Bridging research and practice in operational performance measures for public transit systems in the US
Variations in performance evaluation methods and their appropriate use

Miwa MATSUO

Abstract

In the U.S., Moving Ahead for Progress in the 21st Century act (MAP-21) demands outcome-driven performance-based decision-making for transportation systems. However, practitioners are still struggling to find appropriate measures for performance that are appropriate and feasible. This paper briefly reviews performance measures developed in literature, and map relationships between measurement objectives and measurement methods. The objectives of measurement can be categorized into internal and external communications, and each category is further divided by the scope of measures. Based on the objectives and organization's analytic capacity, three different types of measures should be used appropriately. Simple performance indicators are the simplest and most useful measure in most occasions. However, other types of indicators should be jointly employed due to performance indicator's weaknesses in comprehensive project assessment and in inter-organizational comparison. Specifically, project evaluation through cost benefit analysis should be utilized to evaluate large-scale operational changes or capital investments. Aggregated performance measures would also help monitoring performance. They enables assessing managerial inefficiencies through inter-organizational comparison and the effects of exogenous factors to performance. For successful strategic management with performance measures, monitoring and analytic requirements should be determined based on the capacity of organizations. Supervising agencies such as state department of transportation should work as an information organization that provides feedbacks from the analysis of the data.

1. Introduction

Performance measures are important in planning and managing public transit systems because "we never do anything much about a problem until we learn to measure it." (Moynihan, 1978, p12) Since public transit services in the U.S. depend on financial supports from government, they are responsible for accounting appropriateness of expenditure. In 2012, the responsibility is formally stylized as the Moving Ahead for Progress in the 21st Century Act (MAP-21). MAP-21 requires an outcome-driven, performance-based approach to decision-making processes for transportation projects in the states and metropolitan areas. Currently, the MAP-21 requirements to public transit systems are limited to asset management and safety issues. However, comprehensive performance measures should be developed for transit systems, as is required for highway systems. Performance measurement is beneficial for transit operators themselves in identifying their operation problems, exploring potential for further system development, and improving customer experiences that promotes ridership.

Variations in productivity has also been an interest in economics and business studies. Although economics textbooks say that all the firms optimize their production process all the times, in reality, firms may not adapt to changing environments successfully. Thus, there is a need to develop evaluation methods of productivity, which allows various degrees of success by firm to reflect the empirical findings in business field. From the discussions by Debreu (1951) and Farrell (1957), economists have developed econometric approaches to account for productivity, while business and operations research researchers take nonparametric approaches to measure productivity. In both fields, literature volume of these productivity studies are already enormous.

Despite the development in performance measures in literature, practitioners still rely almost solely on simple performance indicators to measure their productivity. This paper aims to bridge research about performance measures and transit planning and operating practice. By examining both philosophy behind the performance measures and practical needs, this paper proposes appropriate performance measures by scope of measurement and by the capacity of organization.

This paper is organized as follows. In section 2, I review conceptual developments of performance measures. Section 3 briefly summarizes three major performance measurement methods. Section 4 discusses practical needs for measurements with assessment of best practices. Last in section 5, I propose a framework for monitoring and data assessment.

2. Conceptual developments of performance evaluation

2.1 Fundamentals of performance measures in transportation planning

Transit performance measures are typically designed to capture two important dimensions of transit systems, managerial efficiency (or production efficiency) and service effectiveness, or a combination of these two, overall system performance (Fielding, 1987, Chapter 4). As Fielding (1987) describes, the idea of managerial efficiency comes from operations-research tradition, while the measurements of service effectiveness comes from policy-analysis. In transit performance evaluation, efficiency is measured by comparing the volume of service provided with the volume of resource inputs. Namely, efficiency measure assesses whether the operator is making the best use of resources. Effectiveness measures consumption of transit services to evaluate social impacts of the services. In other words, effectiveness measure assesses whether the provided service is achieving its social goal.

Through the assessments of efficiency and effectiveness, supervising organizations of regional transit systems can specify the factors that undermine the performance. Within the sources of inefficiency and ineffectiveness, the effects of internal and external factors should be distinguished because the implications for improvements are quite different. The external factors should be considered carefully in the goal setting of transportation planning, while the internal factors should be utilized in the assessment of day-to-day operation and marketing strategies.

In the Fielding's classic approach, external influences such as demands, land use, and conditions of other transportation alternatives are considered in determining the volume of resource inputs (Fielding, 1987, p3). Managers of transit services consider these external factors as an exogenous factor and ignores endogeneity of consumption volume or surrounding environment (Figure 1). As a result, service consumption (i.e., transit ridership) is considered to be the final output determined solely by production process, rather than the interaction between demand and supply of the service.

Recent literature distinguishes internal and exter-

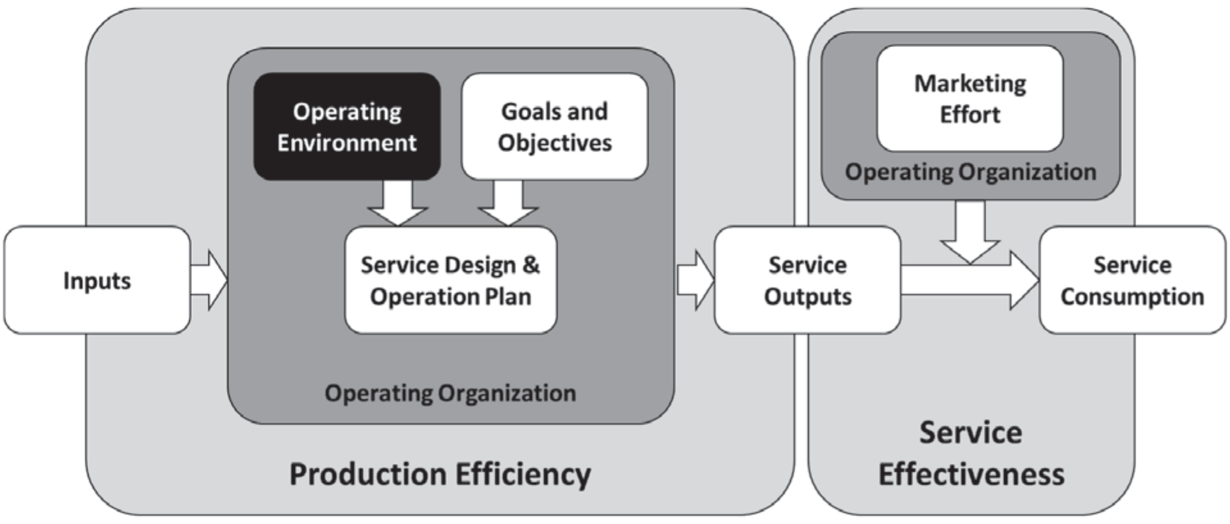

Figure 1.　Fielding's approach for measuring efficiency and effectiveness

nal factors more structurally. Economic approach rigidly separates out demand factor from supply performance analysis (Figure 2). Operations-research approach does not distinguish demand factor separately; however, it often avoids the issue of demand and supply relationship by narrowly focusing on service production process. Regardless the analytic approach, the effects of external factors are accounted by econometric analysis, and the remaining inefficiency and ineffectiveness are labeled as internal under-performance factors. Internal inefficiency (and ineffectiveness) are then decomposed into technical and allocative inefficiencies (and ineffectiveness) for further performance analysis. Technical inefficiency corresponds to the inefficiency in the use of existing resource mix, and allocative inefficiency corresponds to the inefficiency caused by suboptimal input choices.

Last, performance evaluation is not only required for improving transit management but also for accounting. According to Fielding (1987), the U.S. General Accounting Office (1972) addresses efficiency and effectiveness requirements in the program audit guidelines. Specifically, the guidelines describes that auditors evaluate three aspects of transit performance: (1) financial performance and compliance with legislative intent, (2) economy and efficiency, and (3) program results or effectiveness. In other words, as recipients of public funding, U.S. transit systems are required to address whether they are performing well both in service production process and in attaining original objectives (ridership). If they are underperforming, they need to assess the reasons and build plans for improvement.

2.2 Measurements to address efficiency and effectiveness

Fielding (1987) gives general guidelines for measuring service input, service output, and service consumption, and literature seems to follow the guidelines. As measurements for service inputs, Fielding recommends to employ “the quantity of resources expended to produce transit service, expressed in either monetary or nonmonetary terms.” Specifically, the service input measures should include operating cost and employee hours spent for operations, maintenance, and administration, capital investment such as number of vehicles or peak vehicle requirement, and energy utilization measured either by fuel cost or volume. As measurements for service outputs, he states that the measurement should be “the quantity of service produced by a transit operator, expressed in nonmonetary terms.” The measurement examples include vehicle hours, vehicle miles, capacity miles, and service reliability, and service safety measured by number of accidents. As measurements for service consumption, he discusses that the measurement should be “the amount of service used by the public may be expressed in either monetary or nonmonetary terms.” The examples of consumption measures shown in his book are the number of passengers (total, revenue, and special groups), passenger miles, and operating revenue (total and passenger).

Although Fielding’s guidelines are widely accepted both in research and practice, neither collect-

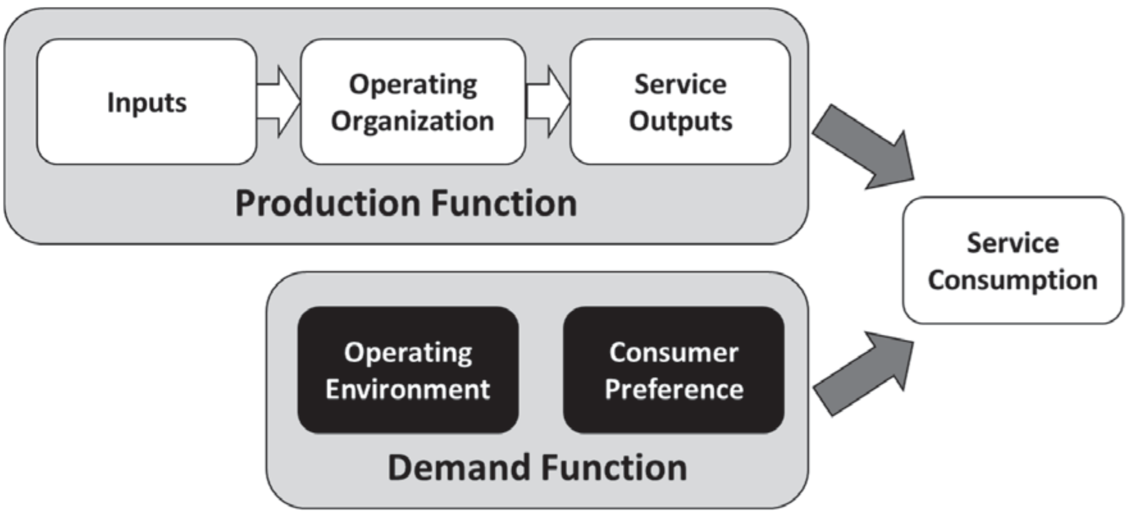

Figure 2. Economic approach to efficiency analysis

ing data nor constructing indicator from data is easy. Collecting desirable data is not easy because (1) there are multiple perspectives in transit services; (2) the inputs and outputs are shared by different types of passengers, and thus serve for different goals simultaneously; and (3) often the case, the most desirable measurements are technically or financially impracticable. For example, goals of bus service often include both reducing private vehicle traffic on congested corridor and social welfare for transportation disadvantaged people. Although they are two different goals that demand two different measures for evaluation, it is difficult to conduct two separate analysis because a single bus route may operate both peak and off-peak hours, run through both congested and uncongested corridors, and may serve for both regular commuters and transportation disadvantaged people. Moreover, frequently used measures such as revenue vehicle hours, revenue vehicle miles, unlinked passenger trips, and passenger-miles are inappropriate for evaluation because they do not account for when, where, and who consumed the service. “Big data” from smart cards may partly address the issue; however, not every agency has financial capacity to introduce such a technology. Even if the system is successfully introduced, it remains difficult how to summarize and assess such a disaggregated data to address defined objectives of the service.

Constructing efficiency and effectiveness indicators is another hurdle. Although a set of multiple simple indicators are preferable for understanding operation status, an aggregated single indicator is needed in interpreting whether their overall performance is good or not. In section 3, I discuss three different approaches in constructing performance measures.

2.3　The scope of performance measures

Before discussing the methods of constructing performance measures, I need to clarify the scope of performance measures because practitioners often mix arguments about performance measures for process (or operation) and outcome. The former is often called as productivity measures or narrowly defined performance measures. The latter refers to either solely output measures or life-cycle project evaluation. Technically, operational performance is measured by performance indicators (PIs), stochastic frontier analysis (SFA), and data envelopment analysis (DEA), while project evaluation is conducted through cost benefit analysis (CBA) or cost benefit ratio (CBR).

Performance measures and project evaluations are similar in that they evaluate how much output we can expect from inputs. However, they differ in the timeframe and the scope of analysis. Performance measures assess relative operational performance in a short-term regular business. Specifically, measures should be calculated for daily, monthly, or yearly productivity in a fixed regular operation by comparing the target agency with other agencies with similar businesses. The assessment suggests where managerial inefficiencies are taking place, and how much more productive its business can be.

Project evaluation evaluates the life-cycle performance compared to the baseline case, which is business-as-usual without project. Project evaluation differs from performance evaluation in three aspects. First, the evaluation score is calculated solely based on the additional costs and benefits associated with the target project. Comparison across the project options may take place after the project evaluation scores are calculated, but the score itself is unaffected by the existence of project alternatives. Second important difference from the performance evaluation is that the variations in productivity among operators are ignored in the project evaluation process. The analysis assumes that day-to-day resource allocation is optimized by operators. In other words, the analysis implicitly assumes certain-level of managerial productivity without question. Third, project evaluation is mainly for the analysis of long-term project outcomes, which typically includes large-scale capital investments. Day-to-day or year-to-year operational performance of regular business are included as costs and benefits of the project but not the main focus of the analysis.

3. Variations of performance measures

In this section, I explore three major methods of constructing operational performance measures: performance indicators (PIs), economic approach represented by stochastic frontier analysis (SFA), and operations-research approach represented by data envelopment analysis (DEA)

PIs are the most commonly used by practitioners (Karlaftis, 2012), and developed extensively in literature from the 1970s (examples including Fielding et al. 1978; Gilbert and Dajani, 1975; Meyer and Gómez-Ibáñez, 1981). For every important aspect of transit performance, one "marker variable" is selected with consideration of availability and reliability (Fielding, 1987). Those marker variables are usually simple and intuitive. For example, Fielding recommends to employ labor efficiency, vehicle efficiency, fuel efficiency, maintenance efficiency, and overall cost efficiency as efficiency indicators. Likewise, he recommends utilization of service, revenue generation, operating safety, public assistance (equity and distribution of funds), and social effectiveness (utilization by targeted beneficiaries) as service effectiveness indicators.

A major problem of PIs is that each PI may yield inconsistent result and does not provide any information about overall performance (Benjamin and Obeng, 1990). The inconsistency issue demands a smaller set of indicators, or preferably, a single indicator to describe transit performance. As a result, research focus has moved towards methods that formally construct production frontier and measure relative (in) efficiency compared to the frontier.

There are two major approaches in constructing efficiency frontier: an econometric (parametric) approach and an operations-research (non-parametric) approach. The standard of the econometric approach is set by Aigner et al. (1977), Schmidt and Lovell (1979), Schmidt and Sickles (1984). The econometric approach often takes probabilistic approach, and the method is represented by SFA. SFA estimates production function, cost function, or profit function that defines a production frontier, and (in)efficiency is measured as the difference between actual and ideal production. Greene (2008, p93) describes the approach as "a formal analysis of residuals from production or cost model." Specifically, outputs of organization i at time t (y_{it}) is described as a function of inputs (x_{it}), unobservable heterogeneity and random components ($v_{it} \sim iidN(0,\sigma_v^2)$), and inefficiency ($u_{it} \sim iidN^+(0,\sigma_u^2)$)

$$y_{it} = \mathrm{F}(x_{it}) * v_{it} * u_{it} \qquad \cdots(1)$$

Production frontier is estimated from econometric analysis, with some uncertainty because of v_{it}. If we consider a production process that requires two inputs, capital (K) and labor (L), the predicted frontier of one-unit of output would be an isoquant line with uncertainty (a dotted line in Figure 3). (In)efficiency score of the production unit 0 (P_0) is measured by the deviation from the predicted production frontier. Namely, comparing P_0 with P_0^*, the efficient production status on the predicted frontier (Figure 3). Cobb-Douglas or trans-log function is typically employed for analysis to make the equation 1 log-linear, with stronger preference to a more flexible trans-log function (e.g., Benjamin and Obeng, 1990; Karlaftis and McCarthy, 2002; Karlaftis and Tsamboulas, 2012). The model may include organization fixed-effects or random-effects to account for time-invariant unobservable heterogeneity (detailed discussions in Greene, 2004).

The effects of exogenous factors are assessed either by one-step or two-step approach (Greene, 2008). In the one-step approach, the observable heterogeneity of interest are included as parameterized functions of exogenous variables. In the two-step approach, efficiency measures are first constructed without considering those heterogeneity. Then, the

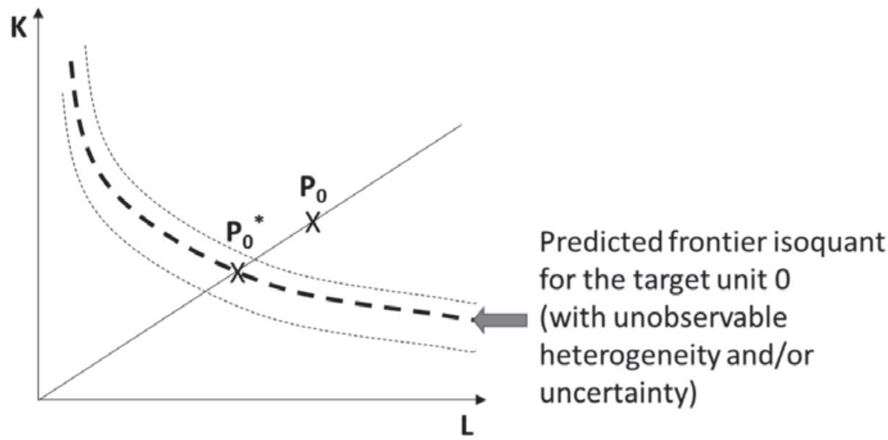

Figure 3. Efficiency estimation by SFA

estimated efficiency scores are regressed with the exogenous factors to assess the effects of exogenous factors to efficiency variations.

Econometric approach is consistent with general economic theory and formally treats random factors. Moreover, recent development in SFA approach allows means and variance of (in)efficiency vary across the firms, treats panel data more appropriately, and deals with endogenous factors using generalized method of moments (extensive reviews in Greene (2008) and van Biesebroeck (2007)). However, the econometric approach is often criticized about assuming independent and identical distribution for v_{it} and u_{it} having a known normal distribution (Charnes et al. 1996). In real data, particularly in transit data, the assumption is often invalid. Schmidt (1985) emphasizes the need to find an estimator whose consistency does not depend on specific distributional assumptions.

Operations-research approach to construct efficiency frontier is represented by DEA. Unlike SFA, DEA is characterized by its deterministic non-parametric approach. DEA is originally developed by Farrell (1957), and further stylized by Charnes et al. (1978) for constant return to scale analysis and Banker et al. (1984) for variable returns to scale analysis. Production frontier is constructed by "enveloping" what is achieved by existing units, using linear functions or hyperplanes (Karlaftis and Tsamboulas, 2012). For example, assume that there are four production units in observation (P_0, P_1, P_2, P_3) that produces one-unit of same output using different mix of two inputs, K and L (Figure 4). The production frontier is constructed by enveloping minimum-input observations, namely, connecting P_1, P_2, and P_3 with straight lines. Production (in)efficiency of unit 0 (P_0) is measured as the deviation from efficient mix of inputs (P_0^*), which is on the constructed frontier.

The effects of exogenous factors are assessed either one- or two-steps in DEA, too. Tobit regression is often employed for the two-step semiparametric analysis because efficiency score is censored. However, Simar and Wilson (2007, 2011) find that many preceding studies that employed simple Tobit analysis are inappropriate. Simple Tobit regression does not yield consistent estimates, and more complex double bootstrap procedures are required for estimating accurate coefficients and standard errors (detailed discussions in Simar and Wilson (2007, 2011)).

Supporters of operations-research approach claim that DEA is more appropriate than SFA in that it is free from assumptions about functional form and error term distributions. They also claim the DEA's advantages in handling multiple outputs (e.g., Karlaftis and Tsamboulas, 2012). Furthermore, recent development in network DEA (NDEA) makes it possible to handle multiple production processes simultaneously, which advanced its application from the original single-process black-box approach (with regard to NDEA methods, Kao (2014) provides an extensive review). Van Biesebroeck (2007) also discusses that DEA is preferred when productivity is highly heterogeneous across firms and scale economies is not constant.

The critiques of DEA, however, point out two serious concerns about DEA: inability to accommodate random variation in the data (Greene, 2004) and not paying enough attention to deal with endogeneity of resources (Orme and Smith, 1996).

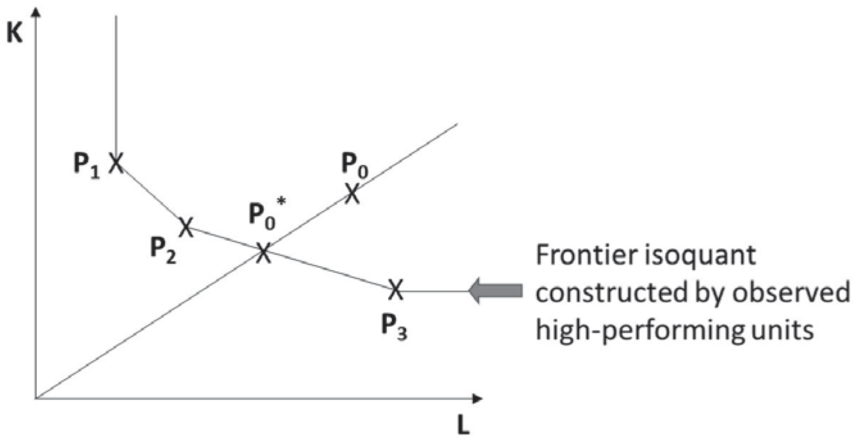

Figure 4.　Efficiency estimation by DEA

4. Practical needs for performance measures

4.1 Performance measures for internal and external communications

Despite the extensive methodological developments in literature, studies pay little attention about how to utilize their methods in practice. In transit planning practice, the objectives and use of operational performance measures can be categorized by whether the information is for internal monitoring or for external communication. Internal monitoring includes short-term operations level, mid-term planning level, and long-term strategic level management (Pangilinan, 2015). The external communication includes communication with direct users, government, and general public.

Operations-level performance measures are for making real-time operational adjustments. For example, New York City Metropolitan Transit Authority (MTA) has introduced performance measure that monitors service status every seconds to hours to maintain high level of service or to respond incidents and emergencies (Pangilinan, 2015.) More specifically, MTA monitors and visualizes subway train arrival times and predicted arrival times to stations. With this information, MTA can maintain more regular headways through making real-time operational adjustments such as skip-stop operation or holding trains.

Planning-level performance measures are used for monthly or yearly operational revisions, and they may lead to recommendations for future capital investments. As planning-level performance measures, MTA monitors wait-time distribution and on-time performance at terminals to check the adherence to operation standards (Pangilinan, 2015.) MTA also assesses the relationships between performance and train load, to understand when and where the troubles are concentrated. Regional Transportation District of Denver (RTD) is another interesting case of utilizing planning-level performance measures. RTD monitors asset condition closely and examined its relationship with service delay statistics. Through the assessment of seasonal patterns of technical break down and delays, they find that aged vehicle stocks are the main source of delays. With further analysis, they show that their capital stock replacement would be cost-effective and beneficial to society as a whole (Cripps, 2015).

Strategic-level performance measures monitor years of operational performance to assess the progress towards long-term goals. As Fielding (1987) emphasizes, measuring the degree of attainment is critical in strategic planning and management processes. The monitored progress will give feedbacks to goal settings and further planning of follow-up projects. In practice, San Francisco Municipal Transportation Agency (SFMTA) clearly sets its strategic goal as increasing accessibility. Based on this long-term strategic goal, SFMTA monitors accessibility changes between before and after the projects for evaluation (Pangilinan, 2015.)

Performance measures needed for internal monitoring is basically PIs, particularly for operations-level. Operations-level performance measures are mostly used for problem identification, which demands simple and intuitive measures for understanding the operational status. For this purpose, inter-organizational comparison is less likely to be useful. Hence, having multiple indicators is not a serious concern, and constructing a single indicator does more harm than good because aggregating indicators often masks the important details of operation.

Planning- and strategic-level performance monitoring also demands simple and intuitive measures, i.e., PIs, but they may do better jobs when considered in combination with aggregated indicators (by SFA or DEA) and/or project evaluation scores (CBA or CBR). In a mid- to long-range planning, operational improvement projects and capital investments under consideration become larger and more complex. As a result, (1) comprehensive project evaluation is essential in choosing an appropriate alternative, and (2) there is more room to learn from other operators about efficient operation and effective marketing. The former officially demands CBA or CBR, and the latter suggests that benchmarking its operational performance through SFA or DEA might be helpful. In the case of RTD Denver, simple PIs are basically used to identify the source of problem, and project evaluation is used to assess costs and benefits of capital upgrading.

To the extent of my knowledge, no organization successfully utilized SFA or DEA scores to assess their relative managerial performance. Even so, inter-organizational comparison with consideration of observable exogenous factors may help developing improvement plans. First, benchmarking their performance relative to other comparable agencies makes it possible to judge whether they need more managerial effort or investment to enhance the service. The analysis suggests the potentials of technical and allocative (in)efficiencies, which, in turn, gives operators better idea what data they should investigate to specify the source of problem. Second, the assessment of the exogenous factor effects makes the inter-organizational performance evaluation fairer. Transit agencies are afraid of being criticized about the things they cannot do anything about, such as low land use density, mountainous geography, and cold and snowy weather. Inter-organizational comparison of performance should explicitly consider such potential exogenous factors to make the lessons from evaluation realistic and useful.

Performance measures are also useful for external communication, including communications to direct users, government, and general public. Communicating service performance to users is important in enhancing customer experience in a short-run and in building trust in a long-run. Real-time operational data are particularly useful in improving customer experience. Massachusetts Bay Transit Authority (MBTA) in Boston developed a smartphone application to communicate real-time operational data (Tribone, 2015.) MBTA also shares some daily statistics on their website, such as passenger wait and travel times, headways, and running-time performance by route segment. Sharing operational information lets customers plan their schedule more smartly, and reduces wasted time on platform. In a long-run, sharing operational information builds better relationship between users and their service. People trust services more if they are accurately and immediately informed whenever troubles happen.

External communication with government is an important role of performance measures. As discussed earlier in this article, transit organizations are responsible for accounting their performance because they receive public funding. As funding distributors and supervisors, state department of transportations (DOTs) are expected to monitor whether the operators spend public money appropriately and whether the service is desirable to the community. Although minor operational issues may be assessed and resolved within an operating organization, state DOTs can help individual organizations improving service further through conducting inter-organizational performance analysis.

Last, communication with general public is also an essential piece of external communication. Federal Highway Administration (FHWA) explains that “public participation is an integral part of the transportation process which helps to ensure that decisions are made in consideration of and to benefit public needs and preferences.” (FHWA, 2015a) It also publishes a detailed official guideline for public participation to clarify the importance of getting feedbacks from every segment of local communities (FHWA, 2015b.) Although FHWA stresses the aspect of receiving information from general public, how to share information about current operations and future investment options is the keys to success in public involvement. Information shared at public meeting often includes technical and complicated matters, thus the data should be appropriately processed to indicators that are easy to understand for non-experts.

With regard to external communication aspects, PIs are easy to understand and appropriate for communication with non-professionals, particularly with general public and legislators. One critical point that promotes public understandings is making the indicators customer-oriented. Non-experts do not understand or care about operational details. Rather, customer experience measures make it clear how the transit is performing and how much improvement they can anticipate. Another key point is that the number of indicators presented to customers should be selective. Every important dimension of service performance should be presented in separate measures so that the interpretation is kept simple. However, if the number of indicators is multiple dozens, they just confuse people. The total number of indicators should be kept minimum for better communication.

4.2 Observed issues about implementing performance measures in transit planning

Although there are strong needs for measuring performance, operators often resist or unable to cooperate with the requirements for performance measures. First and foremost, the operators fear about revealing their poor performance and being punished(1). They suspect that Federal and state governments are trying to exploit the collected data to reduce budget without any technical help. The distrust is a critical issue in collecting accurate and reliable data for performance measure. It is important to promise to use data for helping them to develop better managerial plans, rather than punishing them through simply tightening budget without technical support.

Second, the operators often lack capacity in collecting and interpreting data. Measuring reporting burden is heavy for operators, particularly if they are very small. Even if technology helps operators collecting data, most small operators are incapable of analyzing and utilizing data. Large and high-human capital organizations such as MBTA, MTA, RTD, and SFMTA may be able to assess data in-house, but they are truly exceptional cases. In other words, operators are less likely to benefit from performance measurement unless supervising organizations help them interpreting data. For this perspective, state DOTs seem to be in the position that can transform it into information and give feedbacks to operators. With the expectation of such technical supports from supervising organization, operators are more likely to pay effort in data collection because it now benefits themselves.

5. Discussions and proposal for future monitoring system

As section three shows, PIs are useful for various operational monitoring, which naturally explains their popularity in transit planning practice. Compared to aggregated performance indicators constructed by SFA or DEA, PIs summarize operational and service information without reducing informational dimensions too much. Loss of details is particularly problematic in using for internal minor operational adjustments.

Another reason for PIs' popularity is that it demands less capacity in constructing and interpreting results. Researchers have developed SFA and DEA extensively to improve econometric accuracy and to incorporate complex production structures. However, the methods are now a little too complex for non-researchers. Knowledge and skills in econometrics are essential in conducting carefully designed accurate analysis in SFA or DEA, and not many transit operators or supervising DOTs have such expertise. For PIs, the only difficult stage is the initial decision about what to monitor. After the operators determine the list of monitoring PIs, they just need to collect data and conduct simple calculation.

Despite its popularity and usefulness of PIs, monitoring only PIs does not tell whole story about the performance. First, PIs are unable to tell overall impact of multiple effects of an operational change or capital investment. In the assessment of net benefit of the project, the operator needs to conduct CBA, i.e., conduct project evaluation rather than performance measures. Second, PIs are inappropriate for inter-organizational comparison. Unlike SFA or DEA, PIs do not account for the effects from exogenous factors or scale economy of the systems. As a result, a transit system that may look apparently better in PIs thanks to the transit-friendly environment or scale economies. By using SFA or DEA, assessment can evaluate multiple effects of exogenous factors simultaneously, and score "unobservable managerial efficiencies" of each operator. Needless to say, scoring overall efficiency alone does not help problem identification in operation. If poor performance is observed, it is important to investigate the reason through detailed examination of raw data and simple PIs. Such an inter-organizational analysis is complicated and requires intensive human capital. Since most operators and small metropolitan planning organizations (MPOs) are unlikely to hold such capacity, state DOTs should play roles in the analysis. In reality, it may be difficult to exercise such a regular monitoring of operational efficiency; however, it would be highly effective in raising performance of transit services.

In sum, disaggregated and aggregated performance indicators should be comprehensively monitored and utilized as follows. Operators and small MPOs should monitor a set of simple performance indicators and adjust their operational plans based on the data shown by the PIs. The list of indicators should be defined under the guidance of Federal DOT, so that the operators collect consistent data. Most of the raw data requirement should follow the National Transit Database data reporting requirements to keep the reporting requirement manageable. State DOTs and large MPOs should go one step further towards strategic management. In addition to monitoring and assessment of simple PIs, they should conduct more advanced analysis including inter-organizational comparison. Furthermore, such large organizations should accumulate know-how of analysis and share the infor-

Table 1. Correlations among efficiency and effectiveness scores measured by DEA methods and SFA methods shown by Karlaftis and Tsamboulas (2012)

			Efficiency					Effectiveness		
			DEA		TL			DEA		TL
			E-DEA-C	E-DEA-V	E-TL-F	E-TL-R	E-TL-TV	C-DEA-C	C-DEA-V	C-TL-TV
Efficiency	DEA	Constant Returns to Scale (E-DEA-C)	1							
		Variable Returns to Scale (E-DEA-V)	0.434	1						
	Translog Prod. Func. (TL)	Fixed Effects (E-TL-F)	0.176	0.193	1					
		Random Effects (E-TL-R)	0.521	0.679	0.030	1				
		Time Varying Efficiency (C-TL-TV)	0.317	0.784	0.068	0.916	1			
Effectiveness	DEA	Constant Returns to Scale (C-DEA-C)						1		
		Variable Returns to Scale (C-DEA-V)						0.955	1	
	TL	Time Varying Efficiency (C-TL-TV)						0.772	0.879	1

Reproduced from Tables 6 and 7 of Karlaftis and Tsambroulas (2012)

mation through research platform such as TRB.

Finally, more research should follow van Biesebroeck (2007) in the integration and distinction of appropriate performance measurement methods. Research communities of SFA and DEA are largely different in the fundamental philosophy, and communicate much less than they should. However, they have different strengths and weaknesses, and should complement each other in the applications. Kerstens (1996) and Karlaftis and Tsamboulas (2012) find that efficiency scores depend on the choice of measurement methods. Kerstens (1996) examines how the variation in disposability in inputs and outputs in DEA method affects the efficiency scores, and he finds that the correlation can be as low as 0.4. Karlaftis and Tsamboulas (2012) examine the correlation among efficiency and effectiveness scores measured by variations of DEA and SFA methods. They show that the correlation coefficient can be as low as 0.2, suggesting that the scores differ substantially by the choice of measurement methods (Table 1). The sensitivity to measurement methods demands a theory that explains the criteria for choosing an appropriate methods for measuring efficiency. In the future studies, researchers should develop the discussion even further to address what methods to be employed for inter-operator performance comparison of transit systems.

NOTE

(1) Interview response heard at the 5th international conference on transportation systems performance measurement and data (June, 2015)

Reference

Aigner, D., Lovewell, K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6, 21-37.

Banker, R., Charnes, A., Cooper, W., 1984. Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. Management Science 30 (9), 1078-1092.

Benjamin, J., Obeng, K. 1990. The effects of policy and background variables on total factor productivity for public transit. Transportation Research Part B 24 (1), 1-14.

Charnes, A., Cooper, W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research 2 (6), 429-444.

Charnes, A., Gallegos, A., Hongyu, L., 1996. Robustly efficient parametric frontiers via multiplicative DEA for domestic and international operations of the Latin American airline industry. European Journal of Operational Research 88 (3), 525-536.

Cripps, L., 2015, June. Moving data to move people: Relationship between asset condition and customer experience. Presentation at the meeting of the 5th international conference on transportation systems performance measurement and data, Denver, CO.

Debreu, G., 1951. The coefficient of resource utilization. Econometrica 19. 273-292.

Farrell, M., 1957 The measurement of productive efficiency. Journal of the Royal Statistical Society. Series A General. 120, 253-290.

Federal Highway Administration, 2015a., Public Involvement/ Public Participation. Retrieved from http://www.fhwa.dot.

gov/planning/public_involvement/
Federal Highway Administration, 2015b., Public involvement techniques for transportation decisionmaking: 2015 update. Retrieved from http://www.fhwa.dot.gov/planning/public_involvement/publications/pi_techniques/fhwahep15044.pdf

Fielding, G., Glauthier, C., Lave, C.A., 1978. Performance indicators for transit management. Transportation 7, 365-379.

Fielding, G., 1987. Managing public transit strategically. A comprehensive approach to strengthening service and monitoring performance. Jossey-Bass Publishers, San Francisco, California.

Gilbert, G., Dajani, J., 1975. Measuring the Performance of Transit Service. University of North Carolina Press, Chapel Hill, North Carolina.

Graham, D.J., 2008. Productivity and efficiency in urban railways: parametric and non-parametric estimates. Transportation Research Part E 44, 84-89.

Greene, W., 2004. Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the world health organization's panel data on national health care systems. Health Economics 13, 959-980.

Greene, W., 2005. Fixed and Random Effects in Stochastic Frontier Models. Journal of Productivity Analysis 23 (1), 7-32.

Greene, W., 2008. The econometric approach to efficiency analysis. In: Fried, H., Lovell, K., Schmidt, S. (Eds.), The Measurement of Efficiency. Oxford University Press.

Kao, C., 2014. Network data envelopment analysis: A review. European Journal of Operational Research 239, 1-14.

Karlaftis, M., 2004. A DEA Approach for evaluating the efficiency and effectiveness of urban transit systems. European Journal of Operational Research 152,354-364.

Karlaftis, M., McCarthy, P., 2002. Cost structures of public transit systems: a panel data analysis. Transportation Research Part E: Logistics and Transportation Review 38(1), 1-18.

Karlaftis, M., 2008. Privatisation, regulation and competition: a thirty-year retrospective on transit efficiency. Privatisation and Regulation of urban transit systems. OECD Publishing, Paris, France.

Karlaftis, M., Tsamboulas, D., 2012. Efficiency measurement in public transport: Are findings specification sensitive? Transportation Research Part A, 46, 392-402.

Kerstens, K., 1996. Technical efficiency measurement and explanation of French urban transit companies. Transportation Research Part A 30 (6), 431-452.

Moynihan, D., 1978. The politics and economics of regional growth. Public Interest 51, 3-21.

Orme, C., Smith, P., 1996. The potential for endogeneity bias in data envelopment analysis. The Journal of the Operational Research Society 47 (1) 73-83.

Pangilinan, C., 2015, June. From operations to strategy: Making better decisions through performance measurement. Presentation at the meeting of the 5th international conference on transportation systems performance measurement and data, Denver, CO.

Schmidt, P., 1985. Frontier production functions. Econometric Reviews 4 (2) 289-328.

Schmidt, P., Lovell, C., 1979. Estimating technical and allocative inefficiency relative to stochastic production and cost frontiers. Journal of Econometrics 9, 343-366.

Schmidt, P., Sickles, R., 1984. Production frontiers and panel data. Journal of Business and Economic Statistics 2 (4), 367-374.

Simar, L., Wilson, P., 2007. Estimation and inference in two-stage, semi-parametric models of production processes. Journal of econometrics 136 (1), 31-64.

Simar, L., Wilson, P., 2011. Two-stage DEA: caveat emptor. Journal of Productivity Analysis 36 (2), 205-218.

Tribone, D., 2015, June. It's not me, it's you: Refocusing transit agency data to be about customers and for customers. Presentation at the meeting of the 5th international conference on transportation systems performance measurement and data, Denver, CO.

Van Biesebroeck, J. 2007. Robustness of productivity estimates. The Journal of Industrial Economics 55 (3), 529-569.