Chapter 3

Literature Reviews

Theories of efficiency and reviews of some empirical studies concerning 1) efficiency measurements of bus service under Data Envelopment Analysis approach (DEA), and 2) DEA to measure the performance, efficiency and effectiveness of public transit will be discussed in this chapter.

3.1 An Overview of Performance, Productivity, and Efficiency

Performance of production units is normally described in term of high or low efficiency and productivity.

For the efficiency of a production unit, it means a comparison between observed and optimal values of its output and input. The comparison can take the form of the ratio of observed to maximum potential output obtainable from the given input, or the ratio of minimum potential to observed input required to produce the given output, or some combination of these two. In these two concepts of comparison, the optimum is defined as production possibilities, while efficiency is only technical. In this event, efficiency is economic and is measured by comparing observed and optimum cost, revenue, profit or whether the production unit is assumed to pursue, subject, certainly, to the appropriate constraints on quantities and prices.

Measurements of efficiency and productivity are important owing to two reasons. Firstly, they are successful indicators of performance measurement, by which production units are evaluated. Secondly, by only these ways of measurement, they can also separate their measured effects on each input and output factors. The results can explore hypotheses concerning the sources of efficiency or productivity differentials. Data sources are essential for the policy maker in order to help the institution design policies to improve their performance.

The ability to quantify efficiency and productivity provides recommendation with a control mechanism to monitor the performance of production units under its control.

The principal objective of performance measurement is a guide for improvement, where is appropriate by a comprehensive set of evidence-based goals. Besides, in its popular manifestation, performance measurement leads to league tables but in the context of performance management, it is only the starting point of an exercise in performance measurement. Thus, further details analysis and possibly inspection of the best and the worst performers is necessary for understanding the production process and deriving useful information which may assist both of the worst and the best performers to make further improvements in efficiency.

"Efficiency" is widely used in many aspects, but there are three main definitions employed as follows (Dechpolmat, 2003).

1) Input, cost or allocate efficiency, these efficiency measurements intend to the productivity that uses less inputs to produce more outputs.

2) Process efficiency is the capability of an operating system that should be smooth, fast, and correct.

3) Output efficiency has the purpose is to maximize profit. The process will finish in time and the products have the quality in order to satisfy customers.

The measurement of efficiency can be divided into two types as follows. Firstly, absolute efficiency is the complete process that transforms all inputs into outputs. However, in real situations there are some mistakes which might be caused from the operation, hence; perfect operation is very difficult to happen in reality. Secondly, relative efficiency occurs from the previous reason. Hence, the possible ways to measure the efficiency are to compare the efficiency from different aspects and operations: for example, firstly, comparing between revenue and cost (whether the revenue is less or higher than the expenses), secondly, comparing the operation between organizations which provide the same or similar product, thirdly, comparing the speed of work, fourthly, comparing the quality of products, and lastly, comparing customers' satisfaction.

Production ratio is used from numerous studies to measure the performance or compare the efficiency between the organizations. This kind of measurement shows the proportion of productivity (output/input) or sometimes known as partial productivity. The high score indicates good operation. It can compare between years, or between firms that produce the same product in the same

year. This can apply with either private firm which produces the product such as factories that provide service or non-profit organization such as library. Besides, there are several ways of measurement that could be adopted in different studies depending on product, information availability and the hypothesis of the structure of production and manufacturers' behavior. However, the production ratio is not the good way because it does not has ability to consider about the quality of product while concern about only the quantity.

Coelli et al. (1997) stated the definition of the productivity of a firm as the ratio of the output(s) that it produces to the input(s) that it uses.

Productivity = outputs/inputs

The formula above is simple calculation when the production process involves single input and single output. However, if there is more than one input (which is often the case), then method for aggregating these inputs into a single index of inputs will be used to obtain the ratio measuring productivity.

Referring to total factor productivity, productivity measurement involves all factors of production. Other traditional measures of productivity, such as labor productivity in a factory, fuel productivity in power stations, and land productivity (yield) in farming, are known as partial measures of productivity. These partial productivity measures can provide a misleading indication of overall productivity when considered in isolation because a partial productivity measure is that it relates output to a single factor of production and, therefore, fails to consider tradeoffs among input factors.

The terms productivity and efficiency have been used frequently over the last ten years by various commentators on such as article, report and so on, and they are often used interchangeably, but it is unfortunate that they are not precisely the same things.

The production theories such as the behavior of the producer, production frontier, the economies of scale, and return of production size need to be applied in the study in efficiency in order to find out that which firms are on the line of

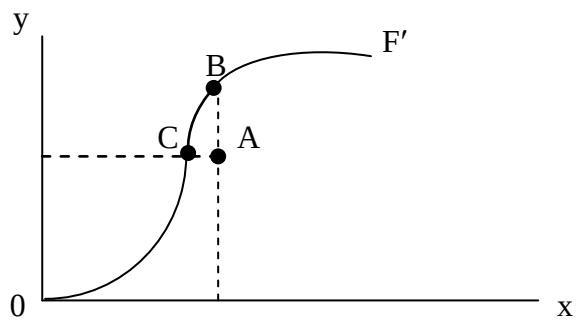
production frontier which means that those firms have more efficiency than others firms.

Figure 3.1 represents line 0F′ is a production frontier which may define the relationships between the input and the output. As the production frontier represents the maximum output attainable from each input level, hence it reflects the current state of technology in the industry. Therefore, firm in that industry operates either on that frontier, if they are technically efficient, or beneath the frontier, if they are not technically efficient.

According to Figure 3.1, Point A represents an inefficient point whereas point B and C represent efficiency points. A firm operating at point A is inefficient because technically an output should be increased to the level that associated with point B without requiring more input.[1]

Figure 3.1

Production Frontiers and Technical Efficiency



Source: Coelli et al. (1997).

And if information on prices is available and behavioral consumption, such as cost minimization or profit maximization, is appropriate, the performance can be considered allocative efficiency.

Economic efficiency consists of technical efficiency and allocative efficiency.

[1] Or alternatively, it could produce the same level of output using less input (i.e., produce at point C on the frontier).

Firstly, technical efficiency reflects the capability of the organization in obtaining the most products with given the level of inputs. Secondly, allocative efficiency reflects the capability of firm to use proper input factors under price constraint for producing a given output level at minimum cost.

3.2 The Developments of Data Envelopment Analysis (DEA)

The efficiency measurement has been an interesting subject of many organizations in their productivity improving. Fifty years ago Farrell (1957) stated in his classic paper about the measurement of productive efficiency that "the problem of measuring the productive efficiency of an industry is an important to both the economic theorist and the economic policy maker. If the theoretical arguments as to the relative efficiency of different economic systems are to be subjected to empirical testing, it is essential to be able to make some actual measurements of efficiency. Equally, if economic planning is to concern itself with particular industries, it is important to know how far a given industry can be expected to increase its output by simply increasing its efficiency, without absorbing further resources (p.11)."

Farrell (1957) also stated that the primary reason that all attempts to solve this problem had failed due to a failure to combine the measurement of the multiple inputs into any satisfactory measure of efficiency. In addition, these inadequate approaches included forming an average productivity for a single input (ignoring all other inputs), and constructing an index of efficiency in which a weighted average of inputs is compared with output. Responding to these inadequacies of separate indices of partial productivities, Farrell proposed an activity analysis approach that could be more adequately dealt with the problem. His measures were intended to be applicable to any productive organization. Unfortunately, he confined his numerical examples and discussion to single output situations, although he was able to formulate a multiple input case.

Twenty years after Farrell's seminal work, Charnes et al. (1978) responded to the need of satisfactory procedures to assess the relative efficiencies of multi-input and multi-output production units, introduced a powerful methodology which has subsequently been titled data envelopment analysis (DEA). The original

idea behind DEA was to provide a methodology whereby, within a set of comparable decision making units (DMUs), those exhibiting best practice could be identified, and would form an efficient frontier. Furthermore, the methodology enables one to measure the level of efficiency of non-frontier units, and to identify benchmarks against which such inefficient units can be compared.

Charnes et al. (1978) proposed deriving appropriate multipliers for a given DMU by solving a particular non-linear programming problem to measure the technical efficiency. They constructed the constant returns to scale (CRS) to measure the relative efficiency. They proposed a model that had an input orientation and assumed constant returns to scale (CRS). The CRS assumption is appropriate when all firms are operating at an optimal scale. However, imperfect competition, government regulations and constraints on finance are the reasons that firm can not operate at optimal scale. Furthermore, Banker et al. (1984) extended the assumptions of CRS by providing variable returns to scale (VRS). This model can measure for all markets, whether the market has the regulations from government or financial regulations.

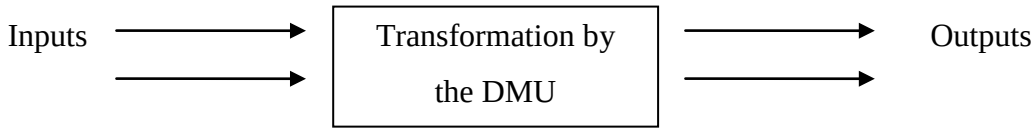
Decision Making Units (DMUs)

Decision Making Unit (DMU), coined by Charnes et al. (1978), in their seminal paper on DEA those DMUs should be homogeneous entities in the sense that they use the same resources to procure the same outcomes albeit in varying amounts.

The characterization of the unit of assessment as “decision making” implies that it has control over the process it employs to convert its resources into outcomes. In DEA the resources are typically referred to as “inputs” and the outcomes as “outputs” and these are the terms we shall adopt. A DMU transforms inputs into outputs in a process depicted in Figure 3.2.

Figure 3.2

A DMU Transforms Inputs into Outputs



Source: Ray. (2004).

The identification of inputs and outputs in DMUs' assessment is difficult due to the fact that the inputs should capture all resources which impact the outputs and that the outputs should reflect all useful outcomes on which are needed to assess the DMUs.

Furthermore, any environmental factors which impact the transformation of resources into outcomes should also be reflected in the inputs or the outputs depending on the direction of that impact.

Measurement of efficiency are based on estimations of the degree to which the DMU concerned could have secured more output for its input levels, or the degree to which it could have used less input for its output level, thus, it need to investigate whether the outset of the DMUs in question have more discretion over input or over output levels (the answer depends on context).

It will turn to be another different efficiency measurement if different inputs or outputs controllable are used. Therefore, before the measurement is introduced, the Pareto efficiency will be defined. Two definitions are given as the output orientation is appropriate when outputs are controllable and the input orientation is appropriate when inputs are controllable.

In this chapter, a set of homogeneous DMUs uses one or more inputs to secure one or more outputs.

Accordingly, for output orientation, a DMU is Pareto-efficient if it is not possible to raise anyone's output levels without reducing at least another's output levels and/or without increasing at least one of its input levels.

For input orientation, a DMU is Pareto-efficient if it is not possible to reduce anyone's input levels without increasing at least another's input levels and/or without lowering at least one of its output levels.

Two measures of efficiency, relate respectively to the output and input orientation above are most normally used as follows.

Technical output efficiency of a DMU is the maximum proportion of its observed output levels that when all outputs are expanded radially as far as feasible, without detriment to its input levels.

The technical input efficiency of the DMU is the maximum proportion of its contracted input levels is of the observed level of that input. All inputs of a DMU are contracted radially as far as feasible, without detriment to its output levels.

Figure 3.3 illustrates the difference between input and output oriented measuring of efficiency. It depicts the case where DMUs produce a single output by using a single input. It could be seen that the line OD is locus of maximum output levels attainable for giving input levels. OD is the efficient boundary of the production possibility set which located between the input axis and OD. Next, DMU A in relation of the definition of Pareto-efficiency will be considered and the measures of the efficiency will be introduced as follow.

According to Figure 3.3, DMU A could have been operated at point D in order to attain maximum output for its inputs level. Alternatively, DMU A could have been operated at point C in order to use minimum input for its output level. Hence, DMU A is not Pareto-efficient. It could deliver more output for no additional input or use less input without detriment to its output level.

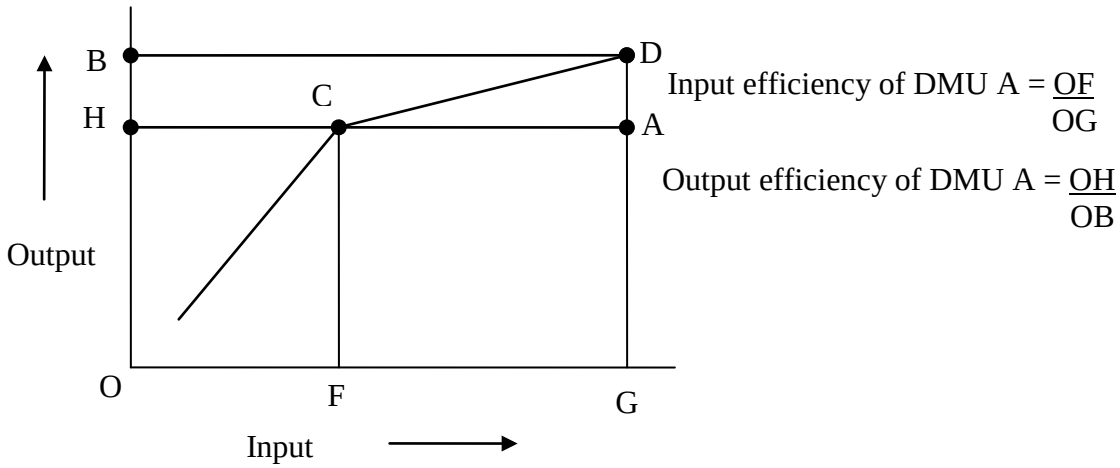
The technical input efficiency of DMU A as defined earlier is $\frac{0F}{0G}$

Similarly, the technical output efficiency of DMU A is $\frac{0H}{0B}$

OF is the minimum input level at which DMU A could deliver its output, while OG is its observed input level.

Figure 3.3

Measures of Input and Output Efficiency



Source: Ray. (2004).

3.3 Literature Reviews of Efficiency Measurement of Public Transit and the Studies using DEA Approach

Bus service is a kind of public goods even it is not a goods that have been exactly characterized as the definition of public good.

Public goods must have two properties. Firstly, the consumption from one person does not detract or prevent others. And secondly, public goods are non-excludability. When a good has been created, it is impossible to prevent other people from gaining access to it.

Publicly provided goods which there is a large marginal cost associated with supplying additional individuals are referred to as publicly provided private good, though of the cost of running, a market provide one of the rationales goods for the public which is not the only or even the most important rationale.

According to Chapter 2, bus service cannot purely be provided by private operators without government subsidy as the private operators must receive the financial assistance from the government for their investment. According to the studies on the effects of subsidy on the performance, efficiency and effectiveness of

public transit and bus service by Fabbri (1996); Karaftis & MaCarthy (1997); Lawrence & Kornfield (1998); Tisato (1998); and Goeverden et al. (2006), although some studies concluded that subsidy will crowd out the efficiency of the performance of public transit; there are some works which argued that subsidy is necessary for public transit and can be considered in terms of redistribution because it might be communally useful to people and also sustain the production of public transit (Fabbri, 1996).

Public ownership of transit systems and the growth of government subsidies eliminated profitability as the standard of public transit performance. Lacking a market oriented guideline to performance improvement, focus shifted to various performance measures that many felt would capture the multi-dimensional nature of transit operations (Kenneth, 2006). Therefore, there are literatures argument on whether the public transits should receive subsidies from government and what the effects of subsidy on their performance are.

Some studies (Fabbri, 1996; Karaftis & MaCarthy, 1997; Tisato, 1998; Lawrence & Kornfield, 1998; Nelson et al., 2006) consider about the effect of subsidy to examine their potential application to performance based subsidy allocation procedures. In addition, many studies adopted the performance indicators to analyze the effects of subsidies on performance and productivity of the transit systems.

Nelson et al. (2006) studied the current benefits and optimal level of provision of transit in Washington, D.C. Since the discrepancy between transits' large share of local transportation resources and its generally low share of local trips raised questions about the use of scarce transportation funds for this purpose. They used a regional transport model consistent with utility theory and calibrated for the Washington, D.C., metropolitan area to estimate the travel benefits of the local transit system to transit users and the congestion-reduction benefits to motorists. From this study they found that (i) rail transit generates congestion-reduction benefits that exceed rail subsidies; (ii) the combined benefits of rail and bus transit easily exceed local transit subsidies; (iii) the lowest-income group receives a disproportionately low share of the transit benefits, both in absolute terms and as a share of total income; and (iv) for practical purposes, the scale of the current transit system is at optimal.

Fabbri (1996) suggested the reasons for subsidizing public transit by following Standard Welfarist arguments which are: economies of scale, second-best problems, redistribution in kind, option value and imperfect information. For all these reasons, it might be socially useful to sustain the production of public transit. Sen et al. (2007) showed the case of Delhi the result in a considerable modal shift that increasing bus transport demand if the price of public transport were to be subsidized and private transport were to be priced optimally.

Some countries in Europe provide free bus service for some groups of their populations such as students and retired population. Witte et al. (2006) examined the effects of free bus transport in Hasselt for the students of Flemish colleges and universities in Brussels comparing with another group (students from French speaking universities and colleges) in 1997 by conducting a survey among the students to examine their present travel behavior (number of trips, motives, …, modal choice) and also the changes of the travel behavior comparing with the previous year. Besides, they compare the current travel behavior between the students who receive free bus service, and people who do not have right for free bus service. According to the study, they found out that there is an incredibly increasing in public transport using among the group of students who benefit from free service.

Goeverden et al. (2006) explored two extremes effects on the axis of cost recovery which are free public transport, and public transport without subsidies. They survey free public transport in four existing cases which could be divided in two cases in Belgium and two cases in Netherlands which is the free transport for students in the city of Hasselt, the Brussels region (for students), the Leiden-The Hague bus corridor. The results of this study showed that the suspension of subsidies for public transport supply operation has strong negative on service quality of public transport while the fares increase substantially. Consequently, demand of bus service tends to decrease.

Nevertheless, some literatures argued that bus service is unreliable on subsidy. Tisato (1998) tried to explore the nature of this relationship between service unreliability and subsidy. The degree of unreliability of bus services and the level of optimal bus subsidy influence cost of users. The over-riding conclusion that can be drawn from the analysis of the impact on subsidy of changes in service unreliability is

unreliability have significant influence on subsidy analysis and also on the level of optimal subsidy. The impact is rather small in cases that a single type of user behavior occurs in either random or planned.

Furthermore, the noticeable point that cannot be forgotten about efficiency and equity is that the subsidy crowd out the efficiency. However, in the public view, equity issue cannot be avoided. In efficiency and equity scheme, Lawrence & Kornfield (1998), mentioned that point between efficiency and equity are more important for the transportation that have to answer the question that the transportation is public good or not, if the consideration will be more on equity. On the other hand if not, the consideration will convert to efficiency. Moreover, the efficiency and effectiveness can also be measured from the performance of public transit, according to Karlaftis & McCarthy (1997) study that measured the performance of public transit systems in Indiana in 1983 – 1994 by using factor analysis methods to generate a set of underlying attributes (factors) which capture the performance of public transit systems.

There are many studies that use the Data Envelopment Analysis (DEA) to estimate the efficiency cost and performance of public sectors. For example, Odecki, & Alkadi (2001) focused on the performance of Norwegian bus companies which has been subsidized by the government. They found out that the efficiency is evaluated by a deterministic non-parametric DEA approach. In addition, Wang et al. (n.d.) used DEA to compare bus service performance; six individual on-time performance measures in three pairs. Besides, Karlaftis (2004) also used DEA and globally efficient frontier production functions to measure efficiency and effectiveness, and tested the relationship between performance and scale economies by using data from 256 US transit systems over a five-year period. The last example that uses DEA to estimate efficiency cost and performance of public sectors is the study of Soteriou & Zenios (1998) that measured the product cost estimates of bank products at the branch level base on the non-parametric benchmarking technique of DEA.

In contrast a major criticism of DEA models is that they are deterministic for input or output data and do not allow for uncertainty therefore some researchers do not use DEA without applying other theory or approach to estimate the

results. For instance, Tzeng et al. (2001) apply the complex theory with DEA model to evaluate the Production efficiency for Taipei city Bus Company. It could be noticed that the DEA model that was used in this paper is too complex because normally DEA model can only be used in condition which the input and output are crisp numbers. In addition, Barnum et al. (2006) measured the performance that compares the efficiencies of subunits within the transport organization by using DEA as they used Stochastic Frontier Analysis to adjust the DEA efficiency scores of environmental factors and presented several brief case studies of the points that were identified as inefficient by DEA.

3.4 Literature Review of Bangkok Mass Transit Authority

Since most people in Bangkok and metropolitans who use the service of BMTA and often face with impolite manners from bus drivers and conductors, thus, the study about the satisfaction of people after BMTA applying ISO 9002 standard in the organization took place in order to improve the service.

Tuntiwasinchai (2002) studies the satisfaction of passengers and employees of BMTA in both administrative department and performer section of bus number 522. The survey showed that after adopting ISO 9002 to improve the service that three groups of samples, most of them agreed that customer satisfaction, quality management system, management responsibility, resource management, and improvement were enhanced especially in part of resource management, nevertheless, product realization was not much be improved.

Bundasak (1984) stated that the incentive wage could create low cost of production. In addition, incentive wage could also enhance workers' willingness to work and faithfulness to the organization. Hence, the author studied about the relationship between the willingness to work and the earned incomes of the bus drivers who earn incentive wages and those who receive only the salary without the incentive wages. This study found out that the bus drivers perceived influence wages as the reward beside from the normal wages, hence, incentive wage is the objective to induce the employees to work harder and feel satisfied with their job. This study also found out that employees who had the incentive wages have better performance than

another group. In addition, the drivers who got incentive have less absence record from their duty than the group that cannot earn incentive because if they do not work, they will be able to earn less money. However, according to the study the work satisfaction between the two groups are not different.

Kraisith (1995) studied the satisfaction and the expectation of salary and benefits of the BMTA regular bus drivers in 9 zones in Bangkok and he found those bus drivers' satisfactions in all kinds of salary and benefits are at the average low rate for example satisfaction of percentage payment, overtime rate, and daily allowance.

Charoenwannaying (1996) studied the level of factors that influenced the opinion of BMTA's bus drivers toward incentive wage payment with the sample size of 374 bus drivers by taking questionnaires and statistical tools. This study found out that BMTA's bus drivers had a normal attitude toward incentive wage payment.

Lapmark (1990) tried to determine the optimum number of buses and trips that incurred the least total cost. The study employed the case of Zone 3 Division 1 by using linear programming technique and sensitivity analysis. She found that the increase in fuel price and fixed cost affected the total cost of operation without any effect on optimal number of buses and trips. But the increasing number of buses by 200 and 10% increase in travelling time per trip affected the optimal number of buses and trips in during both peak and off-peak hours.

Bhasabutra (1994) studied only BMTA's maintenance system because of there is the limitation of time and budget in finding out the most appropriate governance structure to solve inefficiency problem of the organization. In this study, expenditures were divided into 2 parts which are (a) costs of repairing the buses by private bus companies and (b) costs of repairing the buses by BMTA itself. Besides, the transaction cost economics approach was adopted in this study due to the reason that this tool can be applied for any economic activities of any organization. Unfortunately, the result from applying transaction cost economics approach to analyze BMTA bus maintenance activity is failed to suggest about the most appropriate type of governance structure which the organization should adopt in order to solve their inefficiency problem.

In concordance, Mana Mahasuvirachai offered the ideas that the bus service could enhance their efficient and could alleviate the traffic congestion by

improving the bus system. They suggested that the bus reforming should be divided into three main routes which are; (1) main bus routes bus which run on main roads that have more than six lanes. (2) sub-main routes buses which connect roads and Sois with main road and (3) express buses where run only on express roads[2]. There was a survey the opinion of Bangkokian about the bus reform by The National Institute of Development Administration's (NIDA), Social Development Association & the Kirk College's Social Science Institute (1994) about bus system. The results of this survey showed that the majority of more than 10,000 people in Bangkok and suburbs suggested that the city's bus system should be changed to be more efficient in order to reduce traffic congestion problem in the city.

The study from Thammasat University (1997) which focused on the structure of BMTA's revenue and expenditure and the way to subsidize public service obligation (PSO) in three ways. This report found that the main reasons that BMTA has loss since its inception is the lack of working capital for management and investment. This report also suggested the regulation to solve this problem in both short-term and long-term which are to increase the revenue and also to decrease their expenditure, and to transfer to be under supervision of Bangkok Metropolitan Administration respectively.

The recent research conducted by Chulalongkorn University (2007)[3] which studied about the structure of revenues and expenditures of BMTA and analyzed the revenue and expenditure structures in details introduced the way to correct BMTA's loss problem by adopting several solutions, for example, to reduce the expenditure, by changing the fuel from diesel to NGV, and to apply e-ticket to improve the quality of service.

According to the above literature reviews, there are no studies regarding technical efficiency of BMTA. Therefore, this study may encourage the government to realize the significance and develop the efficiency in operations of both BMTA and state-owned enterprises.

---

[2] Manager Magazine. (1996, January).

[3] Chulalongkorn University was hired by BMTA to do this research.

In conclusion, all of the studies that were mentioned previously are useful guidelines for solving BMTA's loss problem and also improving efficiency plan in the future. The studies suggest not only the way to solve problems in term of quality such as ISO 9002 and other technologies, but also in term of service such as bus drivers and other stuffs. In addition, the rest of all studies also concern about the financial condition of BMTA that should be urgently resolved, especially, the inefficiency in operation and management scheme of BMTA as almost of the studies pointed out that the expenditure of BMTA was over the revenue. To be concluded, all mentioned problems indicate that the administration of BMTA is inefficient. Accordingly BMTA should find proper strategies or receive assistance from the government to resolve the problems.