Towards a Passenger-Focused On-Time Performance Metric for Commuter Rail

Ted Rosenbaum
LTK Engineering Services
San Francisco, CA

ABSTRACT

Commuter rail agencies in the US traditionally measure on-time performance (OTP) based solely on trains arriving at the terminal station, often with a permissible margin of 5-6 minutes. Today, while many inbound riders do alight at the terminal, few outbound riders have the final station on a given line as their ultimate destination. Furthermore, non-traditional origin-destination pairs (i.e. off-peak, reverse-peak, and intermediate rides) have grown in recent years, leaving the terminal OTP metric ill-suited to adequately communicate a rail line's reliability to its customers. This paper discusses the history of these on-time performance metrics and proposes a new, more passenger-focused standard for commuter rail. The Caltrain line between San Francisco and San Jose is presented as a case study comparing the new metric with the traditional metric, finding that while the traditional metric offers a relatively good indication of reliability, the new metric offers an enhanced view of rider experience associated with intermediate stations. Finally, this paper discusses future opportunities for the automated collection, aggregation, and distribution of this data as various technologies (e.g. positive train control, automatic passenger counters, etc.) become more widespread in the industry.

INTRODUCTION AND HISTORY

Punctuality is one of the most important features of transit service from a customer's perspective. [1] Reporting of on-time performance is nearly universal among rail transit agencies in the United States, though there is variation in how "on-time" is defined. Some of this variation is entirely appropriate: the use case for transit often varies by mode, and performance metrics should reflect those differences. For riders of a high-frequency subway line who rarely consult a timetable, the important metrics are the length of the wait at the originating station and any delays that occur while in transit. For low frequency rail corridors, where riders often consult a schedule before making travel plans, adherence to that schedule is most important.

Commuter Rail, as currently conceived and operated in the US, typically falls into the latter category. Trains are run according to a timetable, and while frequency may be high during the peak periods in the peak direction, there may often be little to no service outside of these times or in the reverse direction. Many commuter railroads measure their on-time performance by adherence to the schedule at the terminal, usually with a five- to six-minute margin. Guidance from various authorities on the matter do not provide a justification for this method, but cite its widespread use as evidence of its sufficiency. [2] One commuter railroad does venture to say that it was adopted by consensus, stating: "Years ago, railroads across the nation decided that delays of less than six minutes would not be factored into a railroad's on-time performance." [3] But variation in the industry today between a five- and six-minute margin would seem to undercut this assertion.

In reality, the adoption of a terminal-based standard of five minutes was likely first imposed on the railroads of the early twentieth century by regulators. This is not to say that railroads did not track on-time performance before this—much of the industry press in the late 1800s was filled with discussions of which trains "made time" on which railroads. But what was considered "on-time" was rarely elaborated upon, nor was performance consistently reported. This is unsurprising, as Standard time (and higher expectations of precision from the riding public) was not implemented in the US until 1883.

When Congress established the Interstate Commerce Commission (ICC) in 1887, one of the goals enshrined in the founding legislation was the standardization of information being reported by the myriad railroads operating across the country. However, this standardization generally was limited to financial information and usually focused on freight movement. [4] In parallel with the ICC, though, was the growth (in both power and number) of state commissions. These state

commissions responded to local complaints regarding not just rates but performance as well.

In New York, the Public Service Commission first convened in 1907 as a successor to the New York State Railroad Commission. The NYPSC was composed of two districts: the First District was limited to New York City with its main effort being the monitoring of the city's new electric railways, while the Second District regulated the steam railroads and other utilities operating throughout the state. In its first annual report at the end of 1907, the NYPSC Second District cited a wealth of complaints about unreliable passenger train service. The Commission hypothesized that poor performance was either the result of poorly administrated railroads or poor maintenance work, and vowed that it would "endeavor by every means in its power to correct such wretched service." [5]

Upon inspecting the various railroads' locomotive failure reports, the Commission was dismayed at the wide variation in defining what constituted an engine failure. A conference with the mechanical officers of the larger railroads was called, which put forth a single definition: "Any failure…that delays a passenger train five minutes, or a freight train twenty minutes, at an initial terminal, or in arriving at a meeting point, a junction point, or a terminal." [5]

This Commission celebrated its effort as the "first effort which has been made in this country in the direction of a systematic attempt by state authorities to make comparisons between railroads, from one period to another, regarding the efficiency of locomotive operation." [5] But the total prevalence of passenger train delays was still a mystery, so on March 16, 1908, the NYPSC promulgated Circular #34. This Circular required all steam railroad corporations operating in the state to submit Passenger Train Movement Form #83 every month, detailing how many trains were operated in the previous month, how many were delayed, and the cause of the delay. Building on the locomotive failure definition, Circular #34 stated that "Trains shall be considered 'on time' when not exceeding five minutes late at division terminal or end of train run."

Circular #34 applied to all steam railroads, which mostly meant corporations operating inter-city trains. However, the Long Island Railroad was already providing commuter fares for regular riders on its steam trains, and began reporting its performance along with the other railroads. Taking the Commission at its word that this was, in fact, the first effort in the country at aggregating these types of statistics, it marks the genesis of the traditional on-time performance metric used with varying margins by commuter railroads today.

Furthermore, Circular #34 stated that its accompanying form had been "revised after consideration of all the criticisms received from the railroad companies of the state." Some of this criticism was on the grounds that the detailed reporting was overly burdensome, but there is no evidence in the Commission's proceedings either what the other criticisms were or whether there was discussion regarding what the "on time" threshold should be.

A look at early railroad safety rules potentially gives a clue as to why five minutes was chosen: until telegraphed train orders were uniformly used, conductors were instructed to allow a five minute margin for meets with opposing trains. This was on the basis that watches at the time were mercurial objects, and even with a time inspector and a consistently-wound watch, two conductors of opposing trains may not have precisely the same time. [6] A five-minute buffer allowed trains to proceed safely, and conductors arriving late at the terminal as a result may have been forgiven their tardiness.

Whatever the underlying rationale for a five-minute buffer, the NYPSC dutifully collected and disseminated its data monthly for nine years, until the nationalization of the railroads in World War I under the United States Railroad Administration (USRA). While railroads made their reports to the NYPSC, it should be noted that some railroads scrutinized commuter trains more closely. The General Superintendent of the New York, New Haven & Hartford Railroad described in 1916 how the full operations of the railroad were tallied daily and the overall performance of the railroad was compared to the 5-minute standard. But there was also a special daily report compiled by 10am showing the performance of that morning's commuter trains. As apparently the 5-minute standard was not precise enough, this report showed the exact minutes of delay for each commuter train. [7]

Like the ICC before it, the USRA mainly focused on freight (and now troop) movements and financial records. However, under the stewardship of William J Cunningham, the USRA Operating Statistics Section sought to standardize recordkeeping practices throughout the industry. This effort included on-time performance, though it generally omitted suburban (commuter) trains, and used a 10 minute margin for terminal arrivals [8] — another indication that through trains and commuter trains were not meant to be held to the same standard.

The USRA's efforts did, however, succeed in forcing railroads nationally to keep consistent records. This meant that after World War I, as the railroads were re-privatized and began competing for customers once again, they had a wealth of data to draw from. Railroads were quick to tout their on-time performance during this time in their in-house magazines, with the Illinois Central Magazine saying that its 98.8% performance in the first 8 months of

1921 "is a service of which every member of the Illinois Central family may be proud." [9] In 1929 and 1930, the Baltimore & Ohio Railroad consistently published its on-time performance in a chart in its monthly magazine, although—like the Illinois Central—its suburban services were lumped in with their long-haul trains. [10]

Eventually, this type of reporting spread throughout the industry, and as commuter lines were taken over by public agencies in the latter half of the twentieth century, the use of performance metrics as a means of holding officials accountable also spread. For a brief time, the ICC defined on-time performance by terminal arrival time, but it was defined with a 5 minute margin for every 100 miles of operation, and only applied to Amtrak service. [11] Today, nearly every railroad operating in the US publicly tracks on-time performance on at least a monthly basis. Of these, the majority use an acceptable margin at the terminal of 5 minutes 59 seconds; several use 5 minutes, and fewer measure at nonterminal stations. [12]

PROBLEM DESCRIPTION

Of course, the riding public is savvy, and has questioned the use of the traditional on-time performance metric for some time. As early as 1953, letters to the New York Times splashed cold water on the impressive-sounding performance of the local railroads. One rider vehemently protested the Long Island Railroad's proclamation of a 98.64% on-time performance in December 1952, calling it a "statistical hoax… [which] allows the railroad to take credit for trains being on-time at 3 o'clock in the morning or 1 o'clock in the afternoon, when no operating problem confronts them." [13]

That rider was, of course, correct, but neglected two other issues which arguably muddy the on-time performance waters even more. First, grading service exclusively at one station allows agencies to more easily inflate their performance. This can be done systematically, by padding the schedule between the penultimate and terminal stations to ensure that a train arrives "on time" even if it was several minutes late at nearly every intermediate station; or individually, as operators have shown a bias against reporting delays 1-2 minutes beyond the threshold. [1] Second, not all riders ride to the terminal station. While many do in fact reach a terminal (often near the central business district) during the morning peak, few riders remain at the end of the line in the afternoon in the opposite direction. Even if every rider on a given system were to reach the terminal in the morning peak, that's still only half a day's worth of ridership.

In fact, ridership in recent years has seen a marked increase in non-traditional trips on commuter railroads: off-peak trips, reverse-peak trips, and intermediate trips (where neither the origin nor destination is a terminal) have grown on many systems. [14] This confluence of events begs the following questions: is the traditional on-time performance metric an acceptable proxy for passenger experience? Is there another way to measure on-time performance which is more suitable and which can be computed without being a burden on a transit agency?

CASE STUDY

Caltrain is an excellent case study through which to examine these questions. Caltrain is the Commuter Rail agency overseen by the Peninsula Corridor Joint Powers Board (PCJPB or JPB) operating in California between San Francisco and San Jose, with additional limited service 30 miles south to Gilroy. In the last decade-plus as the economies and populations of San Francisco, San Jose, and Silicon Valley have grown, Caltrain's ridership has more than doubled—from 29,760 on an average weekday in 2006, to 62,190 in 2017. [15]

Throughout that time, Caltrain has measured and reported its on-time performance to the Joint Powers Board every month in terms of trains arriving at their terminal station (northbound trains at 4th & King in San Francisco, southbound trains at San Jose Diridon Station, Tamien, or Gilroy) within 5 minutes 59 seconds of the scheduled time. [16] For purposes of tracking mid-line customer experience and to provide a check on the accuracy of the timetable, Caltrain also tracks on-time performance in both directions at Redwood City, which is about halfway from San Francisco to San Jose. Although performance at Redwood City is tracked daily, it is not currently reported publicly.

However, despite the presence of major job centers at both ends of the line, Caltrain's ridership has always derived a large segment of its ridership from intermediate riders. For instance, Palo Alto, home to Stanford University and sitting near the midpoint of the corridor, has long been the second busiest station on the line—after San Francisco and ahead of San Jose. In 2017, the 4th & King terminal accounted for just 58% of all northbound alightings during the morning weekday peak period.

Recognizing that it does not fit into the traditional American "Commuter Rail" paradigm, Caltrain is currently undertaking a massive capital project to electrify the corridor, allowing faster, more frequent trains to serve communities throughout the Peninsula, ultimately transforming Caltrain into a regional rail mass transit provider. [17] Once this change occurs, passengers may

have a different conception of what adequate performance looks like. In the meantime, Caltrain's ridership is as likely a candidate as any Commuter Railroad in the United States to have a divergence between its traditional on-time performance metric and the experience of its riders.

This case study examines that divergence by comparing the on-time performance reported to the JPB each month between April 2016 and April 2017 to an estimate of the number of passengers during that time who arrived at their actual destination station on time.

Data Collection

Although Caltrain does not currently have automatic vehicle locators or automatic passenger counters, it does have a wealth of data available to utilize. Some data, like ridership counts, is human-generated, while station arrival times are automated. Data collection for this case study required developing an estimate of the number of alightings at every stop for each train on a monthly basis, and determining which trains served were delayed at which stations on a daily basis. This information was then aggregated in the final analysis.

Ridership Data

First, Caltrain passenger alightings—broken out by train and station—had to be estimated on a monthly basis. The analysis began with the detailed ridership count performed every February by Caltrain. For several days that month, counters ride every train and tally both boardings and alightings at every station, aggregating the results to smooth out any atypical days. The findings have been published on Caltrain's website going back to 2002, with the raw data also being provided for public use since 2016. Second, every month Caltrain staff report average weekday ridership to the Joint Powers Board. In part because Caltrain provides direct service from the Peninsula to the San Francisco Giants baseball stadium, ridership has a strong seasonality to it, as ridership is on average 5-15% higher from May through September compared with the detailed February counts. (Figure 1)

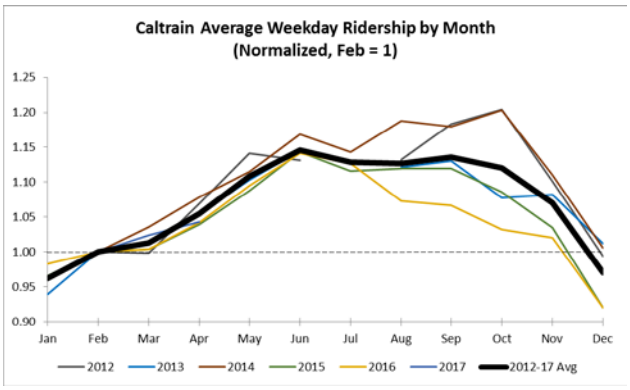

Figure 1. Caltrain Average Weekday Ridership by Month (Normalized, Feb = 1.0).

To estimate the number of passengers alighting from every train at every station throughout the year, linear interpolation was used between the February 2016 and February 2017 count, with a seasonal adjustment added on top of that. As an example, In February 2016, northbound train #233 saw 93 passengers alight at Redwood City, while in February 2017, this figure was 73—a decrease of 20 passengers. To estimate the number of alightings in August 2016—halfway between the two official counts—half of 20 (or 10) passengers were subtracted from the February 2016 count, making 83 passengers. Then, an additional 12.7% (or 11 passengers) were added because August ridership is typically 12.7% higher than February's. This results in an estimated 94 alightings at Redwood City every weekday in August.

Caltrain changed its schedule on April 10, 2017. While still operating five trains per hour per direction in the peak period, the time of trains and stopping patterns changed enough to remove the comparability between the detailed February passenger counts conducted in 2017 and 2018. As a result, no interpolation was used to determine the March 2017 passenger counts; only a seasonal increase was applied for that month. Because the analysis presented here was conducted before the February 2018 passenger counts were available, March 2017—the last full month on the old schedule—is the most recent month analyzed.

Train Delay Data

In parallel with the ridership data, train delays had to be determined at every stop (between 5 and 30 stations depending on the stopping pattern) for all 92 scheduled weekday trains over the 12 months examined in this case study (253 weekdays in all). These data are not currently made available to the public, though Caltrain's contract operator—TransitAmerica Services, Inc (TASI)—was able to provide the data in a usable format, with a separate

file for each month of data. Each file included a 93 column matrix that showed train identifying information (date, run number, locomotive number, and cab car number), as well as scheduled arrival times, actual arrival times, and the resultant delay for every station stop.

Since Caltrain does not have a GPS-connected automatic vehicle locator system or station occupancy circuits, arrival data is estimated. Caltrain's Rail Operations Control System (ROCS) generates track circuit occupancy times, from which arrivals (rounded to the nearest minute) at a nearby station can be estimated. While not a perfect measurement, its automated nature removes the bias sometimes present in human-generated arrival times. To remain consistent with Caltrain's reported terminal-based on-time metric, only trains arriving at least 5 minutes, 59 seconds late to any particular station were considered responsible for delayed passengers.

RESULTS

Passenger destination data was then cross-referenced with train delay data, resulting in detailed counts of how many passengers actually arrived at their destination within 6 minutes of the scheduled time. Passenger arrival was chosen (without regard to excess waiting time or other metrics), as late arrivals have been shown to be the most pressing concern of riders. [18]

To take an illustrative example, it is worth looking at Wednesday, June 22, 2016. Caltrain's midday service was not especially punctual that day, with 3 of its 7 northbound trains arriving more than 5 minutes 59 seconds late at 4th & King in San Francisco. However, these three trains combined delayed 790 passengers—just 1.1% of the day's estimated 71,450 riders. These three trains knocked down Caltrain's traditional OTP by 3.3%, as Caltrain finished the day with 94.57% of its trains arriving on time--just under its goal of 95%. But if Caltrain was judged by how many of its passengers arrived on time, it would have achieved a mark of 95.02% that day—a reversal of the "statistical hoax" described earlier.

Comparison of Traditional vs. Passenger-Focused OTP Metrics

The differences between traditional and passenger-centric OTP metrics can be the result of several factors, some of which work in opposition to each other. As shown above, if several trains are delayed on a single day, but they are all midday trains with low ridership, passenger OTP is likely to be higher than traditional OTP. Working in the other direction, if padding in the schedule allows a train to reach the terminal and be counted as on-time, but most of its passengers disembarked at intermediate stations, passenger OTP will be lower than traditional OTP.

For Caltrain during the study period, these factors resulted in a wide difference between the two OTP metrics. (Figure 2) On any given day, the difference between the traditional metric and the passenger-focused metric ranged from +10.38 (passenger > traditional) to -7.70 (traditional > passenger) percentage points, with the average magnitude of the difference being 1.87 percentage points. For perspective, with Caltrain carrying over 60,000 passengers each weekday, this difference amounts to over 1,100 passengers.

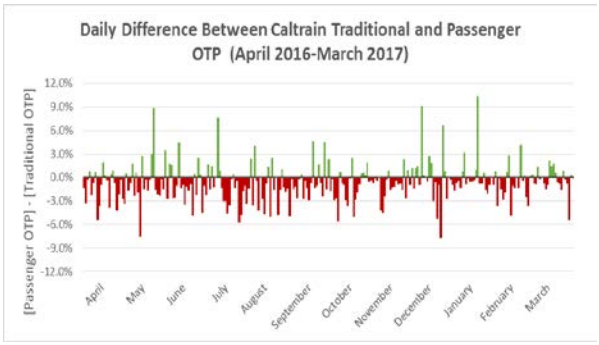

Figure 2. Daily comparison of Caltrain's traditional and passenger-focused OTP

However, when this data is aggregated up to a single monthly value as Caltrain generally reports it (Figure 3), the variations tend to cancel each other out, bringing the two metrics in closer agreement—within two percentage points. At this level it also becomes clear that the traditional OTP metric tends to slightly overstate how many passengers are arriving at their destinations within an acceptable margin.

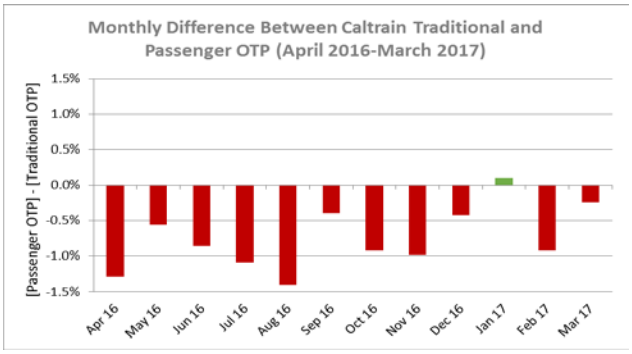

Figure 3. Monthly comparison of Caltrain's traditional and passenger-focused OTP

Aggregating up another level and looking at annual on-time performance makes the tendency for daily variation to even-out most stark: For the twelve months from April 2016 through March 2017, Caltrain's weekday on-time performance was 93.34% if measured the traditional way, and 92.54% if measured based on passengers—a difference of less than one percentage point. The difference between the two metrics shows what was visible in Figure 3; the traditional on-time performance metric yields a slightly more positive view of Caltrain's performance compared with how riders may experience its service.

CONCLUSIONS & DISCUSSION

Similar to the heavy rail results found by Henderson et al., [19] this case study shows that the passenger experience of on-time performance does not meaningfully differ from the traditional, terminal-based on-time performance. This is especially true when reporting on monthly averages rather than daily performance. Caltrain, whose ridership least resembles the traditional American "Commuter Rail" construct, would be most susceptible to a divergence between train and passenger on-time performance. The small one found here indicates that for other agencies the difference may be negligible. As a result, there should be little trepidation on the part of agency staff that switching to a passenger-centric on-time performance metric would expose any deficiencies the traditional metric does not already.

To the contrary, providing a more publicly-understandable and transparent accounting of performance is in the best interest of every public transit agency, as it engenders good faith between it, the riding public it serves, and the elected officials who oversee it. Several heavy rail operators have begun to modify their on-time performance metrics in recent years, usually moving away from the traditional train- and terminal-based metric. [20] While heavy rail has the advantage of turnstiles which continuously collect accurate ridership data, Commuter Rail agencies still have a wealth of data at their disposal which can provide similar insights, especially as automatic passenger counter (APC) technology becomes more widespread in the industry. Even without APC technology, most Commuter Rail agencies collect station-level data at a regular interval. Since Commuter Rail schedules tend to be relatively static compared to heavy rail, passenger counts remain applicable for longer.

Along with ridership data, station arrival data is necessary to complete the type of analysis presented here. It is heartening to see that Commuter Rail agencies have recently begun providing real-time information to passengers, indicating an understanding that to-the-minute (if not the second) performance is expected of them. [21] This is likely to become more widespread, as newer rolling stock typically includes automatic vehicle locator technology. Systems designed to comply with the federal positive train control (PTC) mandate will deliver a wealth of locational data to agencies; even if not made publicly available, it could be used as an input to a passenger-focused on-time performance metric.

Finally, Commuter Rail agencies are no longer responsible for long-distance routes as they were when the New York Public Service Commission first regulated them. Modern customers who ride twice daily for 20-60 minutes each direction have different expectations than riders a century ago who went on multi-hour journeys between cities. This paper is not meant to address the suitability of a 5-minute (or 5 minute 59 second) threshold for lateness. But given the history presented here, agencies would be wise to re-examine their performance metrics and decide for themselves on an acceptable level of precision to which their operations should be held.

ACKNOWLEDGEMENTS

This paper is dedicated to the memory of Bert Frazier of TASI, who provided the raw data for the case study. The author also wishes to acknowledge the generous help and support offered by many on Caltrain's staff, including Michelle Bouchard and Matt Verhoff in particular. Any errors are the author's own.

REFERENCES

1. Nelson, David, and O'Neil, Kay. "Commuter Rail Service Reliability." Transportation Research Record, Vol. 1704 (2000). https://doi.org/10.3141/1704-07.
2. Transit Cooperative Research Program Report 88 – A Guidebook for Developing a Transit Performance Measurement System. Transportation Research Board: Washington, DC (2003). [pg 5-29]
3. "On-time Performance Reports." Metra. Accessed March 2, 2018. https://metrarail.com/about-metra/reports-documents/operations-ridership-data/time-performance-reports.
4. Cunningham, William J. "The Accomplishments of the United States Railroad Administration in Unifying and Standardizing the Statistics of Operation." The Annals of the American Academy of Political and Social Science Vol. 86 (1919). http://www.jstor.org/stable/1013803.

5. New York Public Service Commission Second District. Annual Report. Albany: JB Lyon Company, State Printers, 1907. NY PSC Annual Report, 1907, https://hdl.handle.net/2027/mdp.39015066701759.
6. Pope, SA. “Ask the Conductor; He Has the Right Time.” Southern Pacific Bulletin vol. 12, no. 8, August 1923, p. 3-7. https://books.google.com/books?id=4hAhAQAAIAAJ
7. Droege, John A. Passenger Terminals and Trains, New York: McGraw-Hill Book Company, Inc., 1916.
8. “Performance of Passenger Trains.” Railway Review, Vol. 63, no. 4, July 27, 1918, p. 142. https://hdl.handle.net/2027/njp.32101048990178.
9. “Service as Scheduled.” Illinois Central Magazine, Vol. 10, no. 4, October 1921. Pg. 67-68. (https://archive.org/details/illinoiscentrale21illi)
10. Baltimore & Ohio Magazine, Vol. 17, 1930.B&O Mag, 1929. https://play.google.com/books/reader?id=2l9CAQAAIAAJ&pg=GBS.RA4-PA9.
11. Adequacy of Intercity Rail Passenger Serv., 344 I.C.C. 758, 776 (1973).
12. Various Commuter Rail Title VI Policies; accessed March 2018.
13. Nugent, William A. “Lateness of Commuting Trains.” The New York Times, February 6, 1953.
14. Jaffe, Eric. “Far Beyond Rush Hour: The Incredible Rise of Off-Peak Public Transportation.” CityLab, February 6, 2014, accessed March 4, 2018. https://www.citylab.com/transportation/2014/02/far-beyond-rush-hour-incredible-rise-peak-public-transportation/8311/.
15. “Caltrain 2017 Annual Passenger Count Key Findings.” Caltrain Ridership. Board of Directors Presentations. August 1, 2017, accessed April 3, 2018. http://www.caltrain.com/Assets/__Agendas+and+Minutes/JPB/Board+of+Directors/Presentations/2017/2017-06-01+Annual+Passenger+Counts.pdf
16. “Caltrain On Time Performance & Delay Mitigation.” Caltrain presentation, November 16, 2017, accessed April 2, 2018. http://www.caltrain.com/Assets/BAC/pdf/BAC+On+Time+Performance+$!26+Delay+Mitigation+11.16.17.pdf
17. “Peninsula Corridor Electrification Project Final Environmental Impact Report”. Caltrain, December 2014.
18. Kroes, Eric, et al. 2007. “Value of Punctuality on Suburban Trains to and from Paris.” Transportation Research Record, vol. 2006 (2007). https://doi.org/10.3141/2006-08.
19. Henderson, Gary, Heba Adkins, and Philip Kwong. "Toward a passenger-oriented model of subway performance." Transportation Research Record 1266 (1990): 221-228.
20. “MBTA Back on Track – Reliability.” MBTA Performance Dashboard. Accessed February 6, 2018. http://www.mbtabackontrack.com/performance/index.html#/detail/reliability///
21. Brakewood, Candace, et al. An Analysis of Real-time Commuter Rail Information in Boston, TRB 2014 Annual Meeting. 2014. http://amonline.trb.org/trb-59976-2014-1.2467145/t-0000-1.2476566/326-1.2476418/14-1268-1.2477055/14-1268-1.2477058