

Definition and Properties of Alternative Bus Service Reliability Measures at the Stop Level

*Meead Saberi and Ali Zockaie K., Northwestern University
Wei Feng, Portland State University
Ahmed El-Geneidy, McGill University*

Abstract

The Transit Capacity and Quality of Service Manual (TCQSM) provides transit agencies with tools for measuring system performance at different levels of operation. Bus service reliability, one of the key performance measures, has become a major concern of both transit operators and users because it significantly affects user experience and service quality perceptions. The objective of this paper is to assess the existing reliability measures proposed by TCQSM and develop new ones at the bus stop level. The latter are not suggested as replacements for the existing measures; rather, they are complementary. Using empirical data from archived Bus Dispatch System (BDS) data in Portland, Oregon, a number of key characteristics of distributions of delay (schedule deviation) and headway deviation are identified. In addition, the proposed reliability measures at the stop level are capable of differentiating between the costs of being early versus late. The results of this study can be implemented in transit operations for use in improving schedules and operations strategies. Also, transit agencies can use the proposed reliability measures to evaluate and prioritize stops for operational improvement purposes.

Introduction

Monitoring of the performance measures of public transportation systems has improved since advanced surveillance, monitoring, and management systems have been deployed by transit agencies worldwide. In recent years, service reliability, a key performance measure, has become an important topic for researchers, transit agencies, and policy makers because it significantly affects user experience and service level perceptions. Reliability affects the waiting time of passengers at a stop for a bus to arrive. Reliability also affects total trip time of a passenger. Abkowitz et al. (1978) suggested that reliability is one of the most important factors influencing passenger mode choice. Bowman and Turnquist (1981) found that service unreliability increases operating costs. Reliability is influenced by a number of factors. As listed in the TCQSM (2004), these factors include:

- traffic conditions
- road construction
- vehicle and maintenance quality
- vehicle and staff availability
- transit preferential treatments
- schedule achievability
- evenness of passenger demand
- differences in operator driving skills, route familiarity, and adherence to schedule
- wheelchair lift and ramp usage (generally dwell time)
- route length and number of stops
- operations control strategies
- weather
- incidents

Despite all these, the transportation profession lacked a uniform set of transit-capacity and quality-of-service definitions, principles, practices, and procedures for planning, designing, and operating vehicles and facilities until the publication of the *Transit Capacity and Quality of Service Manual* (TCQSM 1999), First Edition. Also the *Highway Capacity Manual 2000* (HCM 2000) provides a broad range of Level-of-Service (LOS) measures for all the modes, including auto, transit, bicycle, and pedestrian modes. Chapter 27 of the HCM 2000 provides four transit LOS measures: service frequency, hours of service, passenger load, and service reliabil-

ity. Most recently, the *Highway Capacity Manual 2010* (HCM 2010) defines transit service reliability as the “unplanned passenger waiting time at the stop.” Also, chapter 17 of the HCM 2010 suggests that excess wait time reflects transit vehicle reliability. In 2004, Transportation Research Board’s Transit Cooperative Research Program (TCRP) released the second edition of the TCQSM (2004), which contains information about various types of public transportation and provides a framework for measuring transit availability and quality of service from the passenger point of view. The TCQSM introduces a new approach to measure performance of transit service using a two-dimensional LOS framework covering two service quality dimensions (availability and comfort/convenience) for three levels (stops, route segments, and the whole system). Camus et al. (2005) discussed advantages and limitations of the TCQSM method for LOS estimation. The TCQSM reliability measures 1) do not consider the amount of delay but only the number of trips that are late, 2) do not adequately address the effect of early departures on users, and 3) introduce a fixed tolerance around the schedule to estimate the on-time performance.

From different perspectives, it may be of interest to investigate reliability from the standpoint of changes and adaptations necessary in case of system disturbances (such as unavailability of service between certain stations, etc.). This is considered in some publications (see, e.g., Kepaptsoglou and Karlaftis, 2009) but usually is not a topic in transit reliability. Moreover, issues considered in behavioral sciences may also be investigated in public mass transit. For instance, according to Duarte et al. (2010), public transport service impacts the quality of travel experience and the well-being of travelers, as well as their travel behavior leading to the influences of transportation happiness or satisfaction on transport mode choice.

The objective of this paper is to assess the existing reliability measures proposed by the TCQSM and develop alternative complementary reliability measures that account for the interactions among the above-listed factors and capture more characteristics of transit service unreliability. This paper uses the TCQSM definition of reliability: “Reliability includes both on-time performance and the evenness of headways between transit vehicles” (TCQSM 2004). Using empirical data from archived Bus Dispatch System (BDS) data in Portland, Oregon, several key characteristics of distributions of delay and headway deviation are identified, and alternative measures at the stop level are proposed. Toward this end, the results of this study can be fed into the transit operations field for use in improving schedules and operations strategies. Also transit agencies can use the proposed reliability measures to evaluate and prioritize stops for operational improvement purposes.

The remainder of this paper is organized as follows. The second section provides a brief background on transit reliability. In the third section, existing reliability measures in the TCQSM are reviewed. In the fourth section, derivations of the proposed reliability measures are presented. The fifth section presents some empirical analysis results, and the last section concludes the paper.

Background

Various studies build upon the body of research on bus service reliability by employing detailed statistical analysis to measure service reliability using archived Automatic Vehicle Location (AVL) data. AVL technology has been widely implemented in the transit industry in the past decade. Bertini and El-Geneidy (2003) demonstrated robust ways that data collected by a BDS can be converted into potentially valuable transit performance measures.

The Metropolitan Transportation District of Oregon (TriMet) provides transit service in the three-county Portland metropolitan area. TriMet operates 62 million annual bus trips, serving a population of 1.2 million in a 592-square-mile area with 700 vehicles on 98 routes. TriMet's BDS reports detailed operating information in real time, every 90 seconds. In addition, the BDS archives very detailed stop-level data from the bus during all trips (Bertini and El-Geneidy 2003). This includes actual stop time, dwell time, and number of boarding and alighting passengers at every stop. Each geocoded stop has a predefined 30-m (98-ft) stop circle around the stop. The BDS records the arrival time when the bus enters the stop circle and records the departure time when the bus departs the same circle (Bertini and El-Geneidy 2004) (see Figure 1).

Using the archived BDS data, a number of measures can be simply calculated. The scheduled headway at a particular stop can be computed as the scheduled stop time for trip i at a stop minus the scheduled stop time for trip $i-1$ at the same stop:

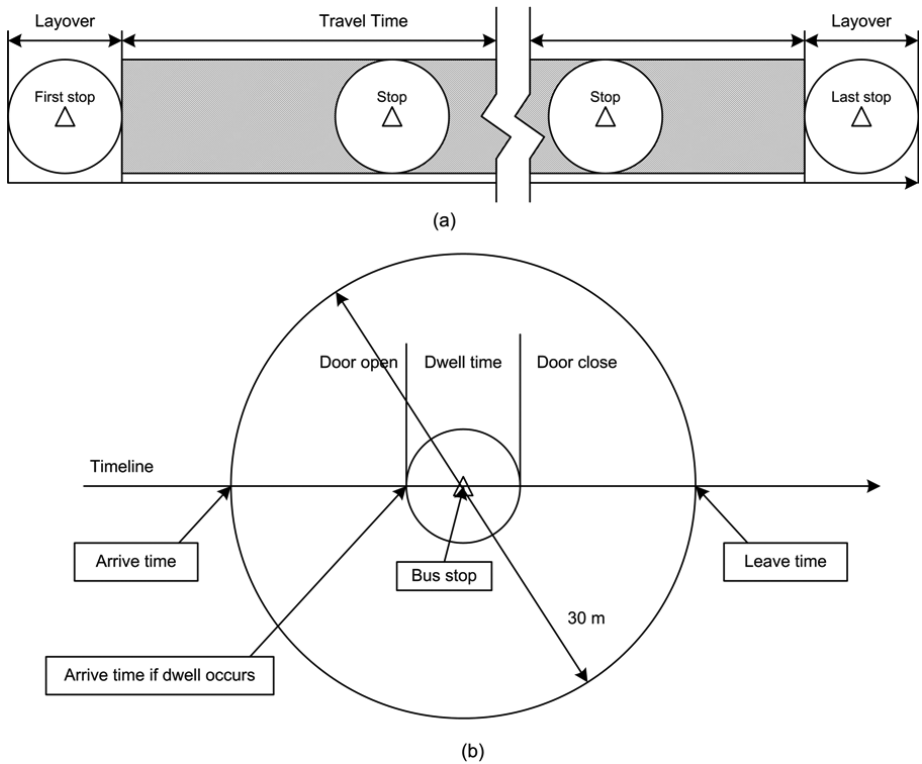
$$\text{Scheduled Headway} = \text{stop time}_i - \text{stop time}_{i-1} \quad (1)$$

Similarly, actual headway, delay (or schedule deviation), and headway deviation can be computed as follows:

$$\text{Actual Headway} = \text{leave time}_i - \text{leave time}_{i-1} \quad (2)$$

$$\text{Delay (schedule deviation)} = \text{leave time}_i - \text{stop time}_i \quad (3)$$

$$\text{Headway Deviation} = \text{Actual Headway} - \text{Scheduled Headway} \quad (4)$$



Source: Bertini and El-Geneidy 2003

Figure 1. TriMet BDS system: (a) time distribution, (b) stop circle description

Lin et al. (2008) used AVL data from Chicago Transit Authority (CTA) bus routes to develop a quality control framework involving Data Envelopment Analysis (DEA). The framework aggregates different service reliability measures into a comprehensive reliability measure. El-Geneidy et al. (2010) used AVL data from Metro Transit in Minnesota to analyze bus service reliability of a few routes at the segment and route levels in Minneapolis. A review of AVL system implementations in the U.S. can be found in El-Geneidy et al. (2010) and Furth et al. (2003).

Several studies have evaluated existing reliability measures and proposed new metrics at different levels (Camus et al. 2005; Xin et al. 2005; Tumlin et al. 2005; Furth and Muller 2006; Fu et al. 2007; Ap. Serratini et al. 2008; Chen et al. 2009). Camus et al. (2005) discussed advantages and limitations of the TCQSM method for LOS estimation and proposed a new reliability measure named “weighted delay index.” Xin et al. (2005) used the TCQSM measures to study several routes and found that

TCQSM measures are sensitive to planning/design variables and can be simply calculated by transit agencies using available data. Tumlin et al. (2005) developed a method to evaluate transit performance in the context of different transportation environments. Furth and Muller (2006) found that traditional transit service measures underestimate the total costs of service unreliability because waiting time and service reliability are analyzed separately. Fu et al. (2007) developed a Transit Service Indicator (TSI) that estimates the quality of service results from the interaction of supply and demand. TSI uses multiple performance measures, including hours of service and service frequency. Ap. Sorratini et al. (2008) investigated measures to assess reliability, such as headway, excess waiting time, service regularity, and recovery time of an urban network, using a dynamic micro-simulation model (DRACULA). Most recently, Van Oort et al. (2012) studied ways to improve reliability by adjusting schedule timetables using holding points. To measure reliability, they used punctuality (deviation from the scheduled arrival time) and probability of departing on time.

Reliability measures are important because they can be used to identify bus bunching. Unreliable routes are more likely to experience bunching. "Bus bunching takes place when headways between buses are irregular leading to longer waiting times for riders, overcrowding in some buses, low numbers of passengers in the remaining buses, and an overall decrease on the level of service and capacity" (Feng and Figliozzi 2011). For additional information on bus bunching see, e.g., Bellei and Gkoulmas (2010).

None of the above-mentioned studies have used empirical cumulative distribution of delay or headway deviation obtained from detailed AVL data. Chen et al. (2009) proposed three reliability measures using data from the Beijing transit system: a Punctuality Index based on routes (PIR), a Deviation Index based on stops (DIS), and an Evenness Index based on stops (EIS). The EIS and DIS measures proposed by Chen et al. (2009) require a coefficient of variation of headway, individual headway deviation at each stop, and boardings at each stop. However, unlike the proposed measures in this paper, they do not fully take advantage of the characteristics provided by a cumulative distribution.

Existing Reliability Measures

Bus service reliability has been defined in a variety of ways, from the perspective of both users and transit agencies. Characterizing the user-perceived service reliability is quite complicated due to the heterogeneity of user preferences, views, and values of time. That is, transit agencies use several different reliability measures. The most widely used of these are on-time performance and headway adherence.

Some agencies also use missed trips and distance traveled between mechanical breakdowns. When buses run at frequent intervals, usually less than 10 minutes, headway adherence becomes more important from the perspective of a passenger. Poor headway adherence causes bus bunching, overcrowding on the lead bus, and longer waiting times. For a passenger arriving shortly before a scheduled bus departure, an early departure is equivalent to a bus being delayed a full headway.

The current reliability LOS proposed by the TCQSM considers on-time performance to be an arrival no more than five minutes after the scheduled time. Early departures are considered on-time only in locations where no passengers would typically board. Most transit agencies consider a bus to be late when it is more than five minutes behind the schedule. Early departures are considered to be as bad as being late. Some agencies allow buses to depart up to one minute ahead of the scheduled time. Transit agencies use on-time performance as a key measure of schedule adherence for evaluating system reliability. Therefore, it is important to differentiate between buses that are late versus early, because the cost of being late is different from the cost of being early. Also, it is necessary to know how late and how early buses are. The on-time performance measure proposed by the TCQSM does not take these factors into account.

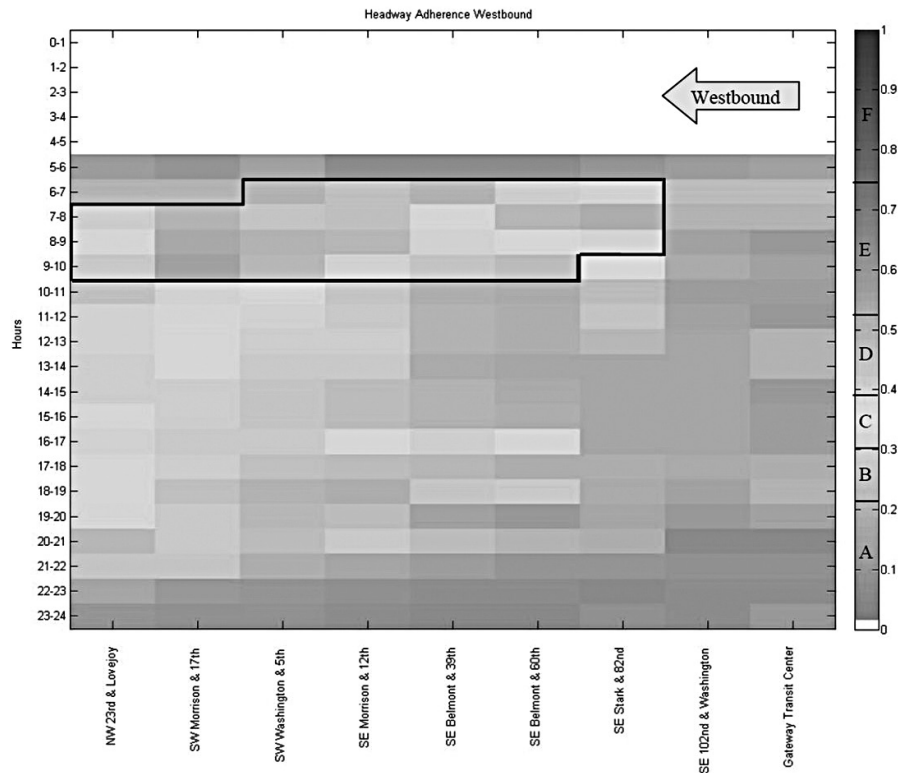
For frequent services, headway adherence is used to determine reliability. As in the TCQSM, headway adherence can be calculated as follows:

$$c_{vh} = \frac{\text{standard deviation of headway deviations}}{\text{mean scheduled headway}} \quad (5)$$

where c_{vh} = coefficient of variation of headways (headway adherence).

Headway adherence is based on standard deviation only and does not capture the extreme cases of unreliability. Also, similar to the on-time performance measure, it does not differentiate between the cost of being early versus late.

Figure 2 shows a color counter time-space diagram of a selected bus route in Portland, Oregon (Route 15 westbound), visualizing hourly calculated headway adherence. The color, ranging from gray to light gray, represents low LOS to high LOS. The white area in the color counter time-space diagram shows that there are no data for those time intervals and the outlined area represents the high frequency service time periods and stop locations.



Source: Feng and Figliozzi 2011

Figure 2. Color counter time-space diagram of headway adherence

Figure 3 illustrates the empirical cumulative distribution of delay at the SE Stark & 82nd stop (solid curve) from the same bus route shown in Figure 2. The dashed curve is the same distribution when altered slightly, representing delay distribution at a hypothetical stop. These two distributions have identical standard deviation (121.5 sec) and, therefore, identical headway adherences. However, they have considerably different width, defined as the 95th percentile of delay minus the 5th percentile of delay. The distribution width of the solid curve is 378 sec, and the distribution width of the dashed curve is 442.5 sec. This implies different unreliability characteristics that cannot be captured by the existing TCQSM metrics and, thus, calls for a supplementary measure.

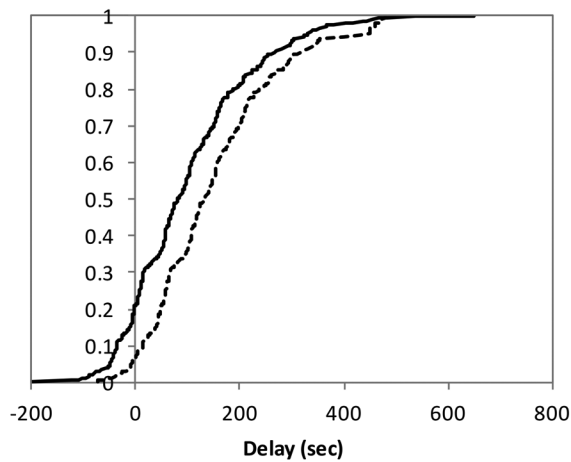


Figure 3. Empirical cumulative distribution of delay at the SE Stark & 82nd (solid curve) and a hypothetical cumulative distribution curve (dashed curve)

Derivation of New Measures for Bus Service Reliability

Focusing on service reliability from the perspective of a transit agency, we propose new reliability measures, using distribution of delays and headway deviations. Here, we use the term delay for schedule deviation. It should be noted that, in some cases, reliability measures from the perspective of a transit agency are entirely different from the user-perceived service reliability. Passenger perceptions of service reliability are partly related to service frequency. Routes with higher frequency may be considered reliable by passengers even if they have poor service reliability. A study by TriMet reported in Kimpel (2001) showed that passengers are more likely to express satisfaction with the performance of bus routes that operate at high frequencies, although later analysis demonstrated that these same routes were among the least reliable. This obvious discrepancy exists because passenger waiting times are still relatively short on high frequency routes with inadequate service reliability, compared to better-performing routes that operate less frequently (Kimpel 2001). Therefore, schedule adherence has been the most important existing reliability measure for infrequent services that operate with headways of more than 10 minutes. For routes characterized by high frequency service, headway variability has been the most important existing reliability measure.

In the remainder of this section, three alternative reliability measures are proposed. For frequent services the distribution of headway deviations and for non-frequent services the distribution of delays are used to capture unreliability characteristics of a bus service.

Earliness Index

The Earliness Index (EI) is defined as the percentile rank of delay/headway deviation of zero. The percentile rank of a particular delay/headway deviation is the percentage of delay/headway deviations in its frequency distribution that are lower or equal to it. Let X denote the delay (for infrequent services) or headway deviation (for frequent services) and $F(x)$ denote the cumulative distribution function of x as follows:

$$F(x) = P(X \leq x) \quad (6)$$

Therefore, EI can be defined as $F(x)$ when $x=0$:

$$EI = F(0) \quad (7)$$

Figure 4(a) is a graphical representation of EI on an empirical cumulative distribution function of x . EI ranges between 0 and 1. For frequent services, an EI of 0 represents the “all behind schedule” condition and an EI of 1 represents the “all ahead of schedule” condition. For not frequent services, an EI of 0 represents the “all late” condition and an EI of 1 represents the “all early” condition. For infrequent services, the theoretical ideal distribution lays on the y-axis of the cumulative distribution function. Buses that are early can be treated as being one headway late, because passengers who are arriving near the scheduled departure time would have to wait for the next bus. Therefore, the “all late” condition is expected to be the achievable ideal distribution for non-frequent services to avoid early departures. Note that the above statement is true only when the theoretical ideal distribution (all “on-time” condition) is not achievable. The closer the EI is to 0, the more reliable is the service. For frequent services, one cannot argue similarly, since maintaining a fixed headway with a small deviation is more important than being ahead of or behind the schedule. Thus, another measure is required to capture the variation of headways.

Width Index

To capture the width of the distribution of headway deviations in frequent services, the Width Index (WI) is defined as the 95th percentile of headway deviations minus the 5th percentile of headway deviations divided by the average scheduled head-

way. Note that the percentiles used here are based on the level of desired reliability and can be adjusted:

$$WI = \frac{F^{-1}(0.95) - F^{-1}(0.05)}{\text{Average Scheduled Headway}} \quad (8)$$

where $F^{-1}(p)$ is the inverse of the cumulative distribution function defined earlier. Similarly, for frequent services, WI can be defined as the 95th percentile of delays minus the 5th percentile of delays divided by the average scheduled headway.

Figure 4(a) is a graphical representation of WI on an empirical cumulative distribution function of x . The ideal width of the distribution of delays/headway deviations is zero ($WI = 0$) when the 95th percentile and the 5th percentile are equal. Regardless of the frequency of the system, average scheduled headway is used in the denominator of the WI to keep the measure unitless.

To distinguish between the cost of being early versus being late, some modifications can be made to Equation (8) as follows:

$$WI = \frac{\alpha[F^{-1}(0.95)] - \beta[F^{-1}(0.05)]}{\text{Average Scheduled Headway} \cdot (\alpha + \beta)} \quad (9)$$

where $F^{-1}(0.95) > 0$, $F^{-1}(0.05) < 0$, α is the weight associated with the cost of being late, and β is the weight associated with the cost of being early. Note that the formulation should be adjusted if needed for other shapes of CDF. If the CDF is completely on the right side of the y axis in Figure 4(a), where $F^{-1}(0.95) > 0$ and $F^{-1}(0.05) > 0$, then the WI can be adjusted as follows:

$$WI = \frac{\alpha[F^{-1}(0.95)] - \alpha[F^{-1}(0.05)]}{\text{Average Scheduled Headway} \cdot (\alpha + \beta)} \quad (10)$$

If the CDF is completely on the left side of the y axis in Fig. 4(a), where $F^{-1}(0.95) < 0$ and $F^{-1}(0.05) < 0$, then the WI can be adjusted as follows:

$$WI = \frac{+\beta[F^{-1}(0.95)] - \beta[F^{-1}(0.05)]}{\text{Average Scheduled Headway} \cdot (\alpha + \beta)} \quad (11)$$

It is also worth mentioning that the width index is capturing more extreme values in a distribution compared to the coefficient of variation (headway adherence).

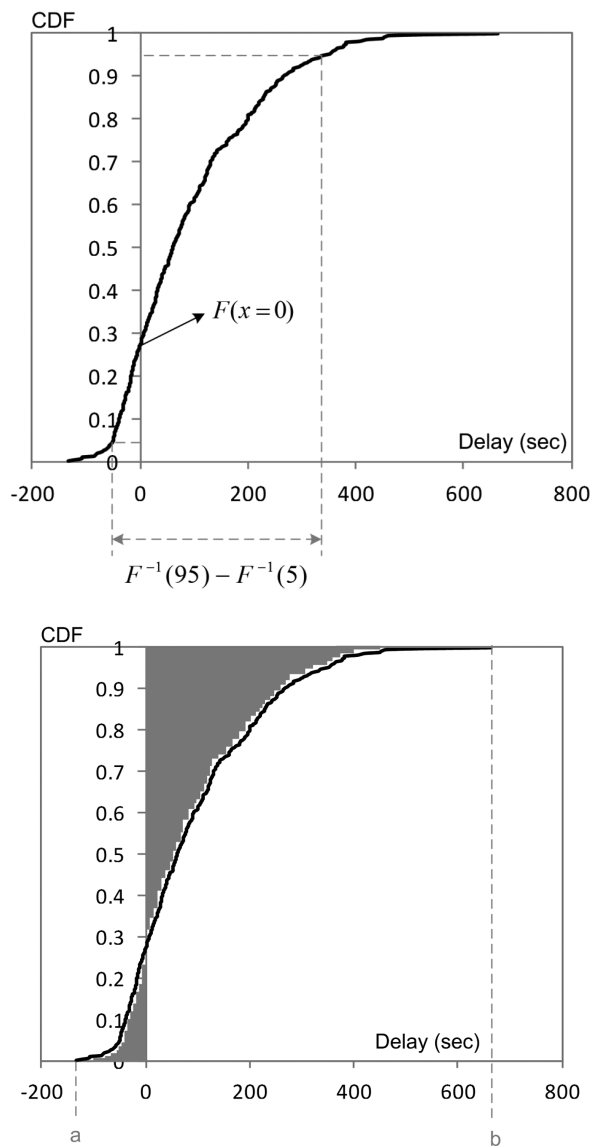


Figure 4. Graphical representation of (a) EI and WI and (b) SSDI on an empirical cumulative distribution function of x (SE Belmont & 60th stop, Route 15, Portland, OR)

Second Order Stochastic Dominance (SSD) Index

To further capture characteristics of the distribution of delays/headway deviations, the second-order stochastic dominance (SSD) concept is applied. Assume we have two cumulative distribution functions of F_A and F_B . A is considered second-order stochastically dominant over B if and only if:

$$\int_{-\infty}^x F_A(u)du \leq \int_{-\infty}^x F_B(u)du, \forall x \quad (12)$$

In other words, A is considered second-order stochastically dominant over B if and only if the area under the curve for A is smaller than B. Let a denote $F^{-1}(0)$ and b denote $F^{-1}(1)$ for a distribution of delays or headways (see Fig. 4(b)). Therefore, the SSD Index (SSDI) is defined as the adjusted second-order stochastic dominance of the distribution of delays/headway deviations as follows:

$$SSDI = \frac{\int_a^0 F(x)dx + \int_0^b (1 - F(x)) dx}{\text{Average Scheduled Headway}} \quad (13)$$

where $a < 0$ and $b > 0$. Note that the boundaries of the integrations should be adjusted if needed for other shapes of CDF—for example, where the CDF is completely on the right or left side of the y axis in Fig. 4(b). SSDI is a unitless measure that is always equal to or greater than zero. The smaller the SSDI is at a specific stop, the more reliable is the service at that stop.

Similar to the WI, the SSDI is capable of distinguishing between the cost of being early and the cost of being late. Following is a modified formulation for SSDI when distinguishing between these costs:

$$SSDI = \frac{\alpha \int_a^0 F(x)dx + \beta \int_0^b (1 - F(x)) dx}{\text{Average Scheduled Headway} \cdot (\alpha + \beta)} \quad (14)$$

where α is the weight associated with the cost of being late, and β is the weight associated with the cost of being early. The SSDI is particularly useful when two or more distributions have similar width but different earliness indices. It is also useful to differentiate between distributions with similar width and similar earliness index but with a different curvature, which represents different unreliability characteristics. For example, the following two distributions, shown in Figure 5, have identical width indices and earliness indices. SSDI helps to differentiate between these two distributions by taking the area under and above the curves as formulated in Eqs.

(13) and (14) and identify the distribution that is closer to the theoretical ideal distribution. In this case, the bus service associated with the dashed distribution curve is more reliable than the bus service associated with the solid distribution curve. The SSDI can also help distinguish between the distributions shown in Figure 3 in which the standard deviations were the same.

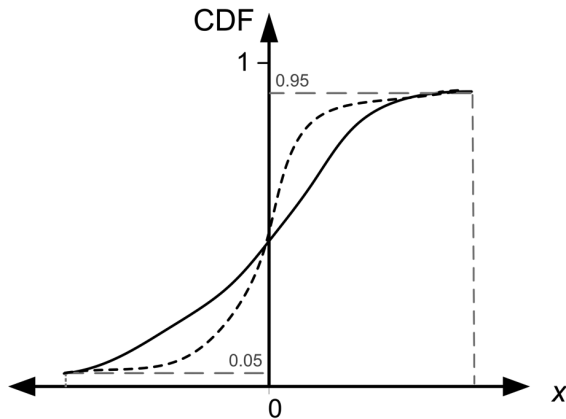


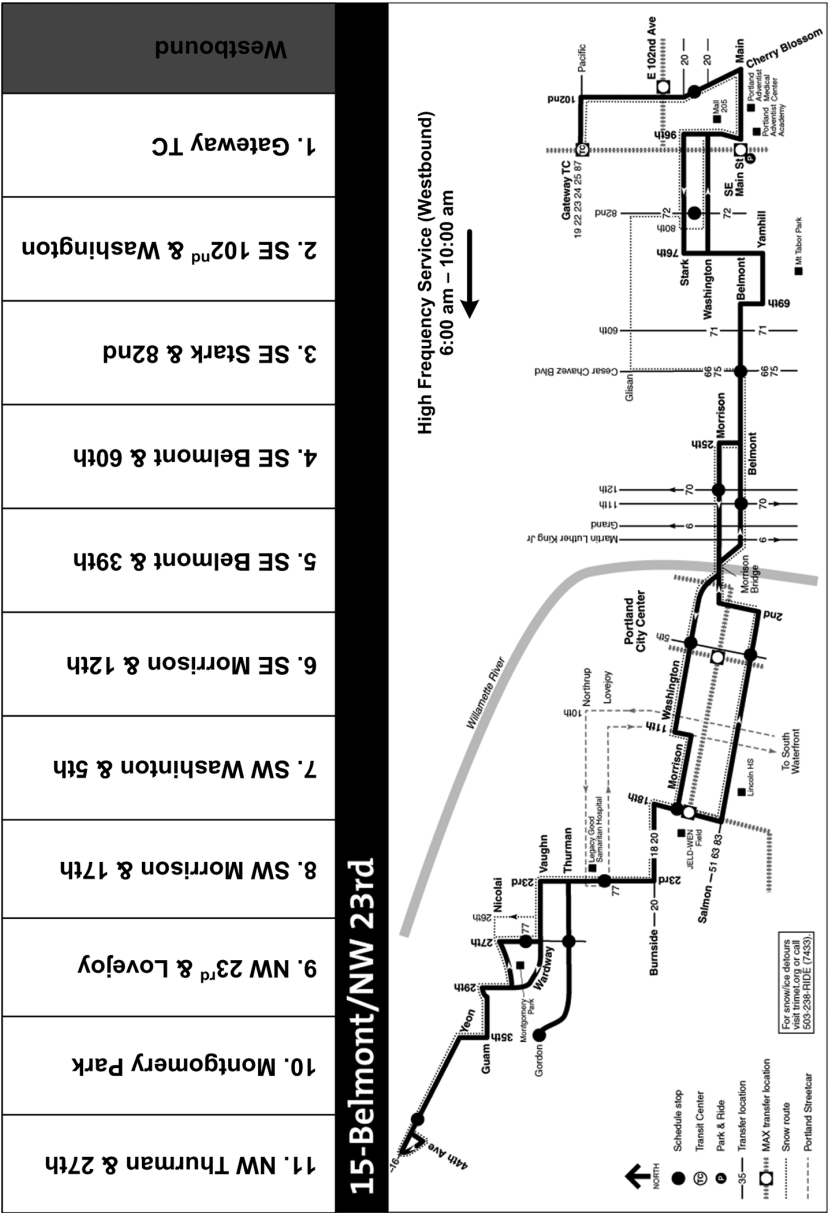
Figure 5. Illustration of usefulness of SSDI when two distributions have identical earliness and width indices

Therefore, the delays/headway deviations distribution can be characterized by three characteristics of this distribution, namely, zero percentile rank, width, and adjusted second-order stochastic dominance. In terms of reliability, the larger EI, WI, and SSDI are, the less reliable bus service is.

Data Analysis Results

For our numerical analysis, we consider the Portland TriMet transit service. More specifically, Route 15 westbound is chosen for this study. This is a heavily-used route that runs through southeast Portland toward downtown during the morning commute period. During the weekday morning and afternoon rush hours, buses on this route run every 15 minutes or better. Route 15 connects Montgomery Park, NW Portland, Portland City Center, SE Portland, and Gateway via Vaughn, Burnside, Washington/Salmon, Belmont/Morrison, Stark/Washington, and 102nd (see Figure 6).

Three major stops were selected for a preliminary analysis: SE Morrison & 12th, SW Morrison and 17th, and SE Stark & 82nd. These specific stops at different time intervals were selected to cover a broad range of geographical locations (in terms of proximity to downtown Portland), demand, and congestion levels throughout the entire



Source: Feng and Figliozi 2011

Figure 6. Route 15 map

route. Archived BDS data for Route 15 from 11/30/2009 to 5/23/2010 including more than 115 weekdays were used. For the selected stops, general descriptive statistics along with the TCQSM reliability measures were computed, as shown in Table 1. The TCQSM reliability measure, including on-time performance and headway adherence, were used to determine the LOS at the selected stops. As can be seen, the SW Morrison & 17th (2-3 PM) and SE Stark & 82nd (9–10 AM) stops had an LOS of F for their infrequent services while the SW Morrison & 17th (9–10 AM) stop had an LOS of E and the SE Morrison & 12th (6–7 AM) stop had an LOS of A for their frequent services.

Current reliability measures do not tell the whole story of the service reliability. To improve the service reliability, transit agencies should know more than just the percentage of buses that are considered “on-time” and the coefficient of variation of headway deviations. The consequences of late buses may be much different than those of early buses. More information, such as how late or early buses arrive at the stop or how much they are ahead or behind the schedule, is needed to effectively improve schedules and operations strategies. Therefore, we recommend that additional measures, such as those presented in this paper, be used as guidance in discussions of reliability.

Table 1. Descriptive Statistics and TCQSM Reliability Measures for the Selected Stops

Service	Stop	Time	Total Number of Trips	Average Scheduled Headway (sec)	Average Headway Deviation (sec)	Stdv. of Headway Deviation (sec)	Headway Adherence	LOS
Frequent	SW Morrison & 17th	9–10 AM	917	403	5.27	241.69	0.60	E
	SE Morrison & 12th	6–7 AM	509	600	-8.76	124.12	0.21	A
	Stop	Time	Number of Trips	Average Scheduled Headway (sec)	Average Delay (sec)	Stdv. of Delay (sec)	On-time Performance	LOS
Not Frequent	SW Morrison & 17th	2-3 PM	404	2100	318.64	264.90	0.53	F
	SE Stark & 82nd	9–10 AM	603	735	80.33	145.98	0.63	F

We carried out an empirical analysis to assess the proposed reliability measures EI, WI, and SSDI for the same selected stops and time periods of Route 15, as shown in Table 1. For frequent services, delay distribution was used and for less frequent (or infrequent) services, headway deviation distribution was used as described in the previous sections.

Figure 7 shows the cumulative distributions of delays and headway deviations for frequent and infrequent service at the selected stops and time periods.

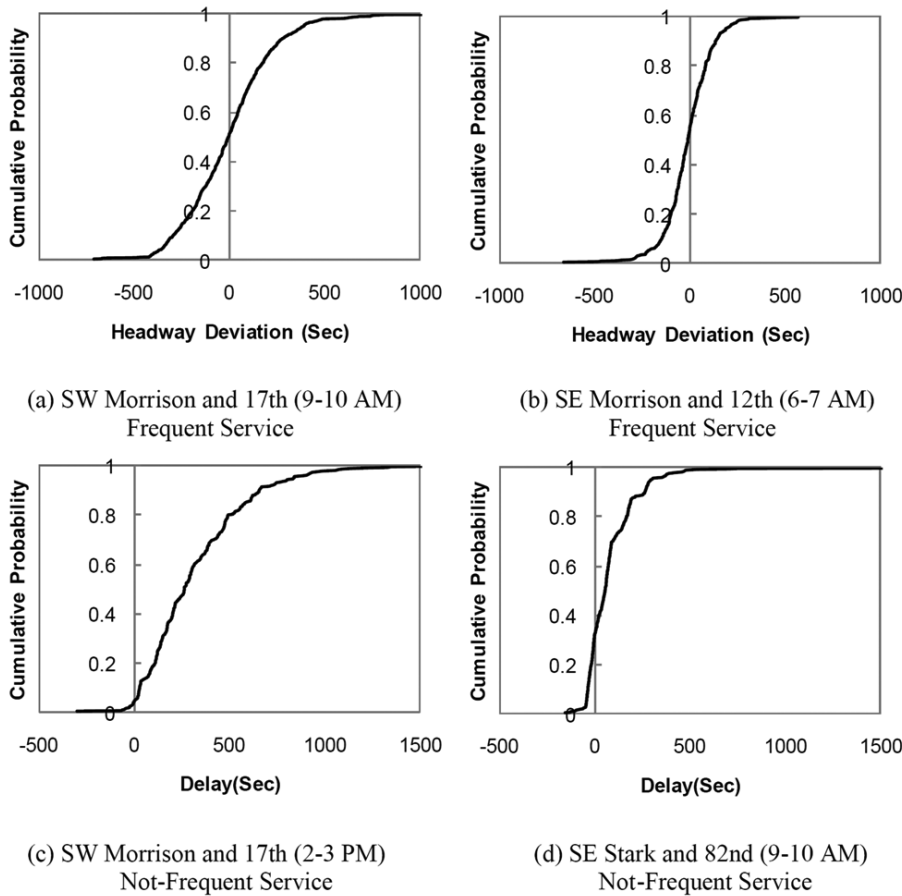


Figure 7. Cumulative distributions of delays and headway deviations for frequent and not frequent services at selected stops and time periods of Route 15

The cumulative distributions of headway deviations shown in Figures 7(a) and 7(b) represent the studied frequent service at SW Morrison and 17th (9–10 AM) and SW Morrison and 12th (6–7 AM) stops. The cumulative distributions of delays shown in Figures 7(c) and 7(d) represent the studied less-frequent service at SW Morrison and 17th (2–3 PM) and SE Stark and 82nd (9–10 AM) stops. Note that TriMet allows its buses to depart up to one minute ahead of schedule. This may explain a large portion of the 30 percent that are not on-time due to early departures at the SE Stark and 82nd (9–10 AM) stop.

As an illustration, Table 2 summarizes the computed reliability measures EI, WI, and SSDI associated with each service at each stop along with its headway adherence and on-time performance. The EI for the less frequent service at SW Morrison and 17th is as low as 0.04, while the EI at SE Morrison and 12th is as large as 0.54. In other words, only four percent of buses at SW Morrison and 17th depart early between 2–3 PM, while 54 percent of buses depart early at SE Morrison and 12th between 6–7 AM. Also, the bus service at SW Morrison and 17th has a WI as large as 1.83 between 9–10 AM, while the WI at the same stop between 2–3 PM is only 0.39. This clearly shows the changing pattern of service reliability in different times of the day. Similarly, the service at SW Morrison and 17th between 9–10 AM has the largest SSDI compared to the other studied services. Note that the SSDI at SW Morrison and 17th reduces from 0.8 between 9–10 AM to 0.15 between 2–3 PM. Overall, the service at SE Morrison and 12th at between 6–7 AM has the worst reliability in terms of EI and SW Morrison and 17th between 9–10 AM has the worst reliability in terms of WI and SSDI.

Table 2. New Reliability Measures at Stop Level

Service	Stop	Time	EI	WI	SSDI	On-time Performance	Headway Adherence
Frequent	SW Morrison & 17th	9–10 AM	0.5	1.83	0.80	-	0.60
	SE Morrison & 12th	6–7 AM	0.54	0.67	0.15	-	0.21
Not Frequent	SW Morrison & 17th	2–3 PM	0.04	0.39	0.15	0.63	-
	SE Stark & 82nd	9–10 AM	0.32	0.46	0.14	0.53	-

To further explore the applicability of the proposed measures in the real world, a more comprehensive analysis was carried out using the same archived BDS data for Route 15 westbound including more than 115 weekdays for 11 AM to 12 PM. The new reliability measures are applied to the whole route to highlight stops that are candidates for operational improvements such as implementing holding strategy, expressing, schedule adjustment, and re-routing. Figure 8 provides a comparison of the proposed

reliability measures with headway adherence at each stop, without differentiating between the cost of being early versus late. Note that travel direction is from Gateway Transit Center to NW 23rd & Lovejoy stop. The last two stops were removed from our analysis due to their geometry, which is inconsistent with the rest of the route.

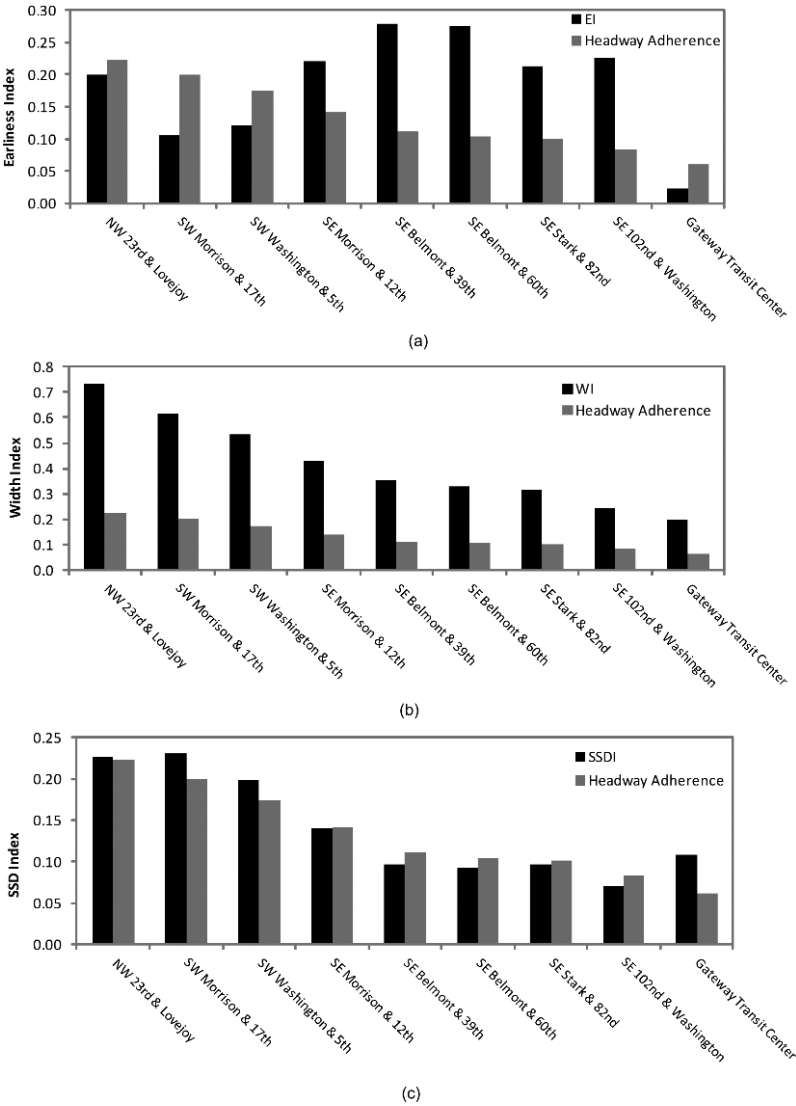


Figure 8. Comparison of (a) Earliness Index; (b) Width Index; and (c) Second Order Stochastic Dominance Index with headway adherence for Route 15 westbound (travel direction from right to left)

The stops with the highest (worst) EI are SE Belmont & 39th and SE Belmont & 60th. Figure 8(a) suggests that there is no correlation between the EI and headway adherence. Also, as shown in Figure 8(b), the stops with the highest (worst) WI are NW 23rd & Lovejoy and SW Morrison & 17th. The WI is following a clear trend throughout the route. As one would expect, the variability of headway adherences are increasing as they move toward the end of the route. This implies the existence of a correlation between consecutive stops in a route. Similarly, the stops with the highest (worst) SSDI are NW 23rd & Love Joy and SW Morrison & 17th. However, the trend in SSDI is not as clear until the middle of the route, where SSDI starts increasing constantly as it moves toward the end of the route.

In this study, an overall consistency was observed among the WI, SSDI, and headway adherence, as expected. To eliminate the impacts of the existing correlation between consecutive stops along the route and quantitatively demonstrate the difference between the proposed measures and headway adherence, a further analysis was performed. Figure 9 shows the percentages of relative difference of selected measures between consecutive stops. The graph shows that SSDI, WI, and headway adherence are capturing different levels of unreliability at each stop relative to its previous stop, despite the general consistency among them. For example, at SE Morrison and 12th (relative to SE Belmont and 60th), the SSDI captures more unreliability compared to WI and headway adherence, while at SE 102nd & Washington (relative to SE Belmont and 60th), the SSDI captures less unreliability compared to WI and headway adherence. To verify this, further research using data from other routes and time periods is required to study situations where unexpected consistency or inconsistency of the proposed measures arises.

Figure 10 illustrates how the EI and On-Time Performance provide different images of service reliability. As shown, On-Time Performance decreases as it moves toward the end of the route, whereas the EI does not follow a specific trend. As explained earlier, arrivals no more than five minutes after the scheduled time and no more than one minute early are considered on-time based on the TCQSM definition, whereas the EI highlights the service unreliability due to the early buses only.

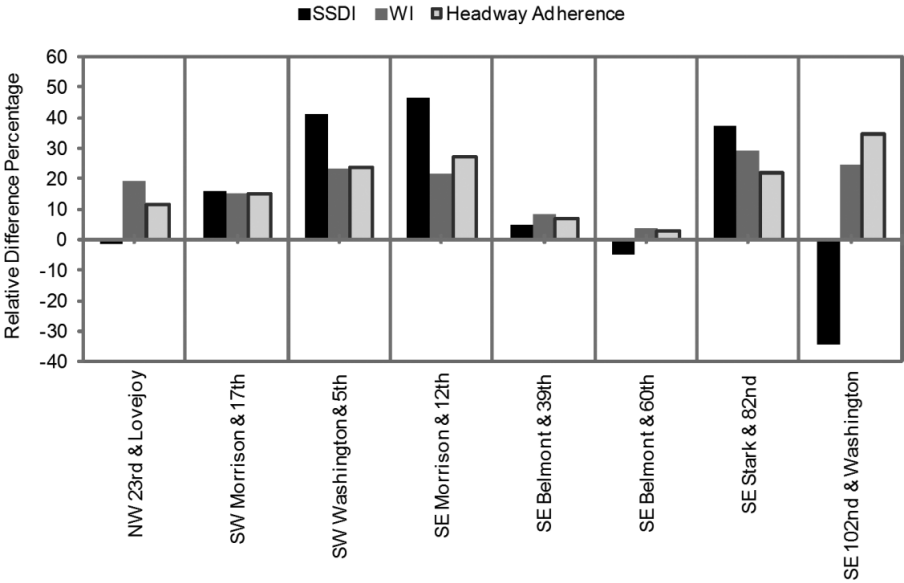


Figure 9. Percentage of relative difference of selected measures between consecutive stops; first stop is Gateway Transit Center (travel direction from right to left)

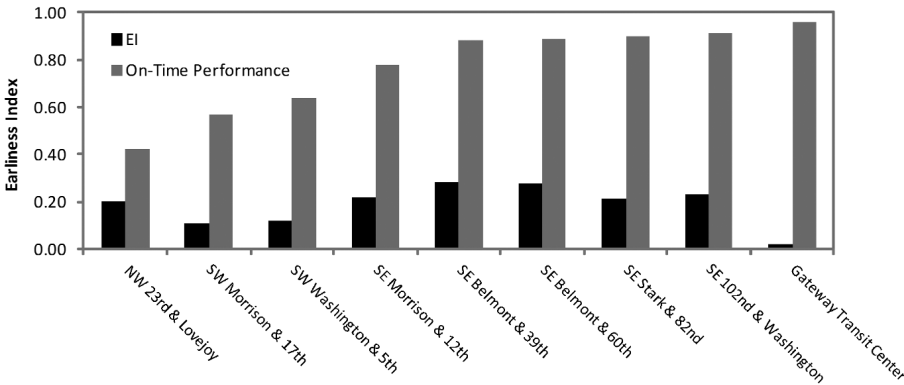


Figure 10. Earliness Index vs. On-Time Performance for Route 15 westbound (travel direction from right to left)

Conclusion and Future Research

This paper evaluated current measures of bus service reliability, specifically the current TCQSM measures, and developed alternative reliability measures at the stop level. Note that the proposed reliability measures are not suggested as replacements for the existing measures; rather, they are complementary. We investigated the distribution of delays and headway deviations on the basis of empirical archived BDS data from Route 15 in Portland, Oregon, for more than 115 weekdays in 2009–10. Findings are summarized below:

For frequent services, the distribution of headway deviations can be used to measure the percentage of buses that are ahead of the schedule or early. For infrequent services, the distribution of delays should be used. The Earliness Index is proposed.

The distribution of delays/headway deviations is often wide. Commonly-proposed reliability measures in the TCQSM only use the standard deviation and do not take into consideration the full width of this distribution. Also, existing measures do not differentiate between the cost of being early versus late. The Width Index is proposed to capture more of the extreme cases of unreliability and differentiate between the costs of being early versus late.

To further capture characteristics of the distribution of delays/ headway deviations, the concept of second order stochastic dominance is used. As a result, second order stochastic dominance index is proposed. This index is particularly useful when two or more distributions have similar width but different earliness indices. It is also useful to differentiate between distributions with similar width and similar earliness index but with a different curvature that represents different unreliability characteristics.

It was found for the studied route and time period that the Width Index and coefficient of variation (or headway adherence) were fairly consistent, whereas the second order stochastic dominance index captured more unreliability in some cases. The Earliness Index provides reliability information that the coefficient of variation does not capture. Eliminating the impacts of the existing correlation between consecutive stops along the route, the difference between the proposed measures and coefficient of variation are quantitatively demonstrated. To further verify the findings and study the situations where consistency or discrepancy of the proposed measures arises, more research using data from other routes and time periods is required.

Transit agencies can use the proposed reliability measures to evaluate and prioritize stops for operational improvement purposes, such as bus holdings or schedule adjustments. This paper exhibits how reliability can vary across stops and how important this variation is in prioritizing stops for improvements. A key topic for future research is defining LOS thresholds for bus service reliability. Furthermore, studying the characteristics of the distribution of delays and headway deviations using more data at stop, route, and network level is required to gain more knowledge of transit service reliability. Causes of service unreliability are also important to investigate.

Acknowledgments

The authors gratefully acknowledge the contributions of Josh Crain from DKS Associates and Charlotte Frei and Madison Fitzpatrick from Northwestern University for their review and feedback. The authors also acknowledge the TriMet and PORTAL team at Portland State University for providing the data. The authors would like to thank Stefan Voss and the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

References

- Abkowitz, M., H. Slavin, R. Waksam, L. Englisher, and N. Wilson. 1978. Transit service reliability. TSC Urban and Regional Research Series, U.S. Department of Transportation, Cambridge.
- Ap. Sorratini, J., R. Liu, and S. Sinha. 2008. assessing bus transport reliability using micro-simulation. *Transportation Planning and Technology* 31(3): 303–324.
- Bellei, G., and K. Gkoumas. 2010. Transit vehicles' headway distribution and service irregularity. *Public Transport* 2(4): 269–289.
- Bertini, R. L., and A. El-Geneidy. 2003. Generating transit performance measures with archived data. *Transportation Research Record* 1841: 109–119.
- Bertini, R. L., and A. El-Geneidy. 2004. Modeling transit trip time using archived bus dispatch system data. *Journal of Transportation Engineering* 130(1): 56–67.
- Bowman, L. A., and M. A. Turnquist. 1981. Service frequency, schedule reliability and passenger wait times at transit stops. *Transportation Research Part A* 15(6): 465–471.

- Camus, R., G. Longo, and C. Macorini. 2005. Estimation of transit reliability level-of-service based on automatic vehicle location data. *Transportation Research Record* 1927: 277–286.
- Chen, X., L. Yu, Y. Zhang, and J. Guo. 2009. Analyzing urban bus service reliability at the stop level, route, and network levels. *Transportation Research Part A* 43: 722–734.
- Duarte, A., C. Garcia, G. Giannarakis, S. Limão, A. Polydoropoulou, and N. Litinas. 2010. New approaches in transportation planning: Happiness and transport economics. *Netnomics* 11: 5–32.
- El-Geneidy, A., J. Horning, and K. Krizek. 2010. Analyzing transit service reliability using detailed data from automatic vehicular locator systems. *Journal of Advanced Transportation* 45(1): 66–79.
- Feng, W., and M. A. Figliozzi. 2011. Using archived AVL/APC bus data to identify spatial-temporal causes of bus bunching. *Proceedings of the 90th Transportation Research Board Annual Meeting*, Washington, D.C.
- Fu, L., and X. Yaping. 2007. A new performance index for evaluating transit quality of service. *Journal of Public Transportation* 10(3): 47–70.
- Furth, P., B. Hemily, T. Muller, and J. Strathman. 2003. Uses of archived AVL-APC data to improve transit performance and management: Review and potential. TCRP Web Document No. H-28, Washington, D.C.: Transportation Research Board.
- Furth, P., and T. Muller. (2006. Service reliability and hidden waiting time: Insights from automatic vehicle location data. *Transportation Research Record* 1955: 79–87.
- Highlights of the *Transit Capacity & Quality of Service Manual: 1st Edition* (Kittleson & Associates, Inc.) 1999. Research Result Digest, Transit Cooperative Research Program.
- Highway Capacity Manual*. 2000. Transportation Research Board.
- Highway Capacity Manual*. 2010. Transportation Research Board.
- Kepaptsoglou, K., and M. G. Karlaftis. 2009. The bus bridging problem in metro operations: Conceptual framework, models and algorithms. *Public Transport* 1: 275–297.

- Kimpel, T. J. 2001. Time point-level analysis of transit service reliability and passenger demand. Ph.D. dissertation, Portland State University.
- Lin, J., P. Wang, and D. T. Barnum. 2008. A quality control framework for bus schedule reliability. *Transportation Research Part E* 44: 1086–1098.
- Transit Capacity and Quality of Service Manual, Second Edition*. 2004. Transportation Research Board.
- Tumlin, J., J. Walker, J. Hoffman, and R. Hutabarat. 2005. Performance measures for the urban village transit network. *Proceedings of the 84th Transportation Research Board Annual Meeting*, Washington, D.C.
- Van Oort, N., J. W. Boterman, and R. Van Nes. 2012. The impact of scheduling on service reliability: Trip-time determination and holding points in long-headway services. *Public Transport* 4(1) 39-56.
- Xin, Y., L. Fu, and F. Saccomanno. 2005. assessing transit level of service along travel corridors using TCQSM—A case study. *Proceedings of the 84th Transportation Research Board Annual Meeting*, Washington, D.C.

About the Authors

MEEAD SABERI (meead@u.northwestern.edu) is a Ph.D. candidate in Transportation Systems Analysis and Planning at Northwestern University, Evanston, IL. He received his M.S. degree in Transportation Engineering at Portland State University, Portland, OR. His main research interests are traffic flow theory and characteristics, urban network modeling, transportation data management, and public transit.

ALI ZOCCAIE K. (ali-zockaie@u.northwestern.edu) is a Ph.D. candidate in Transportation Systems Analysis and Planning at Northwestern University, Evanston, IL. He received his M.S. degree in Transportation Engineering at Sharif University of Technology, Tehran, Iran. His main research focus is on network modeling and reliability, specifically dynamic network assignment of travelers with heterogeneous risk preferences. His other research interests include public transit, network traffic theory, and airline operations.

WEI FENG (wfeng@pdx.edu) is a Ph.D. candidate in Transportation Engineering at Portland State University, Portland, OR. He received his M.S. degree in Transportation Engineering at Beijing Jiaotong University, Beijing, China. His research interests

include traffic flow theory and characteristics, transportation operations, intelligent transportation systems, public transportation, and freight transportation.

AHMED EL-GENEIDY (ahmed.elgeneidy@mcgill.ca) is an associate professor at the School of Urban Planning, McGill University. Ahmed's research interests include land use and transportation planning, transit operations and planning, travel behavior analysis including using motorized and non-motorized modes of transportation, and measurements of accessibility and mobility in urban contexts. Ahmed has a special interest in transportation needs of disadvantaged populations. He received B.S. and M.S. degrees from the Department of Architectural Engineering at Alexandria University, Egypt, and continued his academic work at Portland State University, where he received a Graduate GIS Certificate and earned a Ph.D. in Urban Studies from Nohad A. Toulon School of Urban Studies and Planning. After finishing his Ph.D. he moved to the Twin Cities to work as a post-doctoral research fellow at the University of Minnesota. In 2007 Ahmed moved to Montreal, Canada to start his current position at the School of Urban Planning, McGill University.