OCTOBER 2017

HOW RELIABLE IS THE T?

NEEDED IMPROVEMENTS TO THE MBTA'S METHOD FOR MEASURING SUBWAY RELIABILITY ON THE BACK ON TRACK PERFORMANCE DASHBOARD

MASSPIRG Education Fund

How Reliable Is the T?

Needed Improvements to the MBTA's Method for Measuring Subway Reliability on the Back on Track Performance Dashboard

MASSPIRG Education Fund

Matt Casale and
Divya Odayappan

October 2017

About the Authors:

Matt Casale is a Staff Attorney with MASSPIRG Education Fund.

Divya Odayappan is a summer 2017 Public Policy Intern with MASSPIRG Education Fund. Divya is a sophomore at The George Washington University in Washington, DC.

The authors would like to thank Janet Domenitz of MASSPIRG Education Fund, Tony Dutzik of Frontier Group, Rafael Mares of Conservation Law Foundation, and Chris Dempsey, Charlie Ticotsky, and Josh Ostroff of Transportation for Massachusetts for providing review and insights for this report.

The authors bear responsibility for any factual errors. Policy recommendations are those of the Massachusetts Public Interest Research Group Education Fund. The views expressed in this report are those of the authors and do not necessarily reflect the views of our funders.

With public debate around important issues often dominated by special interests pursuing their own narrow agendas, the Massachusetts Public Interest Research Group Education Fund offers an independent voice that works on behalf of the public interest. Massachusetts Public Research Group Education Fund works to protect consumers and promote good government. We investigate problems, craft solutions, educate the public, and offer citizens meaningful opportunities for civic participation. For more information visit http://www.masspirgedfund.org/.

Cover Photo: Ashmont Train on Longfellow Bridge, Wikimedia Commons.

All photos not attributed in the report were taken by the authors.

Table of Contents

Executive Summary

How dependable is the T? Very dependable, according to the MBTA's "Back on Track Performance Dashboard." Not so dependable, according to commuters.

The Back on Track Performance Dashboard is an interactive website run by the MBTA that publicly reports data on four key performance areas: Reliability, Ridership, Customer Satisfaction, and Financials. It was rolled out shortly after the creation of the Fiscal and Management Control Board ("FMCB"), and is intended to provide the public with accessible and easy-to-understand ratings of the T's services.

The Dashboard reports high reliability ratings for the subway and commuter rail and somewhat lower ratings for bus service. Bus service reliability is usually around 65-70 percent. Commuter rail service reliability usually hovers around or just under 90 percent. Subway reliability is typically the highest rated of all three and is usually at or above 90 percent.

Customers, however, do not rate the T's reliability quite so high. The Dashboard reports that in June 2017, despite the Performance Dashboard's report of high reliability percentages, 44 percent of those surveyed disagreed with the statement that "The MBTA Provides Reliable Public Transportation Services." Another 19 percent only slightly agreed. Similarly, a recent poll of Boston area voters from WBUR/MassINC Polling Group found that 47 percent are dissatisfied with MBTA's reliability. Although some deviation between customer perceptions of reliability and the MBTA's metrics are to be expected, this large of a disconnect cannot be ignored.

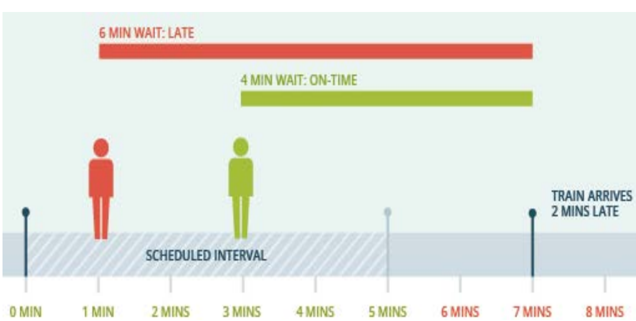
This report takes a critical look at the unique methodology the MBTA uses to measure subway reliability and finds several problems that help explain the gap between the MBTA's rating and customer satisfaction survey data. The report focuses on the subway, because it is the highest rated mode of service according to the MBTA's Dashboard, it is the most widely used mode of service, and because the MBTA uses a unique methodology for measuring subway reliability, which it does not use for either the commuter rail or the bus.

Commuter rail reliability is measured by whether a train arrives to its destination less than five minutes later than it is scheduled to arrive. A reliable bus, when buses are scheduled every fifteen minutes or less, is one that departs and/or arrives no more than three minutes after the scheduled times. When buses are scheduled less frequently, a reliable bus is one that departs and/or arrives no earlier than one minute and no later than six minutes from the scheduled time.

Both of these methods have their own potential issues. Early buses are only counted as unreliable when there is less frequency, but these can still cause problems for riders during times of higher frequencies, especially when buses are scheduled fifteen minutes apart. The commuter rail methodology does not account for cancelled trains or in-station wait times.

This report, however, focuses on the MBTA's methodology for measuring subway reliability. According to the MBTA Performance Dashboard website, reliability of the subway is "the percentage of people who waited the scheduled interval or less." The MBTA calls this metric "Wait Time Reliability."

Subway trains are scheduled on regular intervals, or "headways." A headway is simply a measurement of the distance or time between vehicles in a transit system. Rather than measure subway reliability based on whether a train arrives within the scheduled headway, the MBTA measures reliability based on how long riders are waiting in the station. For example, assuming a five-minute headway, if a rider arrives at the station and waits five minutes or less, then the train is counted as "reliable" for that rider. As the below image demonstrates, this is true even if the train does not arrive within the five-minute headway. As the image also demonstrates, the *same train* in the *same station* can be both reliable and unreliable for the purposes of the MBTA's calculation.



There are several problems with using this methodology to measure subway reliability. First, at a fundamental level, it conflates "headways" with "wait times." Acceptable wait times for reliability purposes are defined based on the headways. However, if the MBTA meets its headway goals, wait times will be significantly lower than the actual headway. For example, assuming a five-minute headway, if riders arrived at the station at the same rate every minute, the average wait time would only be 2.5 minutes.

Second, by only measuring wait time, the MBTA's subway methodology doesn't actually account for on-time performance. A late train can be "reliable." In fact, if in the above illustration riders arrived at the station at the same rate every minute, despite being late, the train would be counted as reliable for five out of every seven riders, and unreliable for only two out of every seven. Moreover, it assumes a reliable trip for riders who only waited a few minutes no matter what happens throughout the duration of their trip. Even if a rider waits a short time to get on a train, her trip still might take twice as long because of service disruptions, backups, or other issues along the way.

Third, the methodology does not account for overcrowding. It assumes that every rider can get on the first train that arrives after he or she gets to the station, which is often not the case during rush hour. This issue is compounded by equating the headways with acceptable wait times. And it is important to also recognize that longer headways themselves lead to more overcrowding, a longer wait time, and a worse customer experience.

To illustrate the problems with the MBTA's Wait Time Reliability metric, this report compiles the text message alerts regarding delays and service interruptions on the Red Line sent out to subscribers between July 14 and July 26, 2017. These alerts are generated using the same real-time data collection used to generate the reliability ratings, but the alerts themselves are not considered by the MBTA's methodology. Over the course of the two weeks, there were over forty alerts on the Red Line alone. The Dashboard reliability rating never dipped below 91 percent. There were days, however, with up to ten alerts. Days that had no alerts, like Sunday, July 23 (93 percent), had the same reliability percentage as days that had ten alerts, like Monday, July 24 (also 93 percent).

The comparison between the alerts and the reliability ratings highlights the problems with the MBTA's methodology. It is confusing and incorporates only a portion of what riders think of when they think of reliability. Most importantly, it is not reflective of the rides customers are experiencing.

The FMCB was created to increase transparency and accountability and to improve customer service. The MBTA has taken steps in the right direction. But a reliability measurement that does not sufficiently reflect the actual reliability of MBTA services is not true transparency. Use of those numbers erodes public trust and confidence in the MBTA. Moreover, the Dashboard should be a tool for that the MBTA can use to identify issues and develop solutions. For it to be an effective tool, it must accurately reflect what is actually happening in the system.

The MBTA's Wait Time Reliability metric has two fundamental issues. First, defining the acceptable wait time as equal to the headway is arbitrary at best and results in late trains that are counted as reliable. This leads to a high reliability rating, even when trains are not adhering to the scheduled headways. The MBTA needs to use a more appropriate metric for acceptable wait times. The metric should reflect that average wait times should be half of the scheduled headway. Second, by only measuring wait times, the MBTA's metric fails to account for delays and service disruptions over the course of the journey, even though these affect riders' overall experiences. A truly comprehensive reliability metric needs to account not only for wait times, but also for on-train time by factoring in delays and service disruptions.

There's sometimes a disconnect between the data that the agencies collect for themselves and the data they present to the riders, and how they present it to the riders. For the most part, most riders don't care about on-time performance, they don't care about headways, they don't care about terminal performance. They want to know how long it's going to take them to get somewhere, when their train or bus is coming, and if there are any problems along the route.

Therefore, this report recommends changes to the reliability measurement reported on the Performance Dashboard that will mitigate some of the current limitations as well as present a metric that is more in line with how the average rider thinks about service. Specifically, the report recommends that the MBTA (1) revise the wait time metric to reflect that average wait times should be half the scheduled headway; (2) incorporate overall travel time, or excess journey time, into the reliability methodology; and (3) not use the metric for policy making decisions until it is further developed.

The MBTA is in a period of reform, and recently hired a General Manager who is focused on the riders' experience and will be hiring a senior level manager dedicated to customer experience. Now is the perfect time for the agency to take a hard look at the Performance Dashboard, and revamp its methodology to more accurately measure what riders are experiencing.

Introduction

The T is a staple in Eastern Massachusetts. Now operated by the Massachusetts Bay Transit Authority (the "MBTA"), Boston's subway was the first of its kind in the United States.[1] In 1897 and 1898, the Tremont Street Subway opened as the precursor to the Green Line and the Main Line Elevated opened in 1901 as the precursor to the Orange Line.[2] Over the years, Boston's transit system developed into one which, on an average weekday, serves over 1 million combined riders on the subway, bus, ferry, and commuter rail.[3] Millions of people now rely on these services to get to and from work and school, to see their friends and families, to get to their doctors and banks, and to go shopping or see a Red Sox game. The MBTA is the backbone of the regional economy.

Never was the importance of this transit system more on display than in the winter of 2015. During one of Boston's snowiest winters on record, the MBTA proved to be wholly unprepared to deal with emergency weather conditions, and was often forced to cancel, suspend, or delay services throughout the winter, leaving people stranded and causing the city to grind to a halt.[4] Following its disastrous performance during the winter of 2015, on April 8 of that year, Governor Charlie Baker appointed a Special Panel, which after 100 days of intensive study, released a detailed report on the MBTA's operations, maintenance and finances, and outlined a plan of action to reform and improve the agency.[5] With the approval of the Legislature, Governor Baker created a five-member Fiscal and Management Control Board (the "FMCB") to enforce new oversight and management support, and to increase accountability.

Soon after the FMCB convened, in February 2016, the MBTA unveiled a new interactive website that, using data, aims to offer an easy-to-digest look at its performance. The website, http://mbtabackontrack.com, houses the MBTA's "Back on Track Performance Dashboard," which is intended to paint a clear picture of how reliable riders' trips are. Noting that the FMCB was created, in part, to improve transparency and accountability, the "About" section of the website says "making our performance information easily and publicly accessible is a key component in [the FMCB's] work. As one of our customers, or just someone interested in the MBTA's performance, you can use this dashboard to track how well we are serving you across a variety of performance metrics."[6] In short, the Performance Dashboard publicly reports the "reliability" of the MBTA's services.

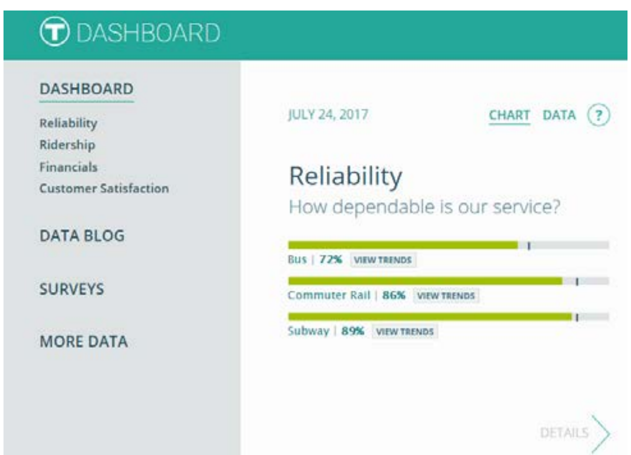
This report examines the MBTA's Back on Track Performance Dashboard, taking a critical look at the way in which it measures reliability for the subway system.[7] The report explains how the MBTA measures reliability and how they collect the data to do so. It also discusses the MBTA's reasoning for how it measures reliability and identifies several problems with its approach.

To illustrate these problems, the report compiles the alerts and delays on the Red Line over a two week period from July 14 through July 26, 2017, and compares the amount, frequency, and severity of the delay notifications with the reported reliability percentages on the Dashboard. Finally, the report offers recommendations on how the MBTA can better reflect customer experience in their reliability measurements

The Back on Track Performance Dashboard

On the Back on Track Performance Dashboard, the MBTA publishes the data for three of the modes of service that make up the MBTA system: the commuter rail, the subway, and the bus. The website is an accessible, colorful, interactive tool to track people's riding experience. Throughout the summer of 2017, the Dashboard reported high reliability ratings for the subway and commuter rail and somewhat lower ratings for the bus. Bus service reliability is usually about 65-70 percent, the commuter rail is usually rated to be around 90 percent reliable, and subway reliability is typically rated as the most reliable of the three, usually at over 90 percent.[8]

Figure 1[9]



Most MBTA riders would likely be surprised at the high reliability ratings, as they have experienced firsthand the *lack* of MBTA reliability. This is reflected in the "Customer Satisfaction" section of the Dashboard, which reports numbers seemingly at odds with the high rates of reliability. For example, for the month of June 2017, despite the Performance Dashboards report of high reliability percentages (hovering around 90 percent), 44 percent of those surveyed disagreed with the statement that "The MBTA Provides Reliable Public Transportation Services." Another 19 percent only slightly agreed. Similarly, a recent poll of Boston area voters from WBUR/MassINC Polling Group found that 47 percent are dissatisfied with MBTA's reliability.[10] Although some deviation between customer perceptions of reliability and the MBTA's metrics are to be expected, this large of a disconnect cannot be discounted.

This section explains how the MBTA measures reliability, how the MBTA collects its data, and discusses some of the explanations for the disconnect between the reliability ratings and the customer satisfaction ratings.

The MBTA's "Wait Time Reliability" Metric

The methods of measuring reliability, for the purpose of the Dashboard, are different for the three major modes of the MBTA's service. For the commuter rail, a reliable train is defined as one that arrives to the final destination less than five minutes later than it is scheduled to arrive.[11] A reliable bus, when buses are scheduled every fifteen minutes or less, is one that departs and/or arrives no more than three minutes after the scheduled times. When buses are scheduled less frequently, a reliable bus is one that departs and/or arrives no earlier than one minute and no later than six minutes from the scheduled time.[12]

While bus and commuter rail reliability are measured somewhat similarly, based on the time the bus or train arrives at its stop,[13] reliability of the subway, which is the most widely used MBTA service,[14] is measured very differently. According to the MBTA Performance Dashboard website, reliability of the subway is "the percentage of people who waited the scheduled interval or less."[15] The MBTA calls this metric "Wait Time Reliability."[16]

To understand this, it is important to understand the way the T is scheduled. Rather than schedule the trains on a timetable — on which a train is scheduled to arrive and depart each station at a specific time — subway trains are scheduled on regular intervals, or "headways." A headway is simply a measurement of the distance or time between vehicles in a transit system. This is a common way for transit systems to schedule subways because it allows for corrections throughout the day without throwing off the entire schedule. The way it works is that each train is supposed to arrive in each station within the headway, or time interval, prescribed by the schedule. For example, if the scheduled headway is nine minutes, as it is on the Red Line during weekday peak hours,[17] a train should be arriving into each station every nine minutes.[18]

Using this system, it should be easy to determine whether a train is on time or late. If the train arrives within the headway, it is on time. If it arrives outside of the headway, it is late. In a post on its "Data Blog," the MBTA used Figure 2 to explain, assuming for the sake of the explanation a five minute headway.[19]

Figure 2[20]

But that's not how the MBTA measures "reliability." On the Back on Track website, the MBTA uses Figure 3 to visually explain its definition of subway reliability, assuming again, a five minute headway.

Figure 3[21]

Under the T's definition of reliability, if the headway is five minutes and a specific passenger waits six minutes, the train is deemed unreliable for that passenger. If a person waits four minutes, the train is reliable for that passenger. So, if passengers are arriving to the station at the same rate every minute during that seven-minute period, even though it is two minutes late, the train will be "reliable" for five out of seven people, and "unreliable" for only two out of seven, meaning late trains contribute significantly to high reliability ratings.

In contrast, if the MBTA made its headway goals for the subway, the average wait time should be half the scheduled headway. If passengers arrive at a steady rate, the average wait time for a subway line where trains came reliably five minutes apart would be 2.5 minutes. In other words, "headway" does not equal "wait time" and the standards for those two should be different. By using this method, the MBTA is incorrectly treating the two as the same.

The MBTA only recently began using the Wait Time Reliability methodology to measure the subway's performance. As noted, the Back on Track Performance Dashboard was launched after the convening of the FMCB in 2016. Prior to the public launch of the Dashboard, the MBTA did not use this method.

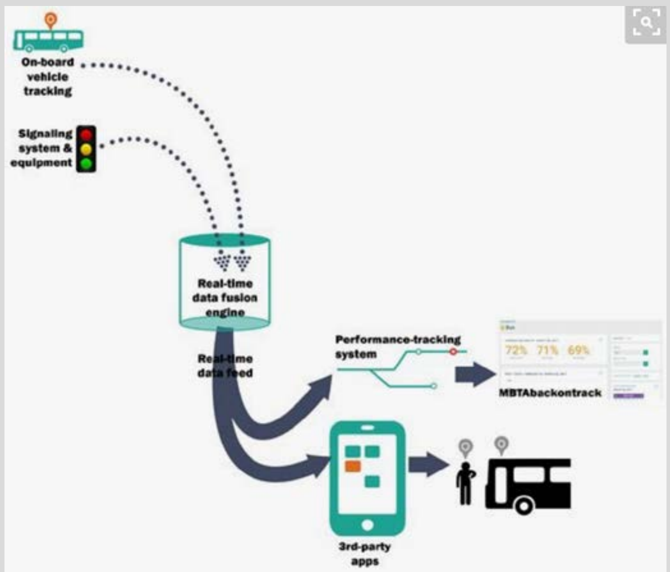
According to a 2015 MBTA presentation, performance was measured by headway adherence. Over the course of 2015, reliability, as measured by headway adherence, hovered around 72 percent.[22] At no point was it above 76 percent and in October 2015 dipped as low as 66 percent.

In other words, when the MBTA began using the Wait Time Reliability method, its reliability performance numbers went way up.[23] Improvements have been made since 2015, but few riders would agree that MBTA reliability has become fifteen to twenty percent better over the past two years.

How the MBTA Collects Data

There are two main software systems involved in the production of the subway reliability metrics shown on the Back on Track Performance Dashboard: the real-time feeds, which produce the vehicle location and arrival prediction data used by app developers and on mbta.com; and the performance tracking system, which uses the real-time feeds as an input and calculates performance metrics. Basically, the MBTA uses "on board vehicle tracking" and "signaling system and equipment" as depicted in the Figure 4, published on MBTA's Data Blog.

Figure 4



Every time a train arrives or departs a station, a record is recorded in a database. From there, the MBTA calculates the time since the last train left the same station and compares it to the scheduled headway. If the actual interval is lower than the scheduled interval, then it assumes that no one who boarded that train waited longer than they were supposed to. However, if the interval between trains is longer than scheduled, it then estimates how many people had an excessive wait time. The estimates are determined by a piece of software developed by researchers at MIT called the ODX model, which stands for Origin, Destination, Transfer. These rates also vary by time of day.

If the arrival rate at a station at a certain time of day is 20 riders per minute, the MBTA can estimate that a two-minute delay caused 40 people to wait too long. Doing this for every train at every station over the course of the day produces an estimate of people with excess waits, which the MBTA divides by the total number of people riding the line to get the percent who waited too long. Subtracting that from 100 is the percent of people with acceptable waits, resulting in the performance metric.

Source: MBTA Data Blog

Problems with the MBTA's Wait Time Reliability Method

The MBTA employs a technical and data-driven method for determining reliability of the subway. Why go through all that trouble? The MBTA is trying to reflect the riders' experiences on the T by weighting performance by the number of riders affected. This is appropriate, because people arrive at different times between trains, and are therefore not equally affected by gaps in service, while there are variations in demand from station to station and hour to hour over the course of the day. Weighting is supposed to result in a more comprehensive and nuanced reliability metric for subway service, according to the MBTA.

But is that what is actually happening? Riders don't necessarily feel that the numbers being reported under this method reflect their experiences. On the Back on Track website, the MBTA also reports the results of a Customer Opinion Panel comprised of several thousand public transit riders who have chosen to share their experiences and feedback with the MBTA once every few months, although this data is *not* used in the determination of reliability. Through this group of riders, the MBTA obtains data regarding their customer satisfaction. For example, for the month of June 2017, the MBTA reports that, despite the Performance Dashboard's report of high reliability percentages (hovering around 90 percent), 44 percent of customers surveyed thought the MBTA did not provide reliable services. Similarly, a recent poll of Boston area voters from WBUR/MassINC Polling Group found that 47 percent are dissatisfied with MBTA's reliability.[24]

The first problem, and one fundamental to the issue, is that the MBTA's method for determining reliability conflates "headways" and "wait time." It assumes that a five-minute headway means riders should not wait more than five minutes. But as noted above, a truly reliable five-minute headway results in a 2.5 minute average wait time

Defining the acceptable wait time as the headway is arbitrary, at best. Why is it acceptable for someone to wait the longest possible time for a train when with regular headways the average wait time should be half the headway?

Second, the MBTA's method doesn't measure on-time performance. In fact, using this method, as demonstrated in Figure 2, the same train could theoretically be both on-time *and* late. But if a train arrives late, it is late, *regardless* of when the rider arrives or how long a passenger is waiting. A rider could conveniently time his or her arrival at a station just two minutes before the scheduled train to save time. But the MBTA's method says it's still okay for them to wait up to five minutes for a train, which could also impact how crowded a train and train station is.

Measuring reliability based on station wait time only, rather than incorporating on-time performance, often means that major service disruptions are not considered by the methodology. A train can be delayed while a rider is on it, and thus can make a person late, or in some cases, *extremely* late, to his or her destination. It doesn't consider situations in which riders are forced to get on a shuttle bus because their station is closed, or when riders have to deboard early and walk or take a bus to get to their destination. If the train arrives within the scheduled headway for a certain passenger, the train is counted as reliable, even if the duration of the trip is much longer than it should be.

Indeed, authors of the MBTA's Data Blog themselves have acknowledged that this is a problematic limitation of the MBTA's Wait Time Reliability Method. "This methodology," the Data Blog authors wrote in March 2016, "does not currently account for major disruptions or diversions where stations are not served at all."[25] The authors claimed at the time that MBTA staff were working with modelers to address the limitation, but more than a year later no changes have been made.

The MBTA's methodology also doesn't consider when people are required to wait for a second train because the first one is full. The MBTA's method assumes that everyone can board the same train, which we know is not necessarily true when there are large gaps during high demand periods. Even if a passenger enters the station less than five minutes before a train arrives, there is no guarantee that he or she will actually be able to board that train. During rush hour on the Red Line, it is not uncommon for riders to have to wait for two to three trains to come through the station before being able to board. If a passenger is unable to board the first train, the ultimate wait time skyrockets, but that is not considered by the MBTA's reliability method. This too has been acknowledged by writers at MBTA's Data Blog.[26]

Wikimedia Commons

The Industry Standard: London's Journey Time Metric

publicdomainpictures.net

Many experts consider the performance measurement system used in London to be the industry standard.[27] For the London Underground, or the London subway system, Transport for London uses a system called the "Journey Time Metric."[28] The Journey Time Metric breaks the journey into four parts: (1) Access, Egress and Interchange Time, (2) Ticket Purchase Time, (3) Platform Wait Time, and (4) On-Train Time. In addition to these components, the Journey Time Metric also considers the effect of line and station closures on customers' journey times.[29]

Access, egress, and interchange time evaluates the walking time required to enter and exit the station and to transfer between two lines.[30] Ticket purchase time is the sum of the queuing time and the transaction time at the ticket office window or the automated ticket machines. The queuing time is surveyed on a regular basis; transaction times are recorded at all windows and for some ticket machines at the busiest stations.[31]

Platform wait time is calculated as the time between the customer's arrival at the mid-point of the platform and the moment the boarded train departs.[32] For the ultimate purposes of determining reliability, scheduled passenger wait time, or acceptable wait time, at a station is calculated as half the scheduled headway for the corresponding section of line.[33]

On-train time is calculated from the moment the train departs the origin station to the moment doors open at the destination station.[34] The scheduled on-train time is calculated from the operating timetable.[35] The actual on-train time is measured using data from the signaling system when available.[36]

The Journey Time Metric also takes into account the effect of closures and disruptions due to incidents and engineering work.[37]

Ultimately, for each of the components, a value is estimated based on schedule, to reflect how long the journey would take if there were no disruptions.[38] The Journey Time Metric then compares the actual journey times to the scheduled ones. The difference between the two is referred to as Excess Journey Time (EJT).[39] EJT is used as an indicator of the journey time reliability in the Underground.[40]

Transit authorities across the United States are beginning to develop systems that seek to follow London's example.

For example, the Metropolitan Transit Authority ("MTA") in New York City recently announced a new system for both tracking delays on the subway as well as informing the public about the reasons for the problems.[41] MTA is unable to fully adopt London's system because, unlike London, where riders pay based on the distance traveled and therefore leave a record of their entire trip, in New York, like in Boston, there is no information on when or where a rider leaves the system. Still, New York's new system measures (1) the on-time performance of individual lines, (2) wait times on platforms, and (3) major incidents that result in a delay that affects at least 50 trains. At some stations in Manhattan at certain points throughout the day, more than 50 trains pass through in an hour. According to MTA, a delay of more than 50 trains results in the type of delay where riders have to "call home because [they] are going to be late for dinner."[42]

The Red Line: July 14 - July 26, 2017

The MBTA uses the same real-time data that it uses to measure reliability to also power an alert system that sends out text messages to subscribing riders with information pertaining to delays, construction, or general updates in T service.[43]

Figure 5[44]

The alerts are designed to inform riders about delays and disruptions in service — in other words, factors that will affect the reliability of their trip.

Wikimedia Commons

As described above, however, these delays are not directly taken into account in the MBTA's reliability determination.[45] In fact, some of these alerts, like major disruptions resulting due to the closing of a station and use of a shuttle bus, aren't taken into account at all.

Though they are not directly taken into account in the MBTA's reliability determination, these disruptions of service directly affect riders' experiences.

Table 1 compiles the Dashboard reliability relating and the MBTA T-Alerts for the Red Line over a two-week period between July 14 and July 26, 2017.

Table 1

Date	Day of Week	Dashboard Reliability Percentage	Alert and Time
7/14/17	Friday	92%	*8:38 am*: Red Line experiencing minor northbound delays due to a medical emergency at Andrew. *2:31 pm*: Red Line experiencing moderate delays due to a disabled train at JFK/UMASS station. *6:38 pm*: Red Line experiencing moderate delays due to an ill passenger on a train at Charles/MGH.
7/15/17	Saturday	93%	*6:09 pm*: Red Line experiencing minor delays due to a disabled train at Davis (southbound).
7/16/17	Sunday	93%	none
7/17/17	Monday	91%	*5:19 am*: Shuttle buses replacing Red Line service between Braintree and North Quincy due to a track problem. *5:36 am*: (update) Red Line will have a delayed start from Braintree to North Quincy. Shuttle buses will be utilized for the start of service. *6:13 am:* (update) Shuttle buses will be replacing Red Line northbound service from Braintree to JFK/UMASS due to a disabled piece of work equipment. *7:02 am*: Regular service has resumed between Braintree and JFK/UMASS. Expect residual delays in service. *7:11 am*: Red Line experiencing moderate delays in service. *8:09 am*: Red Line experiencing minor northbound delays from Ashmont to JFK/UMASS. *10:46 pm*: Red Line experiencing moderate southbound delays due to a disabled train at Harvard. *11:06 pm*: (update) Red Line experiencing minor southbound delays due to an earlier disabled train at Harvard.
7/18/17	Tuesday	91%	*7:51 am*: Red Line (Braintree branch) trains will experience minor northbound delays from Quincy Center to Wollaston due to a speed restriction.

			9:02 am: Red Line (Braintree branch) trains will experience minor northbound delays from Quincy Center to Wollaston due to a speed restriction. *3:30 pm*: Red Line experiencing minor delays due to a signal problem at Andrew. *4:03 pm*: Red Line experiencing moderate delays due to a signal problem at Andrew. *4:05 pm*: Red Line (Braintree branch) trains will experience minor northbound delays from Quincy Center to Wollaston due to a speed restriction. *5:19 pm*: (update) Red Line experiencing moderate delays due to an ongoing signal problem at Andrew. *6:32 pm*: Red Line (Braintree branch) trains will experience minor northbound delays from Quincy Center to Wollaston due to a speed restriction.
7/19/17	Wednesday	91%	*4:18 pm*: Red Line experiencing moderate delays to a disabled train at Porter.
7/20/17	Thursday	92%	*1:14 pm*: Red Line experiencing severe delays due to a medical emergency at JFK/UMASS. *1:28 pm*: Shuttle buses replacing Red Line Braintree branch service between JFK/UMASS and North Quincy stations. Please expect delays as buses are sent. *1:58* pm: All clear (re: Shuttle buses replacing Red Line Braintree branch service between JFK/UMASS and North Quincy stations. Please expect delays as buses are sent.) *1:58 pm*: Red Line (Braintree branch) trains experiencing moderate residual delays due to an earlier medical emergency at JFK/UMASS.
7/21/17	Friday	93%	*10:21 pm*: (update) Red Line experiencing moderate delays due to an earlier disabled train. *11:02 pm*: Red Line experiencing moderate delays due to a disabled train at Davis. *11:02 pm:* (update) Red Line experiencing severe delays due to a disabled train at Davis.
7/22/17	Saturday	95%	none
7/23/17	Sunday	93%	none

7/24/17	Monday	93%	*8:38 am*: Red Line experiencing minor northbound delays from Ashmont due to an earlier disabled train. *9:01 am*: Red Line experiencing minor northbound delays from Ashmont due to an earlier disabled train. *5:43 pm*: Red Line experiencing moderate delays due to a disabled train at Alewife. *6:01 pm*: (update) Red Line experiencing minor delays due to an earlier disabled train at Alewife. *6:13 pm*: (update) Red Line experiencing moderate delays due to an earlier disabled train at Central (northbound) *6:31 pm*: Red Line experiencing moderate delays due to an earlier disabled train at Central (northbound) *6:36 pm*: (update) Red Line experiencing minor delays due to an earlier disabled train at Davis. *7:58 pm*: Red Line experiencing moderate southbound delays due to an earlier disabled train at Davis. *8:33 pm*: (update) Red Line experiencing minor southbound delays due to an earlier disabled train at Davis. *10:38 pm*: Red Line experiencing minor delays due to a disabled train at Ashmont.
7/25/17	Tuesday	94%	*3:42 pm*: Red Line experiencing minor delays due to police action at JFK/UMASS. *4:02 pm:* Red Line experiencing minor delays due to police action at JFK/UMASS.
7/26/17	Wednesday	92%	*5:38 am*: Red Line experiencing minor northbound delays due to a signal problem at Andrew. *7:25 am*: Red Line experiencing minor southbound delays due to a disabled train.

Calculated statistics:

Average reliability: 92.5%
Average Weekday reliability: 92.1%
Average # of alerts per weekday: 4.44

Over the course of the two-week period, there were over forty alerts on the Red Line alone. Most reported either minor or moderate delays to service, and several reported that shuttle buses would be replacing train service at certain stations for varying amounts of time. Some of the delays were caused by medical emergencies or police action, but most were due to signal problems or disabled trains.

Throughout the entire two-week period, the Dashboard reliability rating never dipped below 91 percent. There were days, however, with up to ten alerts. Days that had no alerts, like Sunday, July 23 (93 percent), had the same reliability percentage as days that had ten alerts, like Monday, July 24 (also 93 percent).

Riders need to know what to expect, and what level of service the MBTA is committed to deliver. The current measure is confusing and incorporates only a portion of what riders think of when they think of reliability. Although it is meant to measure each individual rider's experience, the MBTA's method doesn't do a good job of it. Many riders who experienced the delays throughout this two-week period on the Red Line may or may not have waited overly long to get on the train, but at the very least their trip was still longer than it should have been.

Recommendations

The FMCB was created to increase transparency and accountability and to improve customer service. As the FMCB itself has noted, "making performance information easily and publicly accessible is a key component in [the FMCB's] work."[46] Massachusetts deserves a reliable, well-managed, and cost-effective transportation system. Lack of transparency leads to a lack of accountability. A lack of accountability hurts the ability to focus on customer experience.

The MBTA, under the guidance of the FMCB, has made significant strides in the past two years, but there are still improvements to be made. A reliability measurement that does not sufficiently reflect the actual reliability of MBTA services is not true transparency. The Dashboard, however, gives the appearance of transparency. The full picture of how reliability is calculated is buried deeply into the website, and for some information, on the separate MBTA "Data Blog."

Moreover, use of the numbers obtained using the MBTA's method erodes public confidence and trust in the MBTA. In fact, using a reliability measurement that does not sufficiently reflect the actual reliability of MBTA services can be perceived as deceptive. And as is clear from the Customer Satisfaction section of the Dashboard, riders don't find the subway to be as reliable as the Dashboard reports. That is not surprising, given the factors that are not considered in the MBTA's reliability determination. But the whole purpose of the Back on Track website is to increase accountability, customer focus, and public trust. The MBTA is going to have trouble meeting those goals unless it can find a way to better measure reliability, such as reporting excess journey time.

Finally, the Dashboard should be a tool that the MBTA uses to identify issues and develop solutions. In order for it to be an effective tool, it has to accurately reflect what is actually happening on the system. If reliability percentages are inflated because the current methodology is not accurate, then it will be hard for the MBTA to use the Dashboard to identify and implement needed service improvements.

Simply by having the Performance Dashboard and providing detailed, regularly updated performance metrics, the MBTA is ahead of a lot of other transit agencies. The platform that the MBTA has developed for reporting performance metrics deserves commendation.

Still, no matter how good the platform used to report performance metrics is, it does no good if the system for determining those metrics is flawed. The methodology the MBTA uses to determine reliability percentages that are reported on the Performance Dashboard, the Wait Time Reliability metric, contains two fundamental issues. First, defining the acceptable wait time as equal to the headway is arbitrary at best and results in late trains that are counted as reliable. This leads to a high reliability rating, even when trains are not adhering to the scheduled headways. The MBTA needs to use a more appropriate metric for measuring wait times. A logical choice would be to use half the headway as the scheduled average wait time, and then measure actual average wait times against that. Second, by only measuring wait times, the MBTA's metric fails to account for on-time performance or delays and service disruptions over the course of the journey, even though these things affect riders' overall experiences. A truly comprehensive reliability metric needs to account not only for wait times, but also for on-train time by factoring in delays and service disruptions.

Both issues can be remedied by following London's lead.

Specifically, the MBTA should:

1. **Revise the Wait Time Metric**

The MBTA should no longer define the acceptable wait time as equal to the scheduled headway. Instead, the MBTA should follow London's example and design the wait time metric to reflect that *average* wait times should be half the scheduled headway.

Allowing for acceptable wait times that are longer than what the average wait time should be creates a reliability metric that skews results in favor of "reliable," even if trains are not meeting headway goals. When trains are meeting headway goals, meeting the wait time goals should not be a problem. But when trains are not meeting headway goals, average wait times will be above what they should be, and that should be reflected in the reliability metric.

Rather than define an acceptable wait time based on the headway, the MBTA should use half the headway as the scheduled passenger wait time. Then, the actual average passenger wait time could be calculated and compared against the scheduled passenger wait time.[47]

The MBTA should also factor in additional platform wait time for "left behinds," or those who are not able to board the first train to arrive at a station because of overcrowding. The calculation of this additional time can be based on demand levels and the regularity of train service.

2. **Incorporate Overall Travel Time into the Calculation of the Reliability Metric**

Wait Time Reliability, even using a more appropriate acceptable wait time, is an insufficient metric to accurately measure reliability. The MBTA should develop a metric that not only measures wait time, but also incorporates overall travel time, or excess journey time.

The best way to do so would be to develop a metric that measures excess journey time, like used in London. Unlike in London, Boston riders do not "tap-out" when the leave the system, making it somewhat more difficult, but not impossible, to calculate overall travel time. The MBTA already has a subway travel time measure that it uses in its Annual Performance Report (Tracker),[48] but it is not used on the Performance Dashboard.

Given the customer satisfaction survey results, it is apparent that the current Performance Dashboard metric does not measure what riders consider to be "good service."

There's sometimes a disconnect between the data that the agencies collect for themselves and the data they present to the riders, and how they present it to the riders. For the most part, most riders don't care about on-time performance, they don't care about headways, they don't care about terminal performance. They want to know how long it's going to take them to get somewhere, when their train or bus is coming, and if there are any problems along the route.

An excess journey time metric would mitigate many of the limitations of the current metric reported on the Performance Dashboard. It would account for service disruptions, in route delays, and slowed down service. It also would better reflect how the average rider thinks about subway reliability because it measures overall trip time, and how much longer the trip took than it should have. That's what riders think of when they think about reliability. This will not only portray a more accurate picture of the level of service being provided, as well as the improvement needs of the system (be they capital or operational), but it will build credibility with riders, stakeholders, and advocates. To riders, a reliability metric that does not reflect their actual experience seems disingenuous and can undermine the MBTA's statements about the state of the system and its needs.

If the lack of "tap-out" proves to make an excess journey time metric unworkable, at the very least, the MBTA should follow New York's lead and develop a reliability metric that considers wait time, on-time performance and service disruptions – in other words, the factors that affect overall journey time.

3. **Refrain from Using the Reliability Metric in Policy Decisions Until It Is Further Developed**

The problems with the MBTA's current Wait Time Reliability Metric result in a performance measurement that does not accurately reflect what is actually happening in the system or the riders' everyday experiences. Therefore, any policy decisions made using the performance measurement as a basis will be problematic. The MBTA should refrain from using the measurement in policy decisions until it is further developed as recommended in this report.

Conclusion

The FMCB was created to increase transparency and accountability and to improve customer service. A reliability measurement that does not sufficiently reflect the actual reliability of MBTA services is not true transparency. Use of those numbers erodes public trust and confidence in the MBTA. Moreover, the Dashboard should be a tool for that the MBTA can use to identify issues and develop solutions. For it to be an effective tool, it must accurately reflect what is actually happening in the system.

The MBTA is in a period of reform and recently hired a General Manager who is focused on the riders' experience.[49] Now is the perfect time for it to take a hard look at the Performance Dashboard and revamp its methodology.

None of this is to suggest that the MBTA discontinue the Performance Dashboard, or stop reporting reliability numbers. It should, however, do a better job clarifying upfront how the numbers are calculated and identifying the potential issues. It should also work towards developing a metric that is able to consider major service disruptions, closures, and other problems that affect riders' experiences and reasonable wait times (or actual headways). Finally, the MBTA should not use these numbers for policy-making purposes until the methodology is further developed to truly capture the full experience of riding the T.

Notes

[1] *Famous Firsts in Massachusetts*, Secretary of the Commonwealth of Massachusetts, *accessed at* https://www.sec.state.ma.us/cis/cismaf/mf4.htm.

[2] *The Chronicle of the Boston Transit System*, MBTA, *accessed at* http://www.mbta.com/about_the_mbta/history/.

[3] *Ridership: How Many Trips are Taken on an Average Weekday?* T Dashboard, *accessed at* http://mbtabackontrack.com/performance/index.html#/detail/ridership///.

[4] Jack Lepiarz, *With Widespread Delays, MBTA's Long-Standing Issues Come Into Focus*, WBUR, Feb. 4, 2015, *accessed at* http://www.wbur.org/news/2015/02/04/mbta-delays-maintenance-money.

[5] *Back on Track: An Action Plan to Transform the MBTA*, MBTA, *accessed at* http://www.mbta.com/about_the_mbta/MBTABackOnTrack/.

[6] *Welcome to the interactive MBTA Performance Dashboard*, T Dashboard, *accessed at* http://mbtabackontrack.com/performance/index.html#/about.

[7] Although there is a short discussion of reliability measurements of the bus and commuter rail, the focus of this report is the subway.

[8] *Reliability: How Dependable Is Our Service?*, T Dashboard, *accessed at* http://mbtabackontrack.com/performance/index.html#/home. For data on average reliability, information on past reliability data, and trends, click on the "View Trends" button next to the modes of service.

[9] Figure 1 is a screenshot of the "Reliability" section of the T Dashboard taken on July 24, 2017 accessed at http://mbtabackontrack.com/performance/index.html#/home.

[10] Gintautas Dumcius, *Boston voters aren't happy with traffic and the MBTA, new WBUR/MassINC poll says*, MassLive, October 4, 2017, accessed at http://www.masslive.com/news/boston/index.ssf/2017/10/boston_voters_arent_happy_with.html.

[11] *Reliability: How Dependable Is Our Service?*, T Dashboard, *accessed at* http://mbtabackontrack.com/performance/index.html#/home.

[12] *Reliability: How Dependable Is Our Service?*, T Dashboard, *accessed at* http://mbtabackontrack.com/performance/index.html#/home.

[13] To be sure, this is not to suggest that the MBTA metrics for reporting commuter rail and bus reliability are perfect. This report focuses on the subway methodology, but there are questions that need to be raised about the other methodologies as well. For example, related to the bus methodology: Why is a bus arriving early only an issue for the buses with less frequency? If a rider misses her bus that is scheduled to arrive every 15 minutes because it is two minutes early, she ends up waiting 15 minutes, when she was expecting not to wait at all. She would likely consider that unreliable. For the commuter rail, the MBTA measures on-time performance by how long it takes a train to reach its final destination, but most riders do not ride from end to end. It does not account for cancelled trains or trains taken out of service, nor does it account for in-station wait time at all.

[14] According to the T Dashboard, around 60 percent of weekday MBTA trips are on the T. *Ridership: How many trips are taken on MBTA services on an average weekday?*, T Dashboard, *accessed at* http://mbtabackontrack.com/performance/index.html#/detail/ridership///.

[15] *How Are We Measuring Reliability: Subway*, T Dashboard, *accessed at* http://mbtabackontrack.com/performance/index.html#/explanations#reliabilitysubway.

[16] *Explaining Dashboard Metrics: Subway Reliability*, MBTA Data Blog, *accessed at* http://mbtabackontrack.com/blog/36-subway-reliability.

[17] This is nine minutes for each branch of the Red Line, so the central tunnel *should* have a train come through every four and a half minutes at weekday peak.

[18] *Rapid Transit/Key Bus Route Map*, MBTA, *accessed at* https://www.mbta.com/uploadedfiles/Documents/Schedules_and_Maps/Subway/frequency-schedule.pdf.

[19] *Explaining Dashboard Metrics: Subway Reliability*, MBTA Data Blog, *accessed at* http://mbtabackontrack.com/blog/36-subway-reliability.

[20] *Explaining Dashboard Metrics: Subway Reliability*, MBTA Data Blog, *accessed at* http://mbtabackontrack.com/blog/36-subway-reliability.

[21] *How Are We Measuring Reliability: Subway*, T Dashboard, *accessed at* http://mbtabackontrack.com/performance/index.html#/explanations#reliabilitysubway.

[22] *State of the Service: Red Line Heavy Rail*, MBTA (January 25, 2016), accessed at https://www.mbta.com/uploadedfiles/About_the_T/Board_Meetings/StateOfTheRedLine01252016.pdf.

[23] The website began reporting reliability data for the subway in March 2016. In the first two months reliability was reported as being in the upper eighty percent range, around ten points higher than the highest percentage reported in 2015 under the headway adherence method.

[24] Gintautas Dumcius, *Boston voters aren't happy with traffic and the MBTA, new WBUR/MassINC poll says*, MassLive, October 4, 2017, accessed at http://www.masslive.com/news/boston/index.ssf/2017/10/boston_voters_arent_happy_with.html.

[25] *Explaining Dashboard Metrics: Subway Reliability*, MBTA Data Blog, *accessed at* http://mbtabackontrack.com/blog/36-subway-reliability.

[26] *Explaining Dashboard Metrics: Subway Reliability*, MBTA Data Blog, *accessed at* http://mbtabackontrack.com/blog/36-subway-reliability.

[27] Mickael Schil, *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*, Massachusetts Institute of Technology, May 21, 2012, accessed at https://pdfs.semanticscholar.org/9d93/fe1a0c600a7952f1d8aa998efcfdaa95704f.pdf.

[28] Mickael Schil, *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*, Massachusetts Institute of Technology, May 21, 2012, accessed at https://pdfs.semanticscholar.org/9d93/fe1a0c600a7952f1d8aa998efcfdaa95704f.pdf.

[29] Mickael Schil, *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*, Massachusetts Institute of Technology, May 21, 2012, accessed at https://pdfs.semanticscholar.org/9d93/fe1a0c600a7952f1d8aa998efcfdaa95704f.pdf.

[30] Mickael Schil, *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*, Massachusetts Institute of Technology, May 21, 2012, accessed at https://pdfs.semanticscholar.org/9d93/fe1a0c600a7952f1d8aa998efcfdaa95704f.pdf.

[31] Mickael Schil, *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*, Massachusetts Institute of Technology, May 21, 2012, accessed at https://pdfs.semanticscholar.org/9d93/fe1a0c600a7952f1d8aa998efcfdaa95704f.pdf.

[32] Mickael Schil, *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*, Massachusetts Institute of Technology, May 21, 2012, accessed at https://pdfs.semanticscholar.org/9d93/fe1a0c600a7952f1d8aa998efcfdaa95704f.pdf.

[33] Mickael Schil, *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*, Massachusetts Institute of Technology, May 21, 2012, accessed at https://pdfs.semanticscholar.org/9d93/fe1a0c600a7952f1d8aa998efcfdaa95704f.pdf.

[34] Mickael Schil, *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*, Massachusetts Institute of Technology, May 21, 2012, accessed at https://pdfs.semanticscholar.org/9d93/fe1a0c600a7952f1d8aa998efcfdaa95704f.pdf.

[35] Mickael Schil, *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*, Massachusetts Institute of Technology, May 21, 2012, accessed at https://pdfs.semanticscholar.org/9d93/fe1a0c600a7952f1d8aa998efcfdaa95704f.pdf.

[36] Mickael Schil, *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*, Massachusetts Institute of Technology, May 21, 2012, accessed at https://pdfs.semanticscholar.org/9d93/fe1a0c600a7952f1d8aa998efcfdaa95704f.pdf.

[37] Mickael Schil, *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*, Massachusetts Institute of Technology, May 21, 2012, accessed at https://pdfs.semanticscholar.org/9d93/fe1a0c600a7952f1d8aa998efcfdaa95704f.pdf.

[38] Mickael Schil, *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*, Massachusetts Institute of Technology, May 21, 2012, accessed at https://pdfs.semanticscholar.org/9d93/fe1a0c600a7952f1d8aa998efcfdaa95704f.pdf.

[39] Mickael Schil, *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*, Massachusetts Institute of Technology, May 21, 2012, accessed at https://pdfs.semanticscholar.org/9d93/fe1a0c600a7952f1d8aa998efcfdaa95704f.pdf.

[40] Mickael Schil, *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*, Massachusetts Institute of Technology, May 21, 2012, accessed at https://pdfs.semanticscholar.org/9d93/fe1a0c600a7952f1d8aa998efcfdaa95704f.pdf.

[41] Marc Santora, *How Much Time Was Spent on Subway? M.T.A. Can Now Tell You*, New York Times, September 25, 2017, accessed at https://www.nytimes.com/2017/09/25/nyregion/how-much-time-was-spent-on-subway-mta-can-now-tell-you.html?_r=1&mtrref=service.meltwaternews.com.

[42] Marc Santora, *How Much Time Was Spent on Subway? M.T.A. Can Now Tell You*, New York Times, September 25, 2017, accessed at https://www.nytimes.com/2017/09/25/nyregion/how-much-time-was-spent-on-subway-mta-can-now-tell-you.html?_r=1&mtrref=service.meltwaternews.com.

[43] The MBTA also makes the data available to third-party developers interested in building apps, tools, and other services that make riding the T easier. Several developers have taken advantage of the resource to create apps or tools that allow riders to view the real-time data in easy-to-understand formats. The MBTA publishes an exhaustive list on their website. *Rider Tools: App Showcase,* MBTA, *accessed at http://www.mbta.com/rider_tools/apps/.*

[44] *Subscribe to T Alerts*, MBTA, *accessed at* http://www.mbta.com/rider_tools/t_alerts/.

[45] They should have some effect on the reliability numbers because significant delays would result in longer wait times.

[46] *Welcome to the interactive MBTA Performance Dashboard*, T Dashboard, *accessed at* http://mbtabackontrack.com/performance/index.html#/about.

[47] For an example of how this works in London, including the formula used to calculate actual average wait time, see Mickael Schil, *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*, Massachusetts Institute of Technology, May 21, 2012, accessed at https://pdfs.semanticscholar.org/9d93/fe1a0c600a7952f1d8aa998efcfdaa95704f.pdf (page 43).

[48] *MassDOT Tracker: MassDOT's Annual Performance Report Fiscal Year 2016*, MassDOT (November 2016), accessed at http://old.mbta.com/uploadedfiles/About_the_T/Board_Meetings/8%20%20FY16Tracker%20Report.pdf (page 67).

[49] Adam Vacarro, *New MBTA general manager wants to emphasize rider experience*, The Boston Globe, September 7, 2017, accessed at https://www.bostonglobe.com/metro/2017/09/07/new-mbta-general-manager-wants-emphasize-rider-experience/a8HB6z0agytPqQRLaCXnTJ/story.html.

MASSPIRG

Education Fund