1. Introduction

This literature review provides an overview of literature that has been written on the use of performance standards in transit planning. The intent of this review is to provide a comprehensive review of contemporary research on the subject, focusing on the main themes and debates in the field, and also to identify areas where additional research would add value to the field. This review will support my paper, Assessing Transit Service Performance: Recommended Performance Standards for the Santa Clara Valley Transportation Authority.

Section 2 explores the main themes and debates related to the use of performance standards in transit. Section 3 summarizes the important concepts from Section 2 and also identifies areas where additional research would be of value. Additionally, this section describes how the research to be conducted for this project fits in with the research conducted by others. Section 4 is an annotated bibliography of the research consulted for this literature review; the annotation notes whether the source is from a peer-reviewed journal. Section 5 lists the online databases used to search for pertinent literature. For each database, the relevant keywords are listed.

2. Main Themes and Debates

Public transportation agencies in the United States receive a considerable portion of their income from taxpayer dollars — often more than half of an agency’s operating budget comes from general taxes. One might ask why the general public is willing to subsidize transit to such a high degree, given the dominance of car travel over transit. A literature review by Bhatta and Drennan found that most studies show that investment in public transit yields benefits to the community in the form of increased output, increased productivity, lower production costs, higher income, higher property values, higher employment, and reduced noncommercial travel time. These communal benefits might give an indication as to why the public supports transit at the high level that it does. Thompson and Matoff point out that large regional transit agencies owe much of their existence to the tax payments of the wealthy suburbs – precisely those people who are unlikely to use public transit. Policy analysts also believe the expenditure of public funds on transit is justified due to a series of market failures which have caused the private market to undersupply transit service. Since they are so heavily subsidized by public funds, they are under constant pressure to improve the efficiency of their operations by providing the best possible service at the lowest possible cost.

This push towards efficiency has prompted many transit managers to adopt performance measures in an effort to improve efficiency. Indeed, a Transit Cooperative Research Program (TCRP) survey by Benn found that performance measurement usage is growing steadily over time. Theoretically, these performance standards allow transit managers to objectively assess the performance of their services and make better decisions regarding allocation of resources. In reality, considerable disagreement exists concerning the definition of performance measures themselves, as well as how they are used. In effect, most researchers agree on the need to quantify performance, but they do not agree on how to do so. Many researchers suggest that resource allocation decisions should be made based on transit performance indicators, though decision-makers have given this idea tepid support at best. Indeed, a 2004 TCRP report found that there has been no increase in the trend to use performance measures to allocate funds.

The following sections present the main themes and debates in transit performance standards research. In reviewing the research, seven themes emerged as the obvious issues and questions in the field. Each section will present the relevant research on the issue and discuss the differing results. This research focuses on transit performance standards throughout the United States in order to minimize the variability of differing regulatory environments, differing social norms, and differing funding mechanisms. However, there are a few research projects from abroad that do offer interesting advice and conclusions.

A. The Menu of Performance Measures

Transit agencies use a plethora of performance measures to help them assess performance. A study by Phillips categorized a menu of no less than 222 different performance measures in use throughout the industry. Other studies haven’t found quite as many measures (Phillips’ methodology was much more inclusive and detailed), though they do confirm the existence of an abundance of measures. Benn found the use of various different measures across the 111 agencies who responded to his survey. These performance measures span a range of concepts such as cost efficiency, service consumption, service effectiveness, and customer satisfaction, among many others. With so many performance measures in existence, one would think that the industry could readily evaluate its performance, though the reality is quite the opposite.

A transit agency can find itself lost amongst the various measures available for use. A TCRP guidebook by Kittelson & Associates, Inc. et al. recommends an agency limit itself to using a maximum of twenty performance measures. The study found one transit agency that used 150 performance measures, resulting in a “plurality of unfocused, misdirected

---

activities.” Indeed, in their review of industry practices, De Borger et al. learned that there are no “generally accepted” set of measures in the industry.

The research does identify the most commonly used performance measures. Lem et al. identified the most common performance indicators used in California transit agencies as 1) operating cost per revenue vehicle hour, 2) operating cost per passenger boarding, 3) farebox recovery per operating cost, 4) passenger boardings per revenue vehicle mile, and 5) passenger boardings per revenue vehicle hour. A study by the Federal Transit Administration (FTA) concurs with this finding and expands the geography, pointing to the exact same five indicators as being the most typically used throughout the nation. Among these indicators, several studies point to the one that stands out among the rest because of its wide use: passenger boardings per revenue hour. This standard was found to be the single most important measure in use today by Urbitran, Inc. in its 2006 study for TCRP. Similarly, a University of South Florida research report found that 90% of the 63 agencies surveyed used this performance measure. Howard Benn found a similar result in his survey, which he attributes to the fact that wages (which are closely tied to hours of service) constitute 80% of the typical cost to provide transit service.

### B. Types of Performance Measures

Transit performance indicators each attempt to measure a certain objective or goal. Some measures attempt to quantify an agency’s cost efficiency, some attempt to measure customer satisfaction, and others attempt to measure how effective the service is at attracting riders. There is no shortage of research which has attempted to categorize transit performance measures. Grava groups all performance measures into three categories: effectiveness, reliability, and cost. Lem et al. settle on a different set of three: cost efficiency, cost effectiveness, and service effectiveness. Judging from the differing approaches in the body of literature, it appears that there are countless ways to group performance measures. One study in particular, however, gives a particularly convincing argument.

Perhaps the most comprehensive description of the types of performance measures is given by Gleason and Barnum. Their study aims to provide clarity to the often-misinterpreted use of performance measures. They point out that it is quite common for

---

25 Mark Mistretta, Jay A. Goodwill, Rob Gregg and Christopher DeAnnunzio, Best Practices in Transit Service Planning, Center for Urban Transportation Research, University of South Florida (Tampa: Florida Department of Transportation Research Center, 2009): 55.
performance terms to be misunderstood and thus used incorrectly. Terms like “efficiency,” “effectiveness,” and “productivity” are often incorrectly used interchangeably, which is concerning because the performance measures that rely on them are increasingly used for resource allocation decisions. For example, they point out that the indicator “cost per passenger” is considered an effectiveness indicator in one paper and both an effectiveness and efficiency indicator in another. Misusing these terms can lead to skewed understand of the true performance of transit service, resulting in a misallocation of funds. Gleason and Barnum offer clear direction to understand transit performance terms.19

Gleason and Barnum explain that efficiency indicators should be ratios of service inputs and service outputs. For example, vehicle hours per employee or cost per passenger are efficiency indicators that help to illustrate how efficiently an agency is providing its service. Just as we can measure the efficiency of a car engine by its input (gallons of gas) and its output (miles) in an indicator we call “miles per gallon,” we can similarly measure transit service by its inputs and its outputs.20

Effectiveness indicators, Gleason and Barnum argue, are much more widely misunderstood. Effectiveness indicators should not be ratios and should generally not include cost elements. These indicators should be absolute-level indicators which indicate how well agency goals are being met and, as such, should be tied directly to the goals of the agency. The authors point out that “in every case known to the authors, the goal is to maximize ridership.” As such, usually the service options that serve more riders would be considered more effective. However, Gleason and Barnum are correct in pointing out that their perspective on effectiveness and efficiency is not widely held.21

Contrasting research looks at effectiveness differently than Gleason and Barnum. Research by Li and Wachs, for example, uses effectiveness indicators that violate both principles that Gleason and Barnum warn against — they are ratios and they include costs. An example of one of their proposed measures is their cost effectiveness indicator which relates inputs and service consumption.22 This relationship is an opposing perspective to the Gleason and Barnum principle. Gleason and Barnum would argue that the measure of effectiveness should not use the inputs but rather focus entirely on service consumption.

C. Limitations of Existing Performance Measures (Not Multimodal)

Most of the performance measures in use today were developed in the post-World War II “bus era” when bus service was the dominant mode in transit service.23 Rail was not a serious alternative to bus service until the 1980’s, nor were new modes like bus rapid transit or paratransit yet in existence. The relative homogeneity of bus service (in terms of typical costs, service delivery, goals, etc.) meant that transit performance measures were developed with little regard to other modes. This has led to several problems with the performance measures that most agencies still use today.

20 Ibid., 382-384.
Research from Lem et al. found that almost all performance indicators do not include capital costs in their cost components, which limits their ability to be used across modes. Indicators typically consider only the operating costs for a service and ignore the capital costs. This conventional way of looking at performance makes sense to transit agencies because capital costs for a project are usually funded largely through outside sources such as Federal Transit Administration grants or state programs. As a result, agencies are understandably less concerned about capital costs for which they are not responsible for and are more concerned about the ongoing operating costs for which they are responsible.24

Today’s transit industry, however, is becoming more multi-modal, which is a problem for performance measures that do not incorporate capital costs. As the transit industry embraces light rail, commuter rail, paratransit, and bus rapid transit, improvements to the existing performance measures are needed in order to provide accurate cross-modal comparisons. These sorts of comparisons across modes would not be possible with the usual performance measures due to the differing reliance on capital costs between modes. Typical performance measures would unduly favor modes that are less labor intensive, such as light rail (due to its higher reliance on capital costs instead of labor). Research by Lee and Wachs also identified the same problem. They argue that even though capital costs often come from a non-local source, they still are public funds and should be considered if the goal is to improve the resource allocation of all public funds, no matter the source.25

The Lee and Wachs research also points out that existing performance measures do not account for differing capacities between modes, which leads to inaccurate measurements. They make comparisons between modes that can have very different capacities. A light rail vehicle, for example, can have twice the capacity of a standard bus. Even within a single mode, vehicles can have different capacities (such as a small bus when compared to a large bus). In order to accurately compare performance across services with different capacities, performance measures must incorporate vehicle capacity.26

To remedy these deficiencies in the current performance measures, Lem et al. suggest new measures that incorporate both capital costs and vehicle capacities.27 Similar research by Li et al. suggests the same improvements.28

D. One, or More Than One?

The body of research in the field illustrates a disagreement about the number of performance measures an agency should utilize. Mistretta et al. looked at the industry’s “best practices” and suggests that using multiple measures has been most effective for many

24 Ibid., 2.
27 Lewis Lee Lern, Jian-Ling Li and Martin Wachs, Comprehensive Transit Performance Indicators, Institute of Transportation Studies (Los Angeles: University of California at Los Angeles, 1994): 3.
agencies. Barnum et al. also stress the importance of using multiple performance measures which are closely linked to agency objectives. Many agencies have adopted this approach. The Milwaukee County Transit Authority (MCTA), for example, has identified five agency objectives and has developed over a dozen performance measures which are linked to the objectives. However, the MCTA has also developed a new, singular overarching performance measure in an attempt to simplify performance evaluation. Their measure, called the productivity frequency index, is an attempt to move in the direction of using a singular standard to compare service. Khasnabis et al. would applaud a move in this direction, because they believe (in contrast to previous literature) that using multiple measures will undoubtedly lead to conflicting results. For this reason, they suggest using one single metric that can comprehensively evaluate all service. Clearly, there is no agreement on whether a singular indicator is best. Grava convincingly weighs in and counters that “given the complex situation in providing the services and the many stakeholders who have an interest in the effort, it is not likely that such a magic indicator will ever be identified.”

E. Relation to Service Objectives and Goals

Most research points out the need to clearly define the agency’s service goals and objectives before setting out to adopt performance standards. The standards, most agree, need to be tightly related to the performance measures in order to be meaningful. Kittelson & Associates et al., in their Guidebook for Developing a Transit-Performance Measurement System, identify the very first step in the process as defining clear and objective goals. Gleason and Barnum took this idea a bit further, relating the multitude of performance measures to the multitude of transit agency goals and objectives, concluding that there are no universally appropriate measures. Multiple measures must be used, they argue, in order to meet multiple objectives.

Sheth et al. also believe in tight integration between standards and the goals of an agency, but they look at the issue from the perspective of multiple stakeholders. They argue that when formulating performance measures, it is important to consider three perspectives – the provider (agency), the user, and the society at large. Each stakeholder will approach performance from a different perspective. The provider will have an economic perspective and wants to minimize resources used for transit service. The user will have a passenger perspective and wants to maximize the number of miles of service delivered. Finally, society will have a communal perspective and wants to minimize externalities associated with

---

29 Mark Mistretta, Jay A. Goodwill, Rob Gregg and Christopher DeAnnuntis, Best Practices in Transit Service Planning, Center for Urban Transportation Research, University of South Florida (Tampa: Florida Department of Transportation Research Center, 2009): 55.
31 Milwaukee County Transit Authority, Public Transit Service Objectives and Standards, (Milwaukee: Milwaukee County Transit Authority, 2005): 2.
transit service. Researchers that do discuss the issue generally agree that it is important for transit agencies to seek an alternative perspective. Miller et al. went a step further, categorizing 70 broad objectives into three stakeholder perspectives, though they consider the passenger perspective most paramount.37

There is general, though not universal, agreement regarding what objectives are most important to an agency. De Borger et al. argue that social welfare maximization (improving the well-being of society as a whole) is the most important objective, indicating that they value the community perspective highly.38 Most other researchers disagree, preferring a set of objectives that have maintained their permanence over time. MCTA has a fairly standard and un-surprising set of five goals, including simple principles such as “the system should be efficient as a whole.”39 Goals such as safety, comfort, and reliability consistency rate highly in literature.40 These are important goals, because they focus on the riders. Miller et al. reminds us that passengers “are the raison d’etre of transit travel and their perceptions and needs are central.”41 Grava puts the riders first by suggesting that it may be possible to throw out all existing performance measures and adopt a single new one: the number of rider complaints satisfied.42 His idea is interesting, because it would lead agencies to focus on making repeat customers out of riders. The Kittelson study offers some advice on the subject of keeping riders happy, noting that it takes five times the cost to attract a new rider than it does to keep an existing one.43 Kuby et al. offer yet another perspective — the most important goal is to reduce travel time. Their study of factors that influence light rail boardings showed that travel time is the principal concern for attracting riders.44

F. Setting Benchmarks
Once performance measures are set, it will be necessary for an agency to decide what benchmark value that each evaluated service must meet in order to be considered successful. This is no easy task, nor is it well-researched. Mistretta et al. offer some insight via their 63-agency survey. They found that a common method for setting the benchmark is to set it at 50% of the average performance of all services in the category. They found many agencies that used 60% with the same method.45 Barnum et al. offer another method — set...
the benchmark at the 10% percentile among all the services. This way, any route in the bottom 10th percentile will be considered substandard.46 The most comprehensive review of methodologies to set the benchmark comes from Kittelson & Associates. They identify six common methods, including comparison to the average, comparison to a baseline, trend analysis, and comparison to peers.47 The research on the subject suggests that there are plenty of methods that an agency can consider using, and can probably achieve useful results with any number of them.

G. Using Data Envelopment Analysis

There exists a new and promising performance analysis method that is gaining popularity in the research field. Data Envelopment Analysis (DEA) is a mathematical method that is used in many fields to evaluate productive efficiency of a set of “peer entities.” In other words, it is an economic method that measures efficiency across peers. DEA is generic, flexible, and is used across many fields to evaluate entities engaging in a countless number of activities such as baseball players hitting home runs or airlines carrying passengers.48 The process works by analyzing the ratio of inputs to outputs for each entity (called a Decision Making Unit or DMU) and ranking the entities based on efficiency. In a public transit context, the DMUs to be compared can be individual routes, corridors, or even agencies. For example, DEA could analyze a set of park and ride lots based on their inputs (number of parking spaces and operating expenses) and their outputs (number of parked cars and parking revenues). One of the appealing features of DEA is that there is no need to assign weights to inputs and outputs. As long as each DMU uses the inputs and outputs in the same way, the DEA model will “learn” the weights automatically and objectively as it computes the results.49 For these reasons, DEA is appealing to transit managers who are looking for a more comprehensive and objective method to measure performance. Indeed, Barnum et al. identified no less than 32 articles published on the use of DEA in transit performance evaluation.50

One of the first applications of DEA to transit performance was completed by Chu et al. in 1990. In this research, they utilized DEA to come up with a single indicator by which transit agencies could be compared to one another.51 They also use DEA to illustrate the empirical relationship between efficiency and effectiveness and discovered that these variables are not correlated.52 Karlaftis also looks at the relationship between efficiency and effectiveness using DEA in his 2004 report where he looked at data from 256 transit systems over five years. In contrast, he found that efficiency and effectiveness are positively correlated. He also looked at the issue of scale efficiencies and attempted to determine the

50 Ibid.
52 Ibid., 233.
optimal size of a transit agency (due to the principles of returns to scale). On this point he concluded that very large agencies exhibit decreasing returns to scale (meaning as large agencies get larger, they get less efficient).\(^{53}\) DEA research by Odeck and Alkadi expanded on this hypothesis and found that the average transit agency operates under increasing returns to scale.\(^{54}\) This would seem to contradict the results obtained by Karlaftis, though they were looking at different size agencies. The authors note that additional research is needed in order to account for geographical factors, such as operational terrain, weather, etc.

Lao and Liu also recognize the need to add a geographic element to the research and combine DEA analysis with Geographic Information Systems to add the spatial element to transit performance. Their research sought to explain the spatial effectiveness of various transit services at the Monterey-Salinas Transit agency in central California.\(^{55}\)

James Nolan argued that an outside observer could analyze transit agency technical efficiency by using information submitted to the U.S. Department of Transportation. He used this data to compare the efficiencies of 29 average-sized transit agencies, after accounting for variables such as average fleet age.\(^{56}\)

Taxpayers in recent years have often pushed for privatization of transit service, believing it to be more efficient than publicly provided transit. Pina and Torres sought to answer the question of whether the private sector is more efficient at providing transit service. They used data from many European transit services, some private and some public, with DEA analysis to measure their efficiency. While this research focuses on American transit agencies, the Pina and Torres study of European transit service does offer findings not available in the U.S. They concluded that the private sector was not any more efficient at providing transit service, all else being equal.\(^{57}\)

Tsamboulas investigated the same issue again in 2006 and found somewhat different results. His research showed that privately operated services are usually more efficient (meaning they are good at maximizing outputs and minimizing inputs—think “a lot of service at minimal cost”) but that publicly operated services are usually more effective (meaning they transport more passengers). Interestingly, he also revisited the efficiency-effectiveness relationship issue, and he found that the two are generally positively correlated (which agrees with Karlaftis’ findings).\(^{58}\)

Some interesting research has sought to use DEA to explain other factors that influence transit efficiency. Research by Boame in 2003 found that higher average transit speeds contribute to higher efficiencies. She also confirms Odeck and Alkadi’s research in


finding that most transit agencies operate under increasing returns to scale.\textsuperscript{59} Khattak and Yim study the effects of urban sprawl on transit. They find, not surprisingly, that sprawl makes transit less efficient. They use their findings to explore the possibility of transitioning to a personalized demand-response transit service in the San Francisco Bay Area. (This type of system would consist of many automated vehicles on some sort of nodal network; each traveler would be provided his/her own vehicle, which travels directly to the desired destination.)\textsuperscript{60}

3. Conclusion

Transit performance measures are an important part of any plan to objectively evaluate the performance of transit service. There is general agreement that performance measures are valuable and worthwhile, though there is no consensus on what performance measures to use or how to use them. The literature reviewed demonstrates a pattern of somewhat conflicting findings and results.

The field of Data Envelopment Analysis seems promising, however. The objective and thorough nature of this econometric method has broad appeal in transit performance research. More research is needed in this area in order to develop tentative hypotheses into conclusive findings. As more research is done, we may learn more solid truths about the industry and be better able to improve performance.

More research is also needed in benchmark setting. This crucial step is often left out of research, which is a serious omission. Even the most masterfully executed route performance ranking will lose its relevance if we are left wondering how to interpret the data.

Coupled with additional peer review and staff feedback, this research will provide sufficient direction and advice for forming performance measure recommendations for Santa Clara Valley Transportation Authority. This literature review will serve as the basis for the “Measuring Transit Performance” section of the research paper. Using the findings in this literature review, I will identify and evaluate transit performance standards used throughout the industry and at VTA. This in turn will provide a solid foundation for offering recommendations to improve VTA’s performance standards.


\textsuperscript{60} Asad J. Khattak and Youngbin Yim, “Traveler Response to Innovative Personalized Demand-Responsive Transit in the San Francisco Bay Area,” \textit{Journal of Urban Planning and Development} 130, no. 1 (2004): 42.
## 4. Annotated Bibliography

<table>
<thead>
<tr>
<th>Number</th>
<th>Source</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>Kuby, Michael, Anthony Barranda, and Christopher Upchurch. &quot;Factors influencing light-rail station boardings in the United States.&quot;</td>
<td>peer-reviewed journal article</td>
</tr>
</tbody>
</table>


### 5. List of Online Databases, Catalogs, and Keywords

<table>
<thead>
<tr>
<th>Database</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>JSTOR</td>
<td>transit, performance, DEA, measures, indicators, benchmark</td>
</tr>
<tr>
<td>Academic Search Premier</td>
<td>transit, performance, DEA, measures, bus, indicators, benchmark</td>
</tr>
</tbody>
</table>
Google Scholar

- transit, performance, DEA, data envelopment analysis, indicators, measures, Tsamboulas, Determinants of Productive Efficiency, A Test of Inter-Modal Performance Measures
- Li, Odeck, Alkadi, Traveler Response to Innovative Personalized
- Khattak, Determinants of Productive Efficiency
- Nolan, DEA Approach for Evaluating the Efficiency and Effectiveness of Urban Transit

ScienceDirect

- Khattak, DEA Approach for Evaluating the Efficiency and Effectiveness of Urban Transit

TCRPOnline

- transit, performance, bus