Relative change in passenger travel time and relative position are less intuitive
Detailed enough to identify problems and actions	• Headway charts and slow train graphics provide detail to see problems • Slow trains show both when and where problem occurs

Table 9: Summary of Third Iteration of Performance Report

4.5 Lessons Learned about Performance Reporting

The iterative process of developing these reports revealed several important lessons about performance measurement and performance reporting.

Because service is multi-dimensional a single number summary is often inadequate. The information about which dimension is driving performance is important to developing knowledge. Metrics should reflect aspects of service that customers care about. For frequent services this is regular headways and travel time. While dispatchers talked about getting trains back "on-time," they often modified the schedules to re-establish the headways between trains, reflecting their understanding that headway is more important than schedule.

Dispatchers and managers often commented that a bad headway at 11:00 PM was less detrimental to service than a bad headway during the rush hour because it affected fewer passengers. Weighting by passenger demand qualifies service issues by the number of passengers affected. While dispatchers recognize that their objective is to provide high quality service at all times, incorporating passenger volumes emphasizes that the point of transit service is to move people, not just vehicles.

Finally, the process of developing new performance reports shows that design and presentation of performance metrics is just as important as the metrics themselves. The iterative process and pilot programs created a feedback loop that informed the evolution of the reports. Circulating draft reports helps to determine if the information is enabling useful knowledge.

The OCC expressed interest in the report answering two primary questions: 1) How good or bad was service yesterday? and 2) What caused the numbers to change? The top-level summary numbers in the latest revision of the report mimic the single OTP number, but reflect multiple dimensions of service (wait versus travel time) and the degree of passenger impacts. Breaking down the top-level numbers into smaller levels of aggregation does not add as much value as providing detail on every train through visual techniques. Graphs of headways and running times are an effective way to communicate details down to the individual train without overwhelming the viewer with numbers. Graphics can communicate which trains were driving the performance numbers by emphasizing the information that matters for service. The line on the headway graph emphasizes the change from one headway to the next rather than the headways themselves, since the objective is regularity. The only points that are highlighted are the big gaps, since these negatively impact performance. The same is true of the slow train graphs. A cluster of red bars indicates a bigger problem than a single slow train.

These design choices are intended to reflect how customers perceive service. The presentation implies that as long as headways are below a certain threshold, passengers care more about regularity and predictability than they do about the actual headway. Similarly, for running time, only the high running times are colored, since these cause trips to take longer than customers expect.

Performance reporting is an important component in a data-driven performance management strategy. However, reporting alone does not produce performance improvements. Combining the development of new performance reports with operational pilot projects allows the reports to be tested and refined. Additionally, it helps identify other opportunities for change. The performance improvement process and its synergies with the performance reporting process are discussed in the next chapter.

5 Applying Information to Support Change and Innovation

The previous stages of this project have focused on harnessing automated data to create performance reports. While performance information is a necessary part of performance management, its mere existence does not impact service. This portion of the research leverages performance information to make service improvements at the MBTA. The objective is to demonstrate the value of new performance information and build support with those in control of service. The original intention was to establish internal comfort with the reports so that they could be made public. This has evolved to include improving service before releasing numbers publicly. The intended internal audience began as operations control managers and dispatchers, but has expanded to service planning and other management staff. The eventual audience is intended to include the general public. As discussed in Chapter 2, knowledge results from interpreting information and understanding how to influence the present situation. Enabling performance information to impact service thus means getting service controllers to engage with the information, to become comfortable with it, and to trust it. This chapter discusses two pilot projects that modified service based on analyses of automated data. This serves to validate both the understanding that can be drawn from performance information and its ability to capture changes in service quality. This is a departure from the standard research process where the information to knowledge transition is simply assumed to occur.

This chapter discusses the institutional process of designing and implementing these pilot projects within the bureaucratic, public sector context that characterizes the MBTA. It describes how internal support for the pilots was obtained and how this aligns with the literature on innovation in the public sector. It then describes the pilot projects themselves and their results. Finally, it discusses the impact the pilots had on the use and acceptance of performance information by MBTA operations managers.

Altshuler and Zegans' observations about successful innovation in the public sector (Altshuler and Zegans 1997) help to explain why these pilot projects have been accepted by the OCC.

- Firstly, they are incremental modifications to existing operating procedure that are zero or low cost both monetarily and in terms of new work for managers.
- Secondly, they attempt to address existing problems, and if successful they make dispatchers' jobs easier.
- Thirdly, there has been institutional support for making measurable improvements to performance at multiple levels within the organization. It originated from the Secretary of Transportation and Director of Innovation, extended from General Manager down through the OCC Director, and has been espoused by the dispatchers, supervisors, and line managers (see Figure 4 for organizational chart).
- Fourthly, the pilots have been developed in close collaboration with the OCC and incorporate their feedback, underscoring that the intention is not to tell dispatchers how to do their job. They are also low-risk because service in the targeted areas was poor, so failure is not noticeable outside of the OCC. This may have helped to avoid the institutional resistance to change observed by Behn (1997).
- Finally, consistent with the observations from Blau (1963), staff who are dedicated to the mission of the organization like the COO & OCC director, have been supportive of innovations that help it better serve its purpose.

5.1 Generating Institutional Interest in and Support for Performance Improvement

Identifying the stakeholders and partners to provide support within the institution for performance reporting and management is a critical first step the collaborative process followed in this work. The Secretary of Transportation, who was previously GM of the MBTA, has emphasized a need for the agency to become more customer-oriented. To this end he has provided key institutional motivation and support for innovative projects. In this case, the Director of Innovation for the MBTA, who is tasked with improving the customer experience through new technology, became interested in performance management as part of a project to display real-time train arrivals in rapid transit stations. Because the MBTA is now quantifying service for customers by displaying the headways, he felt there should also be an emphasis on improving service (and eventually communicating this improvement to the public). As an agent with the support of the General Manager, he has been able to engage the Director of the OCC who felt that the current numbers were not an accurate reflection of performance and was open to new performance reports. The OCC director has thus become the primary point of contact within operations staff for feedback on the reports, while also opening up contact with the supervisors and dispatchers in the OCC who are directly responsible for service delivery.

Developing the new performance reports in collaboration with operations control managers (as described in Chapter 4) has involved the OCC director and some line managers seeing performance information regularly. This has revealed areas 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[10] – 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. This branching structure creates more operational issues than the Orange and Blue lines. For both of these reasons, it receives two dispatchers, while Blue and Orange have one each.

Specifically, northbound service on the Red Line has been known to be inconsistent, most noticeably in off-peak periods, frequently alternating between short headways and big gaps. Additionally, travel time between Davis (the penultimate northbound station) and Alewife in the peaks is significantly slower than at other times. Working with OCC dispatchers and managers on the reports has provided an opportunity to discuss these observations. They have also identified these areas of poor performance in their experience managing service, but note that their on-time performance is always over 90%. This served as an initial confirmation that the proposed performance measures can identify operational issues more accurately than the OTP reports.

[10] These numbers are based on the AFC 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 transfers are counted on all lines they take.

Figure 18: Red Line Diagram

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 staffed additional drivers at Alewife to help turn trains more quickly in the PM peak. Figure 18 shows the layout of the Red Line, for context.

5.2 Implementation and Results of the Pilot Improvement Projects

As noted previously, working to develop performance reports has revealed segments and times of day that consistently underperform. While OCC staff are aware of these issues, they have had no quantitative evidence of how they impact performance. The same data underlying the performance reports has been analyzed in greater detail to gain further insight into two problems: 1) bunching of northbound service, and 2) northbound delays heading into Alewife. While these issues are not the only ones on the Red Line, they can be addressed by the OCC internally, and thus quickly. Involving other departments may slow the process, since the specialized nature of bureaucracy limits collaboration between departments.

5.2.1 Braintree Offset

The first pilot stems from the observation that the headway graphs in the reports were showing significant variation on the northbound segment of the trunk (after the merger of the Ashmont and Braintree branches), with headways alternating between big gaps and bunches. This suggests poor coordination between the departures of trains from each of the branches. An analysis of historical running times and scheduled running times indicates that too much time has been scheduled between Braintree and the merge point at JFK/UMass. As shown in Table 10, the 90th percentile of the running time distribution is lower than the scheduled running time, even in the morning peak when northbound demand is highest and trains are most likely to be slow. Standard industry practice is to schedule arrivals at non-terminal stops at the median of the distribution, and arrival at the terminal at the 90th or 95th percentile to allow enough time to recover for the next trip (Wilson 2011).

Period	**Percentile of Running Time Distribution**					**Scheduled Time**
	10th	**20th**	**50th**	**80th**	**90th**	
Early Morning	17:26	17:41	18:17	19:17	20:15	20:00
AM Peak	18:03	18:24	19:14	20:34	21:33	22:00
Midday	17:38	17:54	18:33	19:38	20:35	22:00
PM Peak	17:35	17:52	18:33	19:45	20:58	22:00
Evening	17:23	17:37	18:12	19:10	20:18	22:00
Late Night	17:10	17:22	17:53	18:43	19:27	21:00

Table 10: Distribution of Braintree-JFK/UMass Weekday Running Times

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. The OCC has been supportive of this project because they have brought up this issue with the scheduling department previously, and the pilot required only passive input on their part – simply modifying departure times. The pilot initially targeted the off-peak midday, evening, and late-night periods to avoid impacting rush hour service if unsuccessful. The pilot was conducted on four days in the fall of 2012: September 25th and 27th, and October 2nd and 4th.

The results of this initial pilot were positive, though not statistically significant due to small sample size. On the days when the pilot was running, the periods when the schedule was modified generally perform at a better (lower) percentile for coefficient of variation (COV, measuring variability) than the same period on other days and than other periods on the same day. In Table 11 and Table 12, shading indicates a day when the pilot was running, and red text indicates the periods when the schedule has actually been changed. With the exception of Thursday September 27th, big gaps during the pilot periods are at a lower percentile of the distribution for that period, compared to surrounding periods. The results indicate that the pilot schedule seems to have had a larger effect on improving the regularity (measured by COV), than on reducing big gaps. This may be because COV captures changes in the full distribution, while big gaps only capture changes in values around its threshold.

COV thus changes when uneven headways that are below the big gap threshold become more regular. Big gaps only change when a headway crosses the threshold .

Date	Early AM	AM Peak	Midday	PM Peak	Evening	Night
9-24	8.0%	34.0%	87.5%	27.2%	6.8%	56.1%
9-25	20.6%	64.7%	20.4%	30.6%	28.4%	24.7%
9-26	56.3%	5.6%	10.2%	2.2%	7.9%	52.8%
9-27	71.2%	37.5%	31.8%	7.9%	3.4%	22.4%
9-28	85.0%	69.3%	46.5%	10.2%	1.1%	44.9%
10-1	26.4%	48.8%	50.0%	23.8%	25.0%	95.5%
10-2	73.5%	59.0%	22.7%	60.2%	79.5%	68.5%
10-3	75.8%	20.4%	57.9%	29.5%	84.0%	70.7%
10-4	2.2%	12.5%	6.8%	39.7%	53.4%	56.1%

Table 11: COV Performance Percentile by Period for Pilot and Surrounding Days in the Distribution of the Past 125 Days

Date	Early AM	AM Peak	Midday	PM Peak	Evening	Night
9-24	3.4%	94.3%	80.6%	61.3%	15.9%	5.6%
9-25	3.4%	63.6%	23.8%	43.1%	15.9%	42.6%
9-26	55.1%	4.5%	23.8%	0.0%	15.9%	15.7%
9-27	55.1%	40.9%	48.8%	0.0%	0.0%	15.7%
9-28	3.4%	63.6%	23.8%	0.0%	0.0%	15.7%
10-1	3.4%	27.2%	48.8%	43.1%	96.5%	87.6%
10-2	90.8%	63.6%	23.8%	78.4%	15.9%	87.6%
10-3	55.1%	27.2%	48.8%	43.1%	48.8%	56.1%
10-4	55.1%	4.5%	23.8%	78.4%	15.9%	15.7%

Table 12: Big Gap Performance Percentile by Period for Pilot and Surrounding Days in the Distribution of the Past 125 Days

The success of this pilot prompted operations to request a meeting with scheduling to discuss the results. In addition to the analysis of Braintree-to-JFK running times, the OCC director requested an analysis of turning times at Ashmont. The OCC observed that trains frequently did not have enough slack time at Ashmont, particularly since the introduction of single-person train operation (SPTO). Prior to SPTO, trains had a motorperson at the front and a conductor in the middle, which could speed turning a train because the conductor had half the distance to walk to the other end. A single driver must shut down the train while walking to the other end and then start it back up. A conductor on the train can head to the front of the train and switch its driving end while the operator walks forward, saving up to several minutes. There was no pilot for this problem because it would have required broader scheduling changes.

The scheduling department has been receptive to the running time analyses. They lack sufficient staff to analyze rail running times regularly, so these analyses fill 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-peaks and weekends.

Figure 19 – Figure 25 show the distribution of the effective headway metric (measured at Park Street) for the second half of 2012 versus the first quarter of 2013. The effective headway metric is calculated for each period and direction each day, so each day is one observation. The distributions for all periods except the late night have tightened, with peaks closer to the scheduled headway for the period. It is important to note that increasing the turn time for Ashmont trains has required increasing the branch headway from 13 to 14 minutes in the midday (6.5 to 7 minutes on the trunk). Because trains were better spaced on the trunk, the average headway that most passengers experienced actually decreased, despite the scheduled increase. Looking at where the distribution intersects the scheduled headway, 20% of passengers in the midday experience the expected headway, which is a dramatic increase over the 2% in the previous schedule. On Saturdays, the change was also dramatic. Under the old schedule the distribution was almost bimodal, while under the new schedule it is closer to the expected normal distribution.

As shown in Table 13, both the median and 90th percentile effective headway have fallen in almost all periods except the AM Peak. The drop was most dramatic for the weekends. The weekend graphs show a more varied distribution due to their smaller sample size (only one of each day per week), so one bad day can skew it.

Period	Median (minutes) Pre	Post	Change	90th Percentile (minutes) Pre	Post	Change
AM Peak	5.1	5.1	0.0	5.8	7.6	1.8
Midday	7.3	7.3	-0.1	8.0	8.1	0.2
PM Peak	5.5	5.3	-0.2	6.5	6.1	-0.3
Evening	6.4	6.1	-0.3	7.8	7.1	-0.6
Night	7.7	7.8	0.0	9.0	8.7	-0.2
Saturday	9.0	7.9	-1.0	11.7	8.8	-2.9
Sunday	9.5	8.9	-0.5	11.4	10.2	-1.1

Table 13: Change in Distribution of Effective Headways From 2013 Schedule

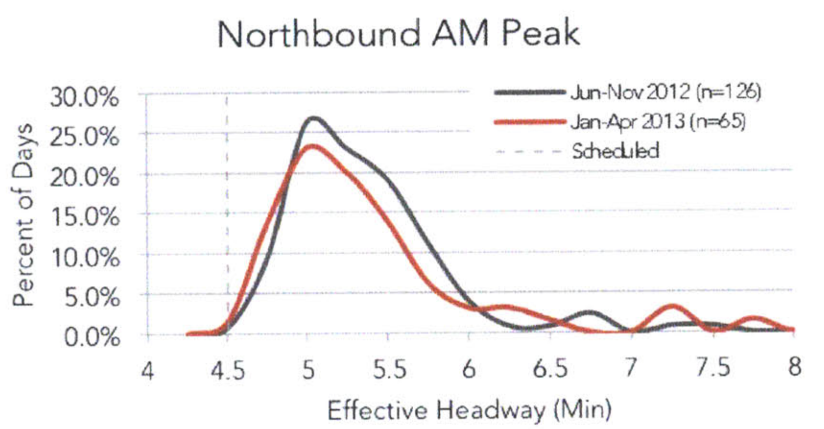


Figure 19: Red Line Effective Headways, AM Peak

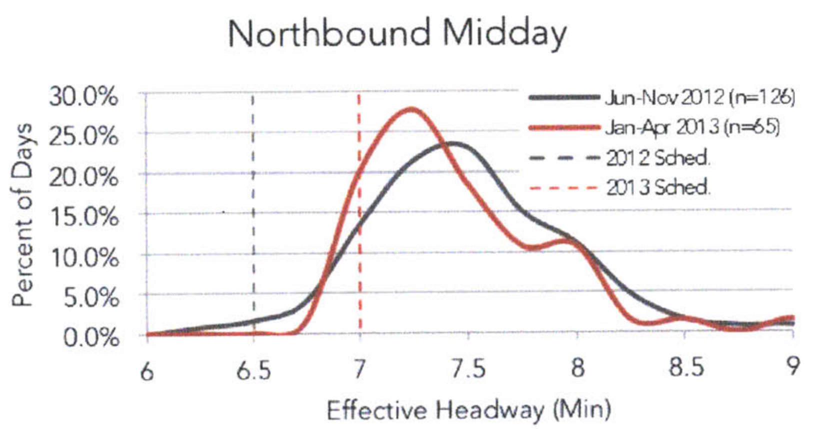


Figure 20: Red Line Effective Headways, Midday

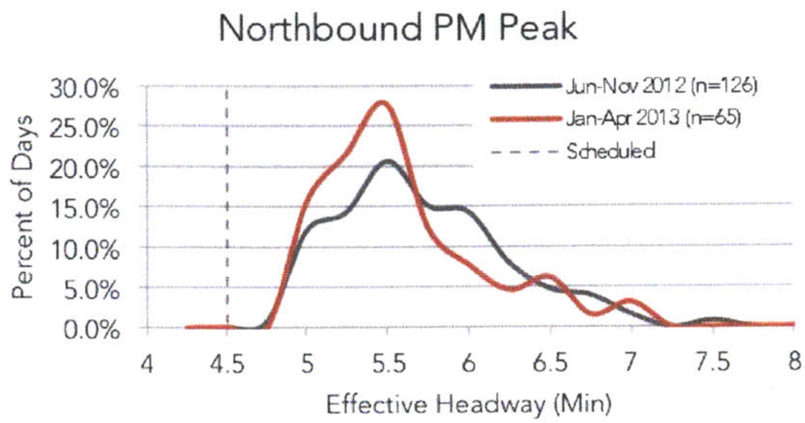


Figure 21: Red Line Effective Headways, PM Peak

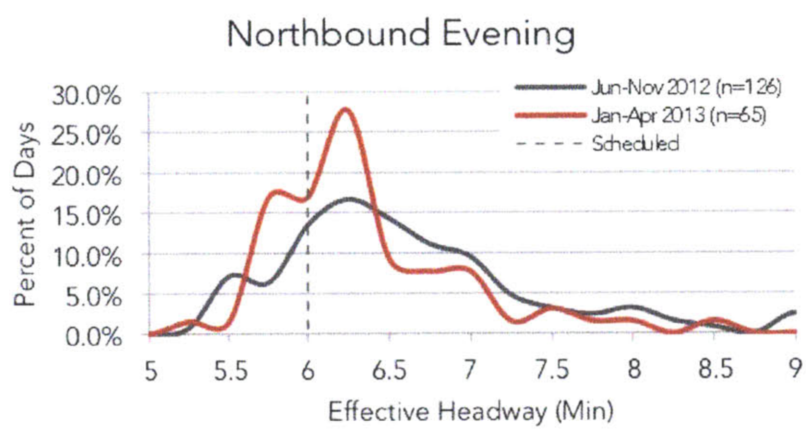


Figure 22: Red Line Effective Headways, Evening

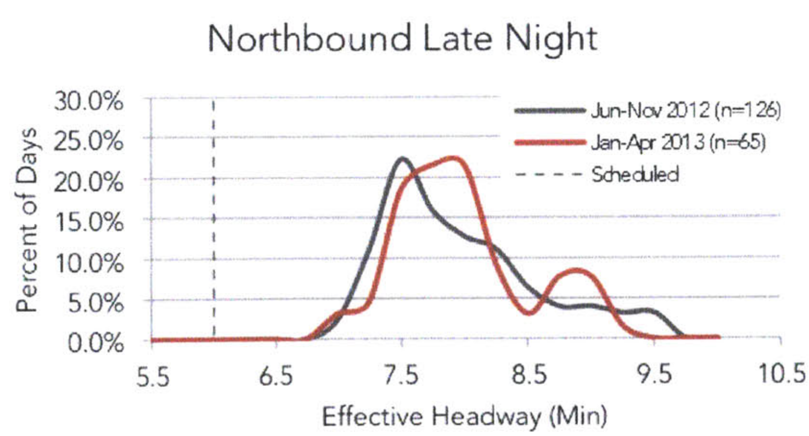


Figure 23: Red Line Effective Headways, Night

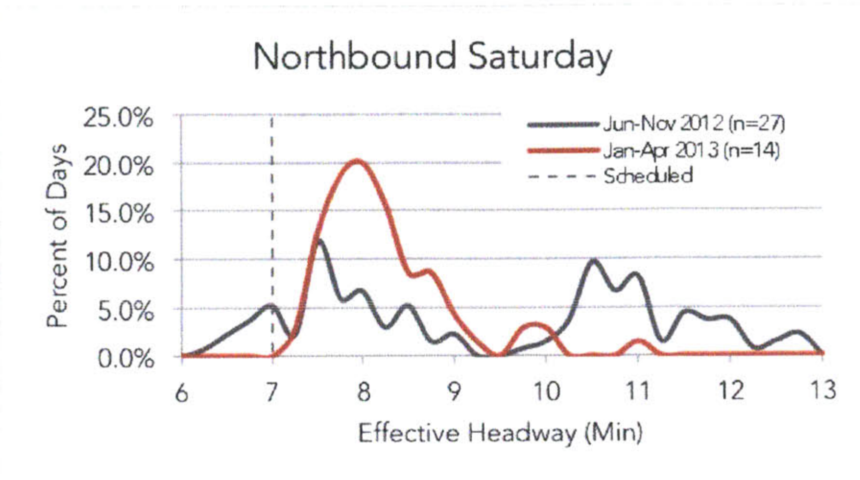


Figure 24: Red Line Effective Headways, Saturday

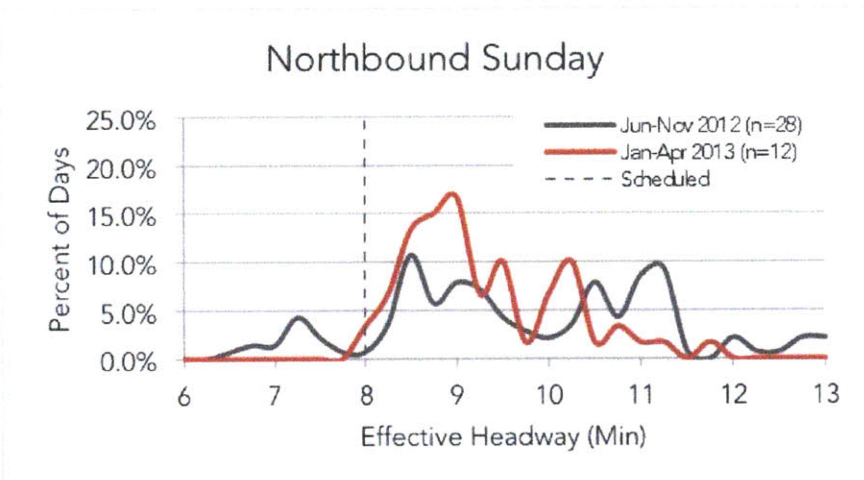


Figure 25: Red Line Effective Headways, Sunday

5.2.2 Alewife Quick Turn

The second pilot originated from dispatcher complaints that northbound trains bunch outside of 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. Because Alewife is a stub-end terminal, trains berthed on its northern track must cross to the southbound track, and crossover speed is limited to 10 miles-per-hour. Clearing the northern platform takes at least a minute longer than clearing the southern track. See Figure 26 for a diagram of the track layout. These factors increase variability and allows headway gaps to ricochet to southbound service. Table 14 below shows the minimum, median, and maximum number of trains held during weekdays from June 1st and October 2nd 2012 (before the pilot).

	AM Peak	Midday	PM Peak	Evening	Late Night	Weekend
Minimum	0%	4%	0%	0%	0%	1%
Median	48%	20%	48%	47%	11%	5%
Maximum	82%	51%	70%	100%	43%	17%

Table 14: Percent of Trains Held for More Than 2 Minutes Outside Alewife Each Day

During the three-hour peaks on a normal day, nearly half of all trains were held outside of Alewife for more than two minutes, inconveniencing thousands of customers. By contrast, on a median day 20% of trains were held in the midday. As shown in Table 15, the average hold lasted 3:47, which is longer than that the uncongested trip between Davis and Alewife.

Non-Delayed	Delayed	Delay Duration
3:21	7:07	3:47

Table 15: Average Travel Time from Davis to Alewife

The pilot project has staffed additional operators[11] 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. The pilot ran from 4:00-7:00 PM on six days in December 2012.

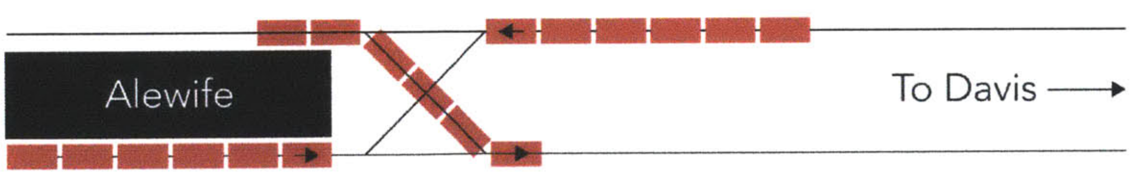


Figure 26: Alewife Track Layout Showing Train Crossing to Southbound Track

[11] 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 that in non-driving positions around the system.

The results of the initial pilot project were positive, and it has been well-received by the dispatchers and OCC managers. Detailed results include:

- The average number of trains delayed[12] fell from 17 to 9 per PM rush. Those delayed trains were held for an average of 45 seconds less, and the distribution tightened, with the 90th percentile of holds falling by more than 2 minutes.
- The average running time from Downtown Crossing to Alewife and back (omitting the dwell time at Alewife) was reduced during the pilot by 3:31 seconds (7%).
- Running times from Central to Alewife fell an average of 116 seconds per train, with the worst day during the initial pilot still reducing running times in this section by 45 seconds on average, and the best day by 2 minutes 32 seconds.
- Total big gaps at all stations between Broadway (northbound) and Downtown Crossing (southbound) were reduced form 8% to 4%
- The COV of headways fell by 10-20% in the area affected by delays: Harvard northbound to Downtown Crossing southbound
- The reliability buffer time (the difference between 50th and 95th percentile) for running time was reduced by 50 seconds per segment on average

The success of this pilot has led the director of the OCC to commit to having reserve operators stationed at Alewife for the PM peak in subsequent crew schedules, 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 of 2013, the median travel time between Davis and Alewife have 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.

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%

Table 16: Running Times between Davis and Alewife (minutes)

[12] Defined as taking longer than 5:20 between arriving at Davis and arriving at Alewife, which is 2 minutes longer than the median travel time

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 outside of Alewife. Before the pilot, 43% of PM Peak trips took longer than 5 minutes, now that is only 32%. The benefits have diminished since the initial pilot, however.

Period	**Taking > 5 min.**	**Taking > 7 min.**
Pre-pilot	43%	14%
Pilot	23%	3%
2013	32%	10%

Table 17: Percent of Slow Davis-Alewife in PM Peak

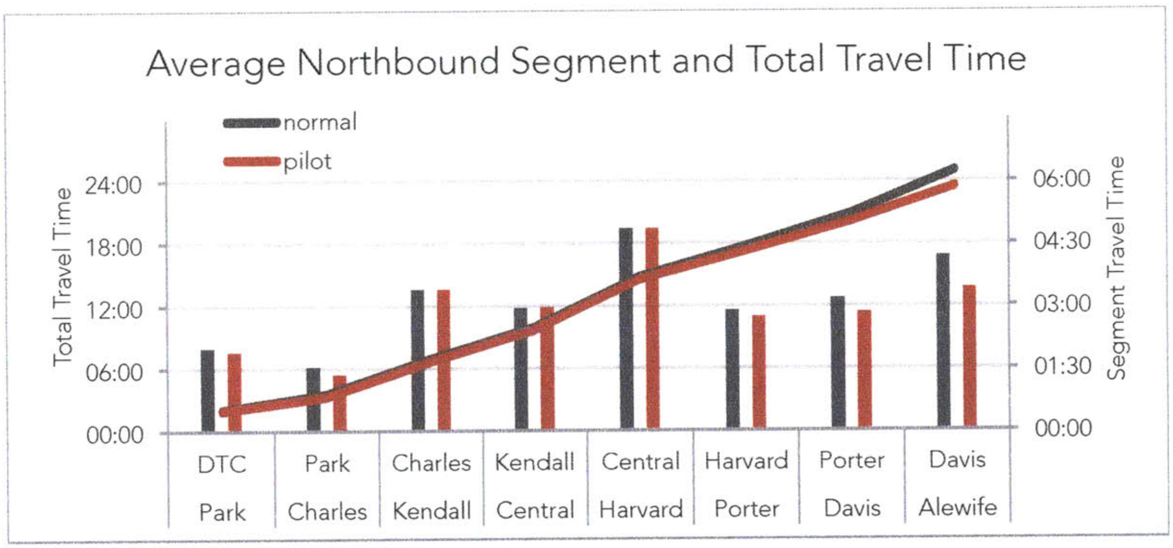


Figure 27: Change in Average Northbound Travel Time

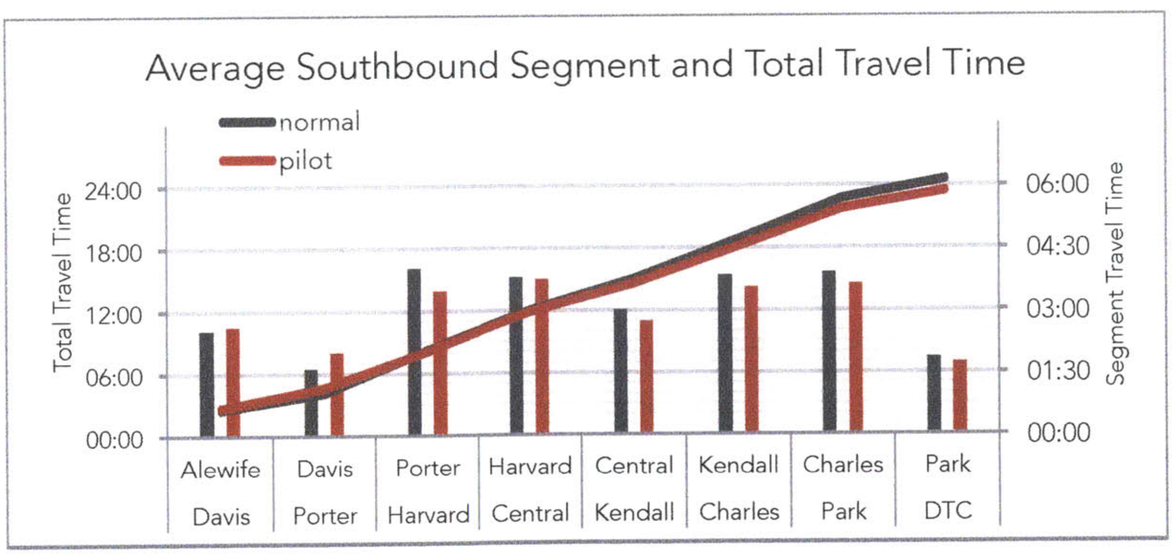


Figure 28: Change in Average Southbound Travel Time

The improvement between Alewife and Davis has positive consequences for the trip from Downtown Crossing to Alewife and back southbound from Alewife to Downtown Crossing. The average northbound travel time in the PM Peak has dropped 1:30, or about 6%, and average southbound travel time has also dropped 5%, as shown in Figure 27 and Figure 28.

This reduction in the benefits from the initial pilot to its extended implementation has several possible explanations. In showing these statistics to the OCC Director, he noted that additional drivers are not always available for quick turns because they are used to cover shifts when drivers call out sick. He also suggested that the high degree of management attention paid to Alewife during the initial pilot was a significant factor in its success. In the initial pilot, there was an instructor and supervisor at Alewife in addition to the reserve operators. The OCC Director stated that management presence often results in more prompt driver performance. In addition to the management presence in the field, there was also additional management attention on the dispatchers in the control room. The OCC director was present in the control room during much of the initial pilot and made it clear that turning trains at Alewife quickly was a priority. MIT researchers were also present. While not authoritative, the presence of outsiders may have also induced dispatchers to pay more attention to the terminal. Making information publicly available may have an analogous effect.

5.3 Impact of the Pilot Projects on Institutional Acceptance of Performance Information

The pilot projects have also served as a test implementation of the performance reports, revealing areas for improvement in the reports that are discussed in more detail in Chapter 4. The performance reports have been used as both a basis for developing the pilot projects and as a tool for measuring their effect. The intention has been to demonstrate the capabilities of performance measures and analysis (1) to address operational issues and (2) to effectively reflect variations in service. The pilot projects have been an opportunity for MBTA personnel to see rail performance information on a daily basis. Receiving reports on a regular basis during the pilot projects has given operations managers the opportunity to see the metrics on various days under different conditions and become familiar with how the numbers and charts relate to their experience in the control center. The fact that both pilots have been extended suggests that MBTA staff view the quantitative information favorably. This is supported by conversations with staff about the reports and pilot programs.

In discussing adjustments to the Braintree departure times, multiple dispatchers have said that they knew the schedule was inaccurate because they often slowed down Braintree trains to stagger arrivals at the junction, while Ashmont trains often did not have enough time to turn around. While OCC managers said they had complained to the scheduling department before, there was no quantitative evidence because the reports showed high OTP. OCC staff may have accepted new performance reports despite them showing worse performance in part because they confirmed their intuition that the trains were poorly scheduled. 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.

Another sign that OCC managers value the performance information was a conversation with the Line Manager for the Red Line. After receiving performance reports regularly as part of the Alewife pilot, he mentioned that they were also useful for investigating customer complaints. When customers would complain about a long wait at a specific time, he would look at the headway chart to verify their claim and respond to it. In his words, "this gives me everything I need to know."

In discussing the results of the extended implementation of the Alewife pilot (first quarter of 2013) with the Director of the OCC, he noted that it was valuable to have quantitative reports of the pilot's impact. He believes the supplementary evaluations of the Alewife pilot (presented above) provide an argument for adding these resources to the Red Line permanently. He was planning on using these analyses to make this case in his next budget proposal.

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 on March 25, 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.

The daily report provides additional quantitative information on service that OCC personnel can relate to information about other circumstances such as track conditions, power problems, disabled trains, medical emergencies, and other service disruptions. For example, construction activity for a new station at Assembly Square on the Orange Line requires trains to move at 10 miles-per-hour through a section of track where there were workers. This shows up every weekday in the report: nearly all trains between Oak Grove and Sullivan Square are counted as slow from 7:00 AM to 3:00 PM (construction hours) and total passenger travel time increased. The OCC Director and COO have expressed frustration that the capital construction department is not sensitive to the operational impacts of their work, and are pleased to have a tool that quantifies customer impacts. Comparing the total passenger travel time before and after construction started, construction is causing 675 customer-hours of delay per day. Figure 29 shows total customer-hours of travel time on the Orange Line for weekdays since October 2012 to present, with a two-week moving average. Construction started in February 2013. Since then, total travel time has trended higher.

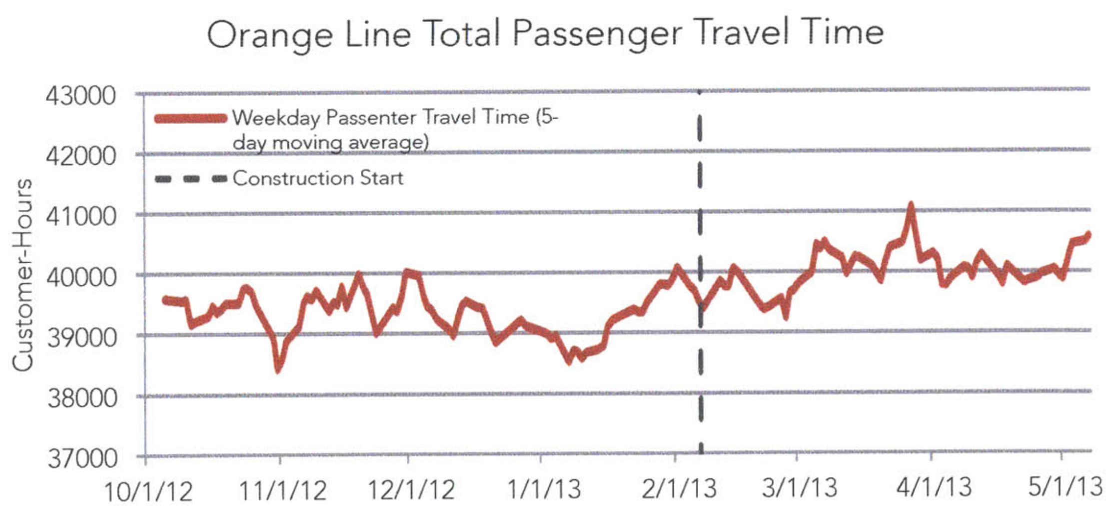


Figure 29: Total customer-hours of travel time on the Orange Line since October 2012

These cases suggest that the OCC has accepted and engaged with the reports as a tool to communicate their experience with other departments. They can use them to make the case for how the work of other departments impacts service and customers. This is an important role for performance information that was not considered explicitly in the initial designs of the performance reports.

In addition to seeing how outside circumstances affect performance, meetings with the COO and OCC Director also showed that they can use them to identify issues with their own management. After seeing performance information for multiple days, the COO noted off-peak performance (midday and night) was unacceptable. He admitted that after the rush hour ends, management pressure eases; less attention is paid to off-peak performance.

The implementation of the programs thus achieved two goals. The pilots have led to operational changes that achieve improvements for passengers, though these do not completely eliminate the problems they were targeting. Pilots also successfully engage MBTA staff with performance information about the rail system. Gaining the trust of operations personnel in the performance information has the possibility of producing additional service improvements as managers see service quality regularly and work to improve it.

6 Findings, Recommendations, and Further Research

The process of generating performance information from automated data, arranging this information into performance reports, and applying it to modify service provides important lessons about performance measurement, reporting, and management. This chapter summarizes and generalizes these lessons, and develops recommendations on how the MBTA and other transit agencies can extract more value from their existing data.

This research provides several key findings about how to take ADCS data, translate it into performance information that is useful for the *people* controlling service, and apply this information to actually change service. These are summarized below and are discussed in detail in the remainder of this chapter.

- Perspective matters: measuring from a customer perspective is a strong basis for evaluating performance
- Process matters: developing measurements 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 read
- Collaboration on performance reporting facilitates introducing service changes because it builds trust and support from both middle and upper management
- Collaboration also improves the substantive quality of the performance indicators by incorporating the domain knowledge of employees
- Proposed service changes are more likely to be implemented when:
    - They address recognized problems

    - They will reduce demands on service controllers' time
    - They are easily reversible
    - They have the potential to benefit a large number of customers and will be visible to them
- Pilot projects and performance reports reinforce one another
    - Pilots show how information can be applied to 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

From these findings, this chapter goes on to make recommendations for how the MBTA can expand the use of automated data and its impact on performance.

- Set performance goals for heavy rail based on the new metrics using historical performance as a baseline from which to improve
- Publish performance information to customers more frequently and in more detail than the current monthly and annual reports
- Reorient metrics for other modes – bus, light rail, and commuter rail – around passengers
- Augment real-time information for heavy-rail operations control to display headways, because this is how both the previous and new reports judged their work
- Engage additional staff such as inspectors in managing headways
- Analyze data at regular intervals to inform infrequent processes like scheduling or fare policy changes
- Assess the customer impacts of unplanned disruptions to quantitatively evaluate capital needs
- Assess the customer impact of scheduled service changes to improve planning for construction and maintenance, sporting events, and special events
- Establish an institutional responsibility for applying data to operations in an employee or office that acts as an internal consultant and coordinator of new initiatives

- Consider planning, and reporting uses of data when specifying and designing future data systems

Finally this chapter proposes several areas for future research that were not addressed but are important to more fully understand how to make automated data more effective.

- Expand to other modes of transit or other agencies
- Expand to dynamic, interactive information for consumption by the general public
- Explore alternatives to weighting performance by passenger volume
- Measure capacity and crowding
- Investigate dispatchers' reactions to different incentives and perspectives of performance information, such as highlighting good rather than poor performance
- Research the impact of calculating performance metrics in real-time and displaying them to dispatchers, rather than showing overall performance the next day.

6.1 Key Findings

By evaluating the existing performance reporting paradigm at the MBTA within the data-information-knowledge-wisdom framework, this research makes several findings about performance measurement. Applying these to develop new performance reports produces additional conclusions about the influence of design on the communicative value of information. Piloting operational changes based on new performance information provides insight into the institutional process of data-driven innovation.

6.1.1 Perspective Matters

How an agency measures service determines what it knows and is therefore able to improve. Evaluating MBTA performance under the current on-time performance standards versus other measures such as big gaps, headway regularity, customers experiencing big gaps, and travel time paint different pictures of service quality. While perhaps an obvious point, it behooves transit agencies to consider the perspective they are taking when designing or updating performance measurements. For the 8% of customers traveling to or from the terminal on one of the Red Line's branches (the perspective implied by the current standard), the MBTA's current performance is acceptable. For the 67% of riders only traveling on the trunk, however, it is irregular. Because this has not been measured previously, it is difficult for the MBTA to address.

Incorporating information on customer demand and 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. For example, given that most of Red Line ridership is on the trunk, the MBTA should strive to maintain an even headway on this segment, rather than managing to branch headways. Even trunk service will also result in even branch headways. 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.

6.1.2 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 (Rowley 2007). If the intention of the performance information is to generate knowledge – an understanding of relationships that can be applied to affect service – then examining how people react to different information is a critical part of the design process. Soliciting feedback from the users of the information provides insight into what they are interpreting from it, which determines what they can do with it. This may, however, create tension when designing a report for multiple audiences, as they may have differing responses to the same information and presentation techniques. Genuine collaboration respects and incorporates the opinions and experience of service managers. In doing so, it improves their acceptance of the metrics, since they know that their input was taken into consideration. Moreover, the collaborative process incorporates the dispatchers' domain knowledge that is critical to making the reports accurately represent service. Without such domain knowledge, the information represents an outside perspective on service quality. Such engagement also emphasizes that performance metrics are intended to be constructive and useful, not punitive.

Collaboration with internal stakeholders has been a labor-intensive undertaking, and soliciting public input may be even more work. Putting this effort into developing reports, however, ensures they are meaningful to operators and customers and enables a distributed knowledge generation model. This may reduce the amount of effort required to identify service improvements in the future because multiple groups have good information from which they can assess service.

6.1.3 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 to base decisions on. 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, given past experience. 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 the MBTA's Smart Bus Mart reporting system provides a flexible reporting tool for bus performance, managers and dispatchers are mostly familiar with the Key Routes On-time Performance Report that is e-mailed to them every day. Smart Bus Mart allows managers to investigate performance issues or seek more detailed information, but this takes time away from actively managing the service.

This creates a temptation to provide a wealth of information about service in a performance report. While performance reports can certainly be information-rich, this also demands that close attention is paid to information design. Excessive amounts of numbers or repeated graphs can result in information fatigue that reduces the communicative power of the reports to generate knowledge, as this research learned through the intermediate drafts of its reports. Using graphics can be a useful technique for condensing information. The early drafts of new performance reports included numbers that quantified the information in the headway graph, but in conversations with OCC staff, it was clear that the graph was better at communicating the headway performance.

6.1.4 Collaboration on Performance Information Builds Trust for Pilot Projects

Support from multiple levels of management is required to implement pilot projects, which can be established through cooperation on the underlying performance analysis. This work has explored the process of implementing changes based on performance information and analysis. This was made possible due to close coordination with MBTA staff, both in the GM's office and in the OCC. Support from senior staff such as the Director of Innovation and OCC Director was essential to getting approval from upper management to do pilot projects. This aligns with Altshuler and Zegans' findings that "being close to clients and relying on them to convey positive messages" facilitates change in bureaucratic agencies (Altshuler and Zegans 1997, 78). The support of the OCC director was also critical in building support among the dispatchers and other operations personnel to implement the pilot projects effectively. Because the pilots were associated with the OCC director who works closely with the dispatchers, they did not dismiss the projects as a micromanaging directive. Instead, they accepted them as an opportunity to make their jobs easier.

6.1.5 Specific Characteristics of Pilot Projects Affect Their Viability

Pilot projects are more likely to be successful if they address recognized problems, reduce the amount of work required by operations staff, are easily reversible, and have the potential to benefit a large number of customers. This is consistent with Altshuler and Zegans' work, which finds that addressing widely recognized problems and proceeding incrementally are common elements of successful changes in public agencies (Altshuler and Zegans 1997). By specifically targeting issues that were recognized as problems by dispatchers, the pilot projects conveyed the value of measurements. They were presented as a tool to make their jobs easier rather than as a report card that would be used to criticize their work.

The OCC was also open to the pilot projects because they were easily reversible if they believed they were harming service to an unacceptable degree. This is part of the reason the Braintree schedule change was tested out as a pilot first: if the change had a negative effect it could be easily reversed. If the change had first been made in the official schedule, it would have been locked in for the next three months. It was easier to pilot a change to an operation like dispatching that occurs every day, rather than in an operation like scheduling that happens every few months.

This work targeted the Red Line because it carries the most passengers out of the MBTA's services, so it receives significant institutional attention. This bolstered internal support for projects to address Red Line issues, particularly among upper management, because improvements would have a large impact on customers and be visible if successful.

6.1.6 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 a operational context. 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 see pilot projects judged as failures, such cases may still have the benefit acquainting operations staff with performance information.

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. Because turning trains around quickly at Alewife required dispatchers and station managers to actively engage in train departures, this added to the benefit of the pilot procedure. As the procedure has become institutionalized, its performance benefits have diminished, in part due to less active management of Alewife. Managers were also not receiving daily performance reports at the beginning of the extended implementation. The lack of feedback information may have contributed to the lower benefits at the start of the full implementation. After making managers aware of the drop in time savings, performance has improved again, though not to the level seen in the initial pilot.

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 consistent high-quality service, though this work did not succeed at taking performance information to this point.

6.1.7 Performance Information is a Communication Tool

These projects also revealed that performance information can facilitate communication between departments. It provides a common basis for identifying and discussing issues, as demonstrated by the pilot to reschedule the Red Line branches. This requires that all parties understand and accept the performance measurements. The existing OTP reports are limited as a communication tool because dispatchers do not trust their information. Providing measurements that conform to both dispatchers' and schedulers' understanding of service has enabled a conversation about how to address issues, rather than disagreeing about whether there were problems.

6.2 Recommendations for Expanding the Use of Automated Data and Performance Information

As a result of this research, some MBTA operations managers are receiving daily performance reports that incorporate customer information for the Red, Blue, and Orange Lines. This is a significant change in the way the MBTA understands service on these lines, and has led to two successful pilot projects to improve quality. In addition to generating better performance information, there are other applications of automated data that could benefit the agency.

6.2.1 Set Goals for Passenger-Oriented Metrics

Firstly, as it has done with its other performance indicators, the MBTA should now set goals for the passenger-centric performance metrics contained in the rapid transit daily reports. Without goals, the power of the information to produce service improvements is limited. These goals should be achievable or they may be ignored because operations personnel cannot see progress towards them. This would be the reverse of the rail OTP reports, which are ignored because service is never judged to be poor. Historical performance information should be used to set goals that are better than the median of the distribution for each metric. These should be revised upwards as performance improves at regular intervals, such as every time the schedule is revised. This provides a basis for continued improvement in the service experienced by customers.

6.2.2 Re-orient Other Modes' Performance Around Customers

The MBTA should also re-orient performance metrics for other services like light rail, bus, and commuter rail, around passengers. Because detailed O-D data for these modes is not yet available, this requires rethinking how to incorporate passenger demand into performance for these modes. This is also an area for future research.

6.2.3 Increase Real-Time Focus on Headway Management

The MBTA should also augment its real-time information to focus on headways for frequent service. Both the old and new performance reports take the headway between trains as their fundamental metric, but the rail dispatchers do not currently have headway information in their standard system view. To manage service effectively, dispatchers need real-time information about service in terms of the metrics used to judge performance, so headway information should be displayed to heavy rail dispatchers. The MBTA should also engage personnel such as inspectors in managing headways. Inspectors in rail stations have real-time information on headways and could make adjustments to avoid bunches and big gaps. Because inspectors do not see the entire system, the OCC would need to establish parameters within which they could exercise control, such as restricting this practice to certain stations at certain periods. Bus inspectors have access to handheld devices that display information similar to what dispatchers see, which they should use to manage headways from their station posts. Having multiple levels of management making service adjustments requires clear and frequent communication between them to ensure that operators are not given conflicting or redundant instructions. An alternative would be to give operators information about the headways of their leader and follower and make it clear that their duties include maintaining an even headway.

This focus could extend beyond management techniques to piloting new operational procedures like those tested in this research. For example, the MBTA could attempt and measure the impact of moving to drop-back scheduling. In a drop-back schedule, a new driver switches onto a train at the terminal and the former driver "drops-back" and takes the next train to pull in. This allows trains to pull out more quickly by overlapping the driver's walking time with the headway.

6.2.4 Analyze Automated Data Regularly

In addition to performance reporting and management, automated data should be analyzed at regular intervals to inform other operational and policy decisions. In analyzing running times on the Red Line, this research has reassessed an assumption that had not been revisited for years. Automated data enables assumptions that go into scheduling to be revisited every time a new schedule is produced, at little marginal cost. The MBTA is currently doing this for bus scheduling through Hastus ATP, and should expand this practice to its rail services.

6.2.5 Evaluate the Impact of and Response to Disruptions

Assessing the impacts of service disruptions is another potentially valuable application of automated data. Combining performance information with records of unplanned disruptions for equipment failures, signal and track problems, medical or security emergencies, allows the MBTA to quantify the effects of these incidents on service. These quantifications can help to make the case for capital investments and inform investment priorities. The Red and Orange Lines suffer from disabled trains multiple times per week, if not once per day. Calculating the number of passengers inconvenienced and the duration of the delays gives an estimate of the passenger benefits of upgrading the rolling stock. A cost per passenger hour saved could be a factor to consider in prioritizing capital and maintenance projects.

Following from the assessment of disruptions, the MBTA could also evaluate and improve the strategies that it uses to recover from disruptions. Analyzing incidents both in terms of their passenger impact and the speed and effectiveness of the recovery enables managers to assess their efforts. In the case of disruptions, managers are focused on safely returning to normal service. Maintaining service quality to the extent possible may be a secondary objective, perhaps a distant one. As the performance reports did with normal service, discussing the analysis of service responses may generate ideas for changes to procedures to improve recovery. It may also enable strategies to be adapted to specific types of disruptions by revealing performance differences between strategies that managers did not notice as they were focusing on the incident. Such analysis would allow managers to augment their standard operating procedures in the face of disruptions and reduce performance impacts.

6.2.6 Evaluate the Impact of Scheduled Service Changes

Evaluating the customer impact of previous planned service disruptions for construction and maintenance will allow the agency to understand their impact on service and customers. This knowledge could be applied to improve planning for construction and maintenance in order to minimize the impact on customers. One example is the daily delay for southbound Orange Line trains between Wellington and Sullivan due to construction of the Assembly Square station. Between 7:00 AM and 3:00 PM, all trains take at least 15% longer than normal. Knowing that this causes 675 hours of passenger delay each day could provide an impetus for switching to a night construction schedule (possibly 8:00 PM to 4:00 AM). This information could be applied to evaluate the customer impact of other construction changes, such as the impending closure of Government Center Station for rehabilitation. A second example would be calculating the total delay created by the reconstruction of the Anderson Bridge (connecting Cambridge to Allston near Harvard) that serves the #66 and #86 buses, which could provide as a basis to design strategies to mitigate the problem. In cases like the Anderson Bridge and other upcoming bridge reconstructions like the Longfellow, River Street, and Western Avenue bridges, the construction is imposed by MassDOT. Documenting the impacts to customers gives the MBTA leverage to request that MassDOT arrange mitigation to avoid the impacts or compensate the MBTA and its customers.

6.2.7 Evaluate Performance During Special Events

The MBTA should also evaluate its performance around sports and special events more closely and make efforts to improve it where necessary. Many of these events occur during off-peak and weekend times, when service levels are lower. Sports and special events produce higher customer volumes that the schedule anticipates. The MBTA adjusts service for New Year's Eve, Independence Day, and the Boston Marathon by providing rush-hour service from the afternoon through close of service. Evaluating performance for other sporting and special events would allow the MBTA to augment service in a more targeted way that does not require as many resources but still provides capacity where needed. Additionally, for some patrons, special events are their only experience on the MBTA. Ensuring that these customers have a high-quality experience may improve their image of the agency and support for transit.

6.2.8 Make Performance Information Available Publicly

While passengers currently have some real-time information about service, more detailed historic performance information should be made publicly available. Increasing transparency about service quality augments the anecdotal impressions of service that currently form the core of the public and political perception of the MBTA. Knowing that data is publicly available also creates an additional incentive for operations personnel to maintain high quality service. Both bus and rail dispatchers are sensitive to creating big gaps because they often result in complaints from customers that are followed by inquiries from their managers. Knowing that customers, advocacy groups, and the press can see more than just the service they experience provides dispatchers with an additional incentive to be concerned with overall service quality. Transit advocacy groups and interested individuals putting pressure on top-level managers would filter down in the same way. Public performance information may also depict MBTA service to be better than some riders believe, counterbalancing their anecdotal impressions of service with a broader perspective. This may generate positive reinforcement and provides a similar incentive to maintain good service.

In making the reports public, the MBTA should solicit additional design input from a focus group or other representatives of the public to ensure that the public-facing reports communicate information effectively. As this research shows, the input of the eventual users is critical to designing information that is comprehensible to them. The management incentive of publicizing performance information, however, could be achieved by simply publishing the current reports because dispatchers would know that their performance was visible outside of the OCC. The effect may be limited if they are not understood by advocates, however, since this limits their understanding of service and ability to express concerns or commendations.

In adapting performance information for public consumption, the MBTA should include indicators that distinguish between what is due to operations management and what is due to circumstances outside of their purview. For example, information on equipment, signal failures, and passenger incidents on the heavy rail system would qualify poor performance. For buses, an indicator of traffic volumes or vehicle shortages due to shuttling would be informative. Without such information, the public may blame the MBTA for service issues that are outside its control. Displaying equipment and traffic issues could also create public pressure on other government bodies such as the state legislature or city traffic departments to address these problems.

6.2.9 Establish Internal Responsibility for Applying Automated Data

If the MBTA intends to expand the application of performance information to impact service, it should establish the institutional responsibility for doing so in an individual or office. While the agency has the data it needs to implement these recommendations, it has only re-oriented processes around the data as ad hoc projects like this research. The experience of this work in piloting new operational strategies suggests several important features for the institutional role of applying performance information to change practices at the MBTA. Firstly, it should be located within an executive-level office (GM or COO) so that it clearly has the institutional backing of upper management. The position should work closely with the departments whose practices will be affected. This research worked closely with the OCC, which has enabled it to incorporate their input into the analyses and to overcome the distrust of outside analysis.

Such a position would operate like an internal consultant, with a dedicated role of examining operations from a perspective that daily managers do not have time to consider. The responsibilities of such a position should be to evaluate how the agency's operations could be improved by analyzing its automated data, starting with those that are most visible to customers. This employee or office would then work with the relevant department to ensure that analyses are appropriately framed and do not leave out important factors. This would allow them to produce recommendations about how to improve operations. Unlike external consultants who leave after making recommendations, the internal position should then coordinate the different departments and individuals whose cooperation is required to implement the projects. This ensures continuity between the initial analyses that identify the problems, the proposed solutions, and their implementation.

6.2.10 Consider Planning and Reporting Uses of Data in New Systems

The MBTA is preparing to track light rail vehicles more accurately as part of the Green Line extension. In designing this vehicle tracking system, it should consider the historical and reporting uses of the data in addition to the real-time display to controllers. The Green Line poses particular issues due to its subway-to-surface operations that will require two separate tracking mechanisms. The data from the two systems should be integrated into a single database to facilitate performance measurement. Additionally, the real-time system should be able to display both the joint headway and the headway for the specific branch. The MBTA should also consider other operational issues it might want to capture, such as being stopped at a traffic light versus at a stop. This example would require the tracking system to record when velocity is zero and if the doors were open. Considering as many future uses as possible when designing the Green Line tracking system may avoid some of the constraints faced in this research.

6.3 Opportunities for Further Research

This work has used a case study of developing and applying performance information for heavy rail transit at the MBTA to draw conclusions about the impact of process and design on the effectiveness of the information. Further research could be conducted on what elements of information design are most effective for other modes of transit, other agencies, or for other sectors of the transportation industry.

6.3.1 Research the Developing Dynamic Information or For Multiple Audiences

This research was limited to generating static, non-interactive information. Additional research should focus on the process and design of dynamic and interactive performance information and how to apply it to service. The reports in this research have been designed with input from a single audience – the MBTA's OCC managers and dispatchers. This group has a fairly uniform understanding of service and background knowledge that has guided the design of the reports. While some thought was given to comprehensibility for a general audience, public input was not solicited. Future research should consider how to design transit performance information for multiple, diverse audiences. Because they have differing amounts of background knowledge and different interests in the information, this may imply a alternate design processes, graphic techniques, media, or even different metrics.

6.3.2 Explore Alternate Approaches to Weighting Performance

The approach taken in this research to incorporate passenger information into performance measurement weights performance based on customer demand. The implication here is that the segments with the most passengers are the most important. In the peak periods, a primary function of a transit system is to provide high capacity into dense employment areas like downtown business districts. In this scenario, weighting by passenger volume is appropriate.

In other periods of the day or in other parts of a city, the primary function of the transit network is providing mobility, not simply capacity. Weighting by passengers places a lower priority on segments with lower demand and may hide poor performance on these segments. Moreover, it does not take into account customer need. For example, the Ashmont branch serves approximately half the number of customers as the Braintree branch, but many more are transit-dependent. Whether a higher demand segment that serves mostly choice riders should receive more weight than a lower-demand segment with more captive riders is a subjective judgment. Likewise, in the off-peak periods, services run less frequently, so poor performance may result in high total trip times, which is not taken into account by measuring service on each route individually. Future research should consider how to incorporate such concepts into performance metrics.

6.3.3 Estimating Crowding from Automated Data

The passenger weighting approach used in this research assumes that all passengers board the first train that arrives. In the peak hours on the MBTA, this is inaccurate on its face – trains reach capacity. Future research should explore how to identify vehicle crowding both in real-time and in historical data. In the latter case, a gross measure of crowding comparing total boardings to total capacity in a given period would allow service planning to know where they needed to add service. This is likely possible with existing historical AFC data for rail and passenger counter data for bus. Neither of these sources is available in real-time at the MBTA, so measuring crowding for operations purposes may be more challenging.

6.3.4 Explore the Impact of Different Types of Information on Management

Further research could also study what incentives different types of performance information provide and how these influence dispatcher behavior. Much of this research assumes that seeing a quantification of poor service provides an incentive to address its causes, which in turn improves quality. Future work could evaluate if operations controllers respond differently to metrics framed in other ways, such as positively oriented measurements that focus on what is working well.

This work could be extended to examine the impact of producing performance metrics in real-time to provide immediate feedback to dispatchers and the public. While they currently have real-time information on the state of the system, there could be an additional benefit to evaluating the day's performance as it occurs, rather than viewing it in the past. This would serve as a live score for the day's performance, a barometer that operations personnel are able to affect in real time. Such a real-time quantification of overall performance may be more effective in influencing dispatcher behavior in the moment, while daily reports are more relevant for managers.

6.3.5 Incorporate Passenger Information Into Performance Metrics for Other Modes

As mentioned previously, incorporating passenger information into metrics for other modes presents additional challenges. On many bus and light rail systems, passengers are only recorded on entry. Headways can still be weighted by passenger arrivals, since these can be estimated. Since passenger destinations are not known for these modes, slow trips cannot be calculated for specific O-D pairs. Instead, a vehicle that takes longer than scheduled to run its route could be weighted by the total number of customers boarding that vehicle, since AFC records are linked to the vehicle for surface light rail and bus. MIT is currently working on applying the methodology from Gordon (2012) to the MBTA, which would provide O-D and transfers for its entire network. Success in this project would eliminate the need for alternate methods of incorporating passenger information at the MBTA, though they may be useful for other systems.

6.4 Final Thoughts

The opportunities for making powerful information out of simple spatiotemporal data for vehicles and customers 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. As the number of data sources and dimensions of analysis increases, so does the need for collaborative information design. Adding more 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.

References

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.

APTA. "2012 Public Transportation Fact Book." American Public Transportation Association, Washington D.C., 2012.

Barry, James, Robert Newhouser, Adam Rahbee, and Shermeen Sayeda. "Origin and Destination Estimation in New York City With Automated Fare System Data." *Transportation Research Record* (Transportation Resaerch Board) 1817 (2002): 183-187.

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.

Blau, Peter Michael. *The dynamics of bureaucracy; a study of interpersonal relations in two Government agencies.* Chicago: University of Chicago Press, 1963.

Boddy, David, Albert Boonstra, and Graham Kennedy. *Managing Information Systems: an Organizational Perspective* . 2nd. Harlow: Pearson Education Limited, 2005.

Chen, Cynthia, Hongmian Gong, Catherine Lawson, and Evan Bialostozky. "Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study." *Transportation Reserach Part A* 44 (2010): 830-840.

Christian, Brian. "The A/B Test: Inside the Technology That's Changing the Rules of Business." *Wired.* Apr. 25, 2012. http://www.wired.com/business/2012/04/ff_abtesting/all/1 (accessed May 20, 2013).

Evans-Cowley, Jennifer. "There's an App for That: Mobile Applications for Urban Planning." *Social Science Research Network.* Oct. 9, 2011. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1951069 (accessed May 20, 2013).

Falconer, Gordon, and Shane Mitchell. "Smart City Framework: A Systematic Process for Enabling Smart+Connected Communities." *Cisco Internet Business Solutions Group.* Sept. 2012. http://www.cisco.com/web/about/ac79/docs/ps/motm/Smart-City-Framework.pdf.

Frumin, Michael. "Automatic Data for Applied Railway Management: Passenger Demand, Service Quality Measurement, and Tactical Planning on the London Overground Network." Massachusetts Institute of Technology, Cambridge, 2010.

Gerstle, David. "Understanding Bus Travel Time Variation Using AVL Data." Massachusetts Institute of Technology, Cambridge, 2012.

GIRO. "ATP: Run-Time Analysis & Improved Punctuality." *GIRO.* Apr. 11, 2011. http://www.giro.ca/docs/public/powerpoint/hastus/en/GIRO-MKT-DUBAI201104-ATP.pdf (accessed 05 21, 2013).

Gordon, Jason. "Intermodal Passenger Flows on London's Public Transport Network." Massachusetts Institute of Technology, Cambridge, 2012.

IBM. *What is Big Data.* 2013. http://www-01.ibm.com/software/data/bigdata/ (accessed May 4, 2013).

Issenberg, Sasha. "A More Perfect Untion: How President Obama's campaign used big data to rally individual voters." *MIT Technology Review.* Dec. 16, 2012. http://www.technologyreview.com/featuredstory/508836/how-obama-used-big-data-to-rally-voters-part-1/ (accessed May 20, 2013).

Kennedy, Stephen. "Transforming Big Data into Knowledge: Experimental Techniques in Dynamic Visualization." Massachusetts Institute of Technology, Cambridge, 2012.

Malikova, Alexandra. "MBTA Green Line 3-Car Train Operating Plans to Enhance Capacity and Reliability." Massachusetts Institute of Technology, Cambridge, 2012.

MBTA. "Ridership and Service Statistics." Massachusetts Bay Transportation Authority, Boston, 2010.

MBTA. "Service Delivery Policy." Massachusetts Bay Transportation Authority, Boston, 2010.

MTA. *Key Performance Indicators.* Metropolitan Transportation Authority. 2013. http://www.mta.info/persdashboard/agencies/nyctsubway/cp/392289_indicator_narrative.htm (accessed May 5, 2013).

NECN. "MBTA makes bus data available." *NECN Boston.* New Englang Cable News. Sep. 9, 2010. http://www.necn.com/09/09/10/MBTA-makes-bus-data-available/landing.html?blockID=307317 (accessed May 21, 2013).

Pangilinan, Christopher. "Visualizing Reliable Public Transportation Service." *TRB Visualization Committee.* Jan. 15, 2013. http://teachamerica.com/VIZ13/abj95TRB13Pangilinan/ (accessed May 20, 2013).

Rowley, Jennifer. "The wisdom hierarchy: representations of the DIKW hierarchy." *Journal of Information Science* 33, no. 2 (2007): 163-180.

Schil, Mickaël. "Measuring Journey Time Reliability in London Using Automated Data Collection Systems." Massachusetts Institute of Technology, Cambridge, 2012.

Shireman, Matthew. "Using Automatically Collected Data for Bus Service and Operations Planning." Massachusetts Institute of Technology, Cambridge, 2011.

Sledge, Matt. "NYPD License Plaate Readers Will Be Able to Track Every Car Entering Manhattan." *Huffington Post.* Mar. 13, 2013. http://www.huffingtonpost.com/2013/03/13/nypd-license-plate-readers_n_2869627.html (accessed May 5, 2013).

Spear, Steven, and H. Kent Bowen. "Decoding the DNA of the Toyota Production System." *Harvard Business Review*, Sep.-Oct. 1999: 97-106.

Stopher, Peter, and Stephen Greaves. "Household travel surveys: Where are we going?" *Transportation Research Part A* 41 (2007): 367-381.

TCRP. *TCRP Report 88: A Guidebook for Developing a Transit Performance-Measurement System.* Transit Cooperative Research Program, Washington D.C.: Transportation Research Board, 2003.

Tétreault, Paul, and Ahmed El-Geneidy. "Estimating bus run times for new limited-stop service using archived AVL and APC data." *Transportation Research Part A* (Elsevier) 44 (2010): 390-402.

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.

Weber, Max. "Bureaucracy." In *From Max Weber: Essays in Sociology*, edited by H. H. Gerth and C. Wright Mills, 196-244. New York: Oxford University Press, 1946.

Wilson, Nigel H. M. "Frequency Determination." Massachusetts Institute of Technology, 2011.

—. "Roles for Richer Data and Better Analysis in Improving the Effectiveness of Transit Systems." Lausanne, Sep. 2012.

Wilson, Nigel H. M., and John Attanucci. "Performance of a Single Route." Massachusetts Institute of Technology, 2011.

Appendix A: Sample Existing Single Bus Route OTP Report

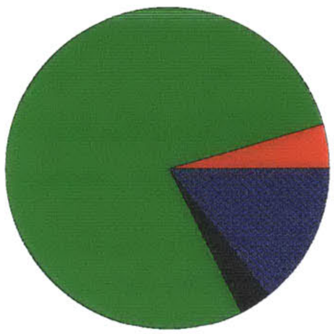
Route OTP Report

Route: 01 **Periods:** One Day: 02/22/2013 (Friday)

Day type: All days

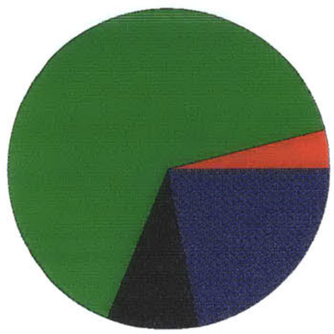
Route		
01	67	1900

Startpoint



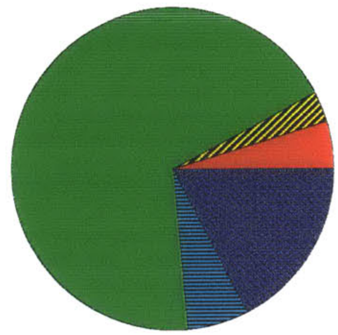
Late	14.4%
Headway gap	3.4%
On time	77.9%
Early	4.3%
Total:	100.0%

Midpoint



Late	21.8%
Headway gap	9.3%
On time	64.8%
Early	4.0%
Total:	100.0%

Endpoint



Late	17.6%
Trip too slow	6.0%
On time	68.5%
Trip too fast	3.2%
Early	4.6%
Total:	100.0%

Inbound by variation

Variation	0% On-Time Performance (%) Goal: 75% 100%	Total known
01-	63	927

Outbound by variation

Variation	0% On-Time Performance (%) Goal: 75% 100%	Total known
01-	70	951

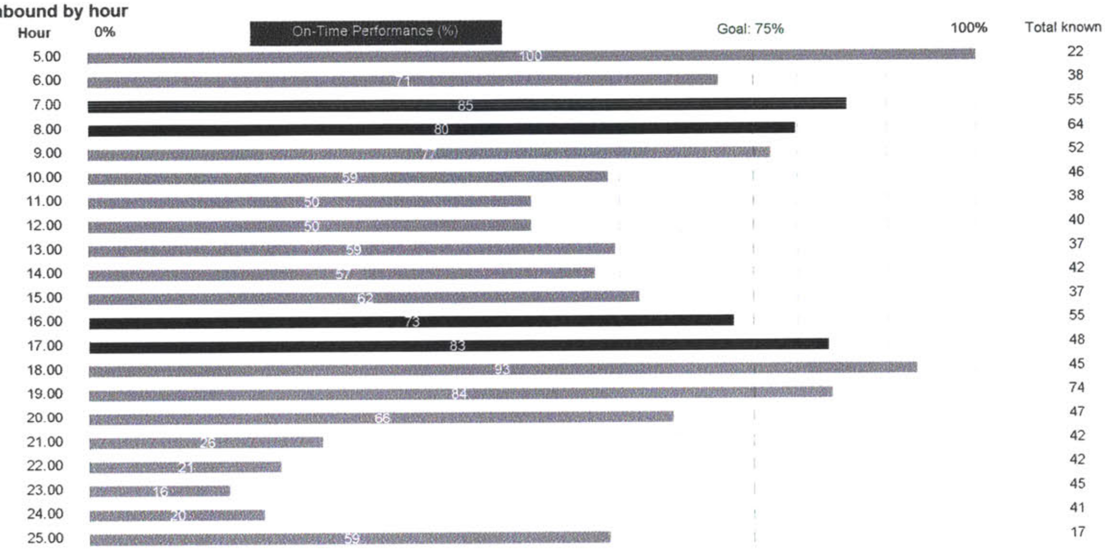
Route OTP Report

Route: 01

Periods: One Day: 02/22/2013 (Friday)

Day type: All days

Inbound by hour



Outbound by hour

Route OTP Report

Route: 01 **Periods: One Day: 02/22/2013 (Friday)**

Day type: All days

Inbound by timepoint

Timepoint	On-Time Performance (%) (0% – 100%, Goal: 75%)	Total known
01: hhgat	80	98
02: maput	75	100
03: cntsq	70	105
04: mit	64	105
05: hynes	58	105
06: masta	54	102
07: Wasma	52	105
08: Melwa	53	103
09: Dudly	61	104

Outbound by timepoint

Timepoint	On-Time Performance (%) (0% – 100%, Goal: 75%)	Total known
01: Dudly	76	106
02: Melwa	83	105
03: Wasma	75	105
04: masta	72	107
05: hynes	69	107
06: mit	63	105
07: cntsq	60	105
08: maput	56	103
09: hhgat	76	108

Printed on 2/25/2013
SmartBusMart

Route OTP Report

Route: 01

Periods: One Day: 02/22/2013 (Friday)

Day type: All days

Cabot runs

0% | On-Time Performance (%) | Goal: 75% | 100%

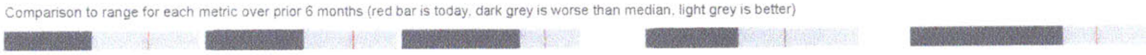
Run	On-Time Performance (%)	Total known
1009	79	63
1012	64	36
1017	76	72
1019	31	36
1023	100	63
1027	77	35
1036	44	72
1053	74	53
1057	94	54
1058	34	53
1070	42	26
1072	89	54
1076	74	54
1079	79	19
1080	98	54
1086	89	36
1096	50	12
1101	31	36
1104	62	45
1109	74	58
1116	76	88
1119	63	79
1124	69	58
1132	57	47
1149	28	36
1156	90	52
1160	6	54
1162	100	20
1167	71	98
1174	74	27
1176	57	72
1184	29	70
1186	48	81
9063	100	24
9065	67	36
9067	75	8
9101	94	34
9106	77	31
9108	97	32

Appendix B: Sample New Performance Reports for Heavy Rail

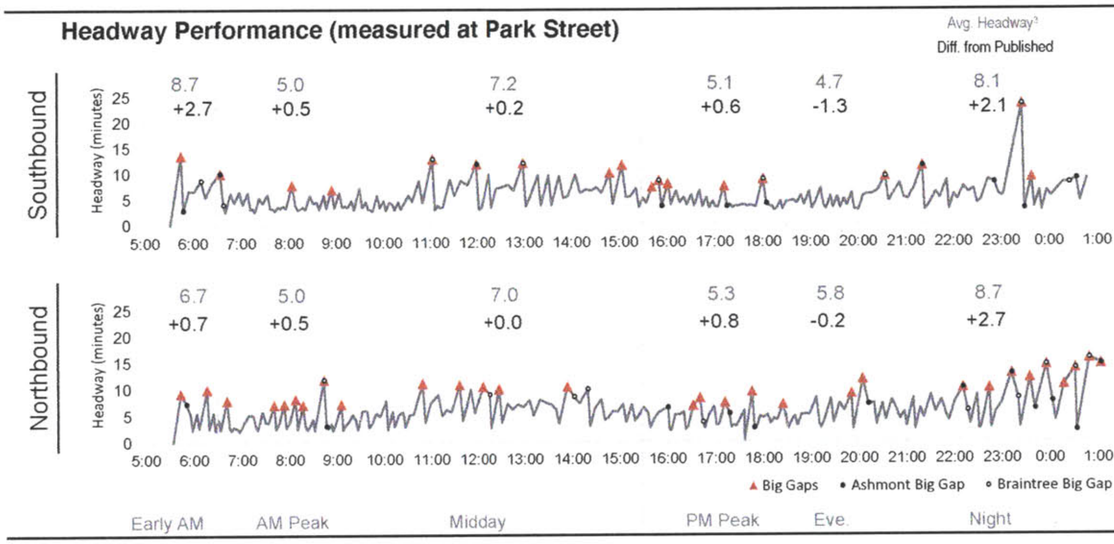
Daily Performance — Red Line — Monday, 05/20/13

	Long Waits		Passenger Travel Time[2]	Passenger Wait Time[2]
11%	**2%**	**0%**	**-3%**	**-2%**
32K pax	7K pax	1K pax	-2.2K hrs	-0.4K hrs
Headway	**Big Gap**	**2X Headway**	**below median**	**below median**

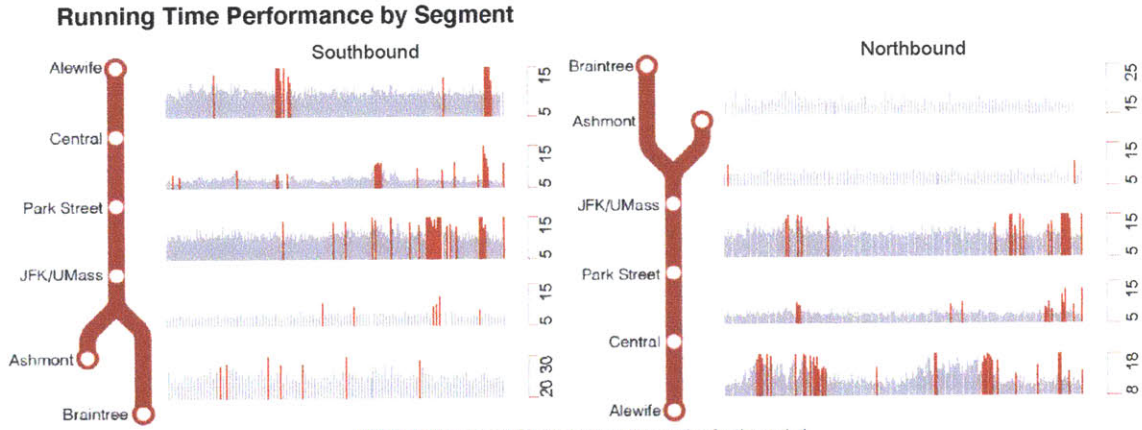
Comparison to range for each metric over prior 6 months (red bar is today, dark grey is worse than median, light grey is better)



Headway Performance (measured at Park Street)



Running Time Performance by Segment



Highlighted times are 15% higher than the median for the period

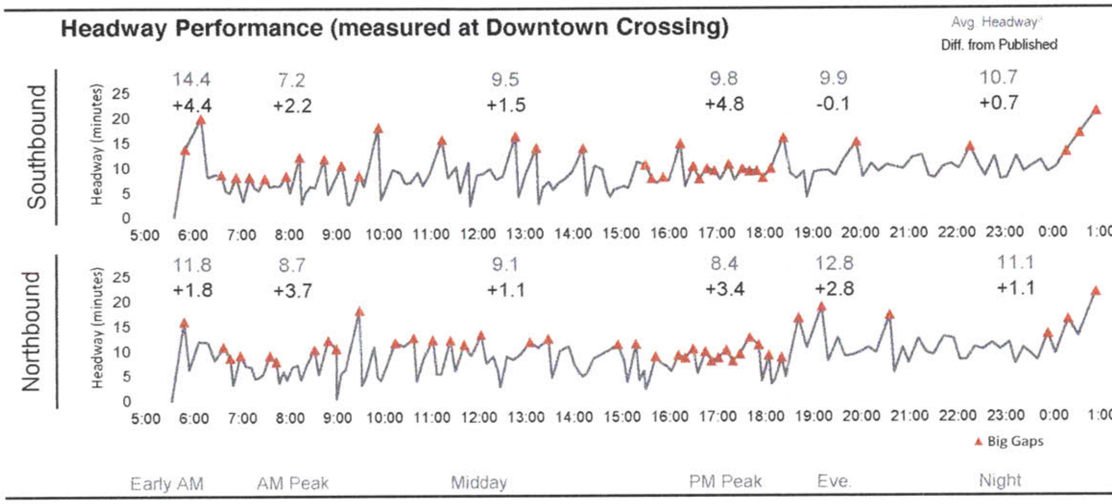
1. The standard for a big gap is either 1.5 times or 3 minutes greater than the scheduled headway, whichever is lower.
2. Passenger travel and wait time are based on average passenger demand rates per period. I.e. 18000 people entering a station during the peak is a demand rate of 6000/hr or 100/min, which are further divided by destination. The rate is multiplied by the headway of a train to get the number of people boarding that train and by its travel time to get passenger travel time. *It does not account for people not being able to board a train due to crowding.*
3. Weighted average headway accounts for the fact that fewer people end up experiencing a short headway than a long headway, since fewer passengers arrive between trains.

Daily Performance — Orange Line — Monday, 05/20/13

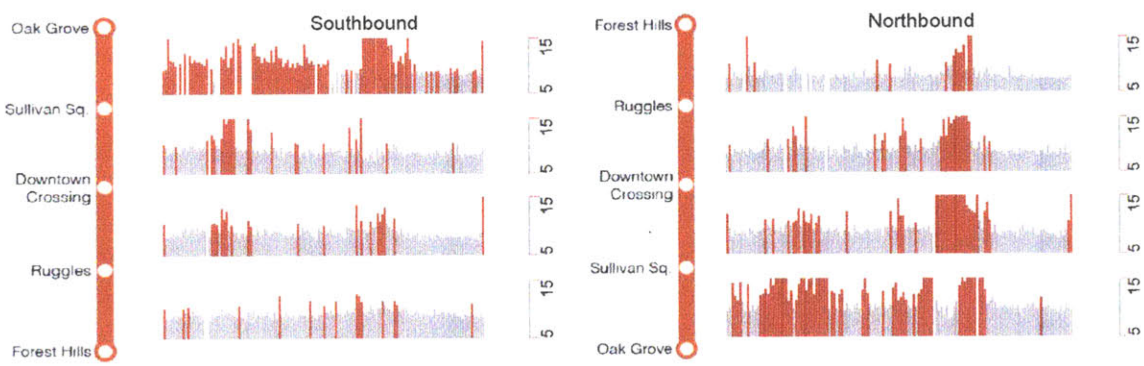
Long Waits			Passenger Travel Time[2]	Passenger Wait Time[2]
25%	9%	2%	+16%	+12%
49K pax	17K pax	4K pax	+6K hrs	+1.6K hrs
Headway	Big Gap	2X Headway	above median	above median

Comparison to range for each metric over prior 6 months (red bar is today, dark grey is worse than median, light grey is better)

Headway Performance (measured at Downtown Crossing)



Running Time Performance by Segment



Highlighted times are 15% higher than the median for the period

1. The standard for a big gap is either 1.5 times or 3 minutes greater than the scheduled headway, whichever is lower.
2. Passenger travel and wait time are based on average passenger demand rates per period. I.e. 18000 people entering a station during the peak is a demand rate of 6000/hr or 100/min, which are further divided by destination. The rate is multiplied by the headway of a train to get the number of people boarding that train and by its travel time to get passenger travel time. *It does not account for people not being able to board a train due to crowding.*
3. Weighted average headway accounts for the fact that fewer people end up experiencing a short headway than a long headway, since fewer passengers arrive between trains.

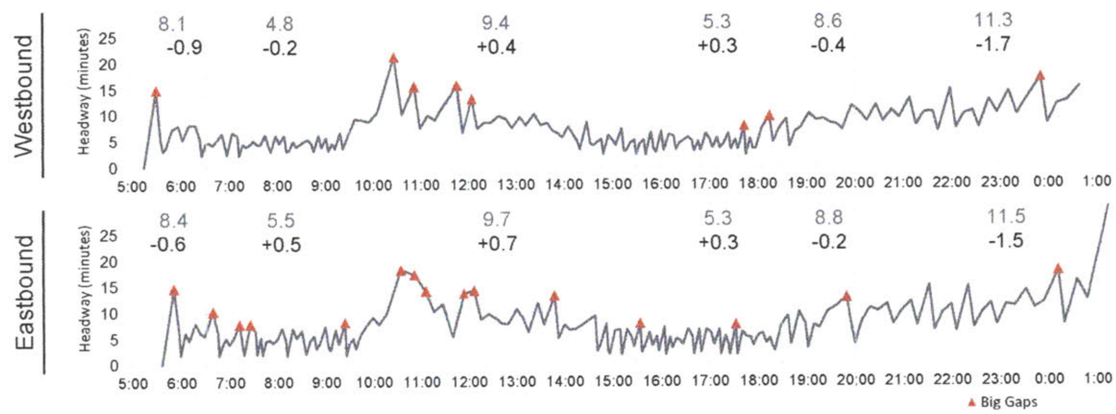
Daily Performance **Blue Line** Monday, 05/20/13

	Long Waits		Passenger Travel Time[2]	Passenger Wait Time[2]
8%	**2%**	**0%**	**+1%**	**+3%**
5K pax	1.3K pax	0.1K pax	+0.1K hrs	+0.1K hrs
Headway	**Big Gap**	**2X Headway**	**above median**	**above median**

Comparison to range for each metric over prior 6 months (red bar is today, dark grey is worse than median, light grey is better)

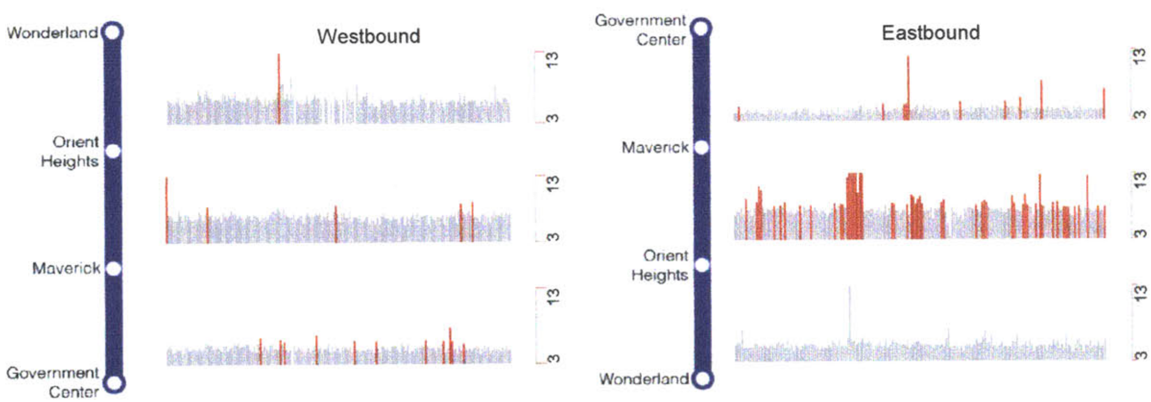
Headway Performance (measured at Airport)

Avg. Headway[3]
Diff. from Published



Early AM AM Peak Midday PM Peak Eve. Night

Running Time Performance by Segment



Highlighted times are 15% higher than the median for the period

1. The standard for a big gap is either 1.5 times or 3 minutes greater than the scheduled headway, whichever is lower.
2. Passenger travel and wait time are based on average passenger demand rates per period. I.e. 18000 people entering a station during the peak is a demand rate of 6000/hr or 100/min, which are further divided by destination. The rate is multiplied by the headway of a train to get the number of people boarding that train and by its travel time to get passenger travel time. *It does not account for people not being able to board a train due to crowding.*
3. Weighted average headway accounts for the fact that fewer people end up experiencing a short headway than a long headway, since fewer passengers arrive between trains.