Title Page (with Author Details)

Application of an incentive for bus drivers to achieve an improvement in the quality of service

Alejandra Chaparro[a*], Patricia Galilea[b], Juan Carlos Muñoz[c] and Joaquín Poblete[d]

[a] *[Av. Vicuña Mackenna 4860, Macul, Santiago, Chile)] [+56 956671223] [aechaparro©uc.cl]*
[b] *[Av. Vicuña Mackenna 4860, Macul, Santiago, Chile)] [+56 223544270] [pgalilea@ing.puc.cll]*
[c] *[Av. Vicuña Mackenna 4860, Macul, Santiago, Chile)] [+56 223544270] [jcm@ing.puc.cl]*
[d] *[Av. Vicuña Mackenna 4860, Macul, Santiago, Chile)] [+56 223544350] [joaquinpoblete@uc.cl]*

*Corresponding author

JEL classification: (minimum of 3, maximum 6)
D24, D86, R49

Keywords: (minimum 5, no maximum, separated with a semi-colon)
Multitasking; incentives; contracts schemes; bus drivers; public transport systems.

ABSTRACT

Concession contracts of operating companies of the public transport system of Santiago, consider important fines if companies fail to comply with the operating plan, regularity and other operational variables included in those contracts. On the other hand, drivers receive a fixed payment with no pecuniary incentive related with their performance.

The main objective of this paper is to analyze the application of a monetary incentive for bus drivers focused on increasing the number of passengers transported to test the existence of multitasking, specifically checking the behavior of drivers regarding bus speed.

We conducted a field experiment with an operator of Transantiago and we used a difference in differences analysis to show that with the pecuniary incentive tested, drivers raised their transported passengers in 9% when riding in long bus routes. We found some evidence of multitasking associated with a decrease in speed of 3%. Thus, our research provides suggestive evidence that inefficiencies may be occurring in the operation because of the lack of adequate incentives for drivers.

Application of an incentive for bus drivers to achieve an improvement in the quality of service

1. Introduction

With the arrival of Transantiago in 2007, a new contract structure between operators and the transport agency was implemented, consisting of concessions awarded through competitive tendering (Gómez-Lobo & Briones, 2013). Another relevant novelty of the new system was that the drivers of the operating companies started earning a fixed salary, leaving aside the old structure that corresponded to a payment per passenger transported (Gómez-Lobo & Briones, 2013; Puga, 2017; Tiznado et al., 2014). The effects of drivers' fixed salaries were twofold: on the one hand races between buses for collecting passengers in the streets were eradicated (which was the main security problem of the old transport system), but on the other hand fixed salaries also eliminated incentives for drivers to control fare evasion and attend the demand of passengers (Gómez-Lobo & Briones, 2013).

Drivers fulfill a significant task in the operational functioning of the bus system (De La Vega, 2018; Tiznado et al., 2014). They are the ones who make the decisions in situ of how to operate a bus, they have direct contact with the users, and they must know how to face the different unexpected situations of the transit of the city. The quality of service perceived by system users depends on their behavior (Gómez-Lobo & Briones, 2013). Therefore, by recognizing the importance of the driver's role in the operation of public transport buses, understanding that their behavior has a direct influence on the quality of service offered and considering that it is necessary for it to carry out various tasks during its working day, it is important to study how an incentive scheme applied to drivers may alter their behavior and thus obtain measures on its influence on bus operation.

The current concession contracts with the operating companies of Transantiago depend on five different indexes measuring quality of service. According to the fulfillment of each one, the payment and fines that will be applied to each operating company are defined. Given the above, the operating company must distribute its efforts in different aspects to ensure a minimum quality of service is established. While part of this task corresponds to the operational areas of the company, the driver has also an important role to comply with these indexes. Consequently, for the operator it is very important that the driver makes a list of tasks in addition to driving the bus. This turns the driver's workday into a job with a potential presence of multitasking elements (Holmstrom & Milgrom, 1991).

The main objective of this paper is to analyze the application of a monetary incentive for bus drivers focused on increasing the number of passengers transported to test the existence of multitasking, specifically checking the behavior of drivers regarding bus speed.

For this research, a field study will be carried out in a company operating the public transportation system, Transantiago. A monetary incentive will be applied to a group of drivers and we will proceed to measure various aspects of their behavior as a worker and as a bus operator. We used a difference in difference analysis to show that with the pecuniary incentive tested, drivers raised their transported passengers in almost 10% when driving long routes. We found some evidence of multitasking associated with a decrease in speed of 3%.

Globally, several public transport systems have implemented an operator payment scheme that depends on the number of passengers transported. For instance, Vigren & Pyddoke (2020) studied the impact on bus ridership of passenger incentive contracts in Swedish public transport. Sepúlveda & Galilea (2020) examined payment schemes related to the frequency of transport services and number of passengers transported, showing the relative importance of those incentives. Nevertheless, other payment schemes for bus operators have been studied: Hensher & Stanley (2003) and Hensher & Houghton (2004) propose a contract based on operational variables (performance based contract, PBC), taking into account reforms of public transport system done in different cities, while Rojo et al. (2015) apply that PBC framework to Spain, including new quality criteria. Van de Velde et al. (2008) analyzed competitive tenderings done in the Netherlands. Another example is the work done by Gagnepain and Ivaldi (2002), recommending fixed-price contracts for better-informed regulators. Other empirical studies on public transport contracts focus on the estimation of cost functions and/or efficiency measures (Gagnepain and Ivaldi, 2002; Batarce and Galilea, 2018; Dalen and Gómez-Lobo, 2003; Avelani et al., 2016, 2018) and the analysis of subsidies for Public Transport (Börjesson et al., 2019, Basso & Jara-Díaz, 2010; Basso & Silva, 2014, among others).

However, few have analyzed the implementation of driver incentives, an exception is Hartman, Kurtz & Moser (1994), who present cases such as the Golden Gate Transit, in San Francisco, The Massachusetts Bay Transportation Authority, Capital Area Transportation Authority in Lansing, Michigan and Metropolitan Transit Commission serving the Minneapolis-St. Paul area, all systems that achieved successful results from applying different incentive programs to their bus drivers. In our study, the implementation and analysis of the results of the application of a driver incentive will be carried out. In addition, econometric methodologies will be included to assess impact that have been very little used in the transport sector, which is something not reported in the existing literature reviewed.

This paper is organized as follows: we will briefly describe the microeconomic models related to contract theory and multitasking which inspire our experiment and the experimental design theory needed for our field experiment. Then, we will describe our experiment followed by the results obtained. Finally, we will conclude and draw some future lines of work.

2. Methodology

The methodology used in this research considers a microeconomic model that emerges from contract theory with the effect of multitasking. Then, we describe the field experiment, which uses an experimental design with a stratification treatment to control the selection bias.

2.1 Contract theory, multitasking and current context of drivers in Chile

The present work is based on contract theory. Commonly, in agent and principal contractual relationships, the aim is to fulfill two objectives: assign risk and reward for the work the agent does. The payment received by the agent is for developing a task. In the case of this research, the principal corresponds to the operating company and the agent to the driver. In 1991, Holmstrom and Milgrom introduced in contract theory the phenomenon of multitasking, presenting for the first time the possibility of a contractual relationship in which the agent performs more than one task.

The most common model of contractual relationship is based on the fact that an agent is required to perform a single task. What is desired by the principal could be that this task is carried out at the greatest possible volume. If the volume of the task is easy to measure, for example, expeditions made by the driver, then the agent will make as many trips as possible. However, this may mean a loss in the quality of the task. Therefore, it would be necessary for the incentive scheme to be able to distribute the agent's attention among different tasks. This type of multitasking models works when the task to be performed cannot be divided into two people. Note that this will be the definition of multitasking in contract theory, the presence of various tasks in the work of the agent that are important for the principal.

In the case of Transantiago, the operating company is regulated by a public agency according to five different indicators and according to the performance in these indicators, revenues, fines, and discounts are calculated. To obtain a good performance the company focuses on achieving a correct and efficient operation of buses, in large part this task depends on how the driver performs his work (Tiznado et al., 2014). However, the salary of drivers is fixed, thus, they are hired and paid to drive, but the company not only cares about driving. The operating company not only expects the driver to drive, it expects them to obtain a good performance on the indicators. For example, drivers should be aware of the separation from the bus that precedes it (regularity), buses must alight at the stops with passengers waiting to ride the bus, drivers should complete their cycle time as expected, bus expeditions should be done at the times indicated, among others. Therefore, the profile of a driver’s job has several elements of multitasking and the salary should not be fixed.

Since the payment received by operators depends on compliance with indicators related to the operation, they have monitoring systems that allow them to control their performance in an aggregate manner. For instance, they do monitor the frequency of a certain bus line. However, there is no individual monitoring of the driver regarding the performance on these indicators. Although drivers have two monetary bonuses in this company in which the experiment was applied, none of them is related to their driver performance: One depends on how punctual they are to start their workday and the second depends on whether or not they turn on the bus console (provides GPS pulses) when they start a trip. This, this research explores the possibility of monitoring individual performance related to service quality.

At the same time, Bolton y Dewatripont (2005) indicate that if the salary and the security of the worker is independent of the performance of the agent, then an incentive scheme is required. Given that currently drivers have a fixed salary regardless of their performance the first condition is fulfilled. Furthermore, the driver knows there is a low risk of losing his job because of his performance. For instance, if the driver does not stop at a stop this does not necessarily mean a dismissal. Moreover, according to data from 2016 there is a deficit of 7% in the number of drivers and it is estimated that by 2026 this will grow and reach a value from 16% to 20% (SECTRA, 2016). Finally, it is important to consider that even if a driver gets fired, he will most likely be hired in another of Transantiago's operating companies. As commented by one of the Transantiago operators, some basic rules of Chilean labor law are more flexible in the case of drivers because of the existing deficit of drivers (Vega, 2019).

2.2 Experimental design, stratification, and difference in differences

To develop the experimental design for this research, we followed the steps recommended by Duflo, Glennerster & Kremer (2007). First, for the research question: What is the effect of an incentive on an individual's behavior? We should try to answer the following: how would the people who participated in the incentive experiment have reacted in the absence of the experiment? The complexity of the response is evident. Comparing the results of the same person over time will not necessarily provide reliable information regarding the impact of the experiment, since other factors that affect the results may have changed. Thus, it is not possible to have an estimate of the impact of the experiment on an individual, but it is possible to obtain an average impact on a group of people when compared with a group of similar characteristics, but that were not part of the experiment.

To obtain the average impact described above, two groups of individuals must be considered: the first is the control group, which is composed of individuals who have not been exposed to treatment and who should, in theory, have similar results to those who, if they received the treatment, although in practice they do have differences. The second group consists of those who were exposed to the treatment. Then, any difference in outcomes between the groups can be attributed to both the impact of the program and the pre-existing differences, which is commonly called "selection bias". Without a reliable way to estimate the size of this, you cannot decompose the difference between the treatment effect and the bias term (Duflo et al., 2007).

We define Y_i^T as the average result of the variable that we wish to analyze from the group i if was exposed to treatment. Y_i^C as the average result of the variable of interest of the same group i if was not exposed to treatment. In addition, we define Y_i as the result of the variable of interest that is actually observed for the group i. We are interested in evaluating the difference $Y_i^T - Y_i^C$, which results in the effect of acknowledging the treatment in the group i. Since it is not possible to observe group i with and without the treatment at the same time, the individual effects of the members of the group in the presence of the treatment cannot be observed. However, it is possible to know the expected value of the average effect that the treatment has that is defined as $E\ [Y_i^T - Y_i^C]$.

If we had access to a large data source that had the results of the variable of interest for a large number of groups i, then you could take the average of the groups that are exposed to the treatment and the average of those that are not. In a large enough sample, this will converge to:

$$D = E\left[Y_i^T \middle| Exposed\ to\ treatment\right] - E\left[Y_i^C \middle| Not\ exposed\ to\ treatment\right]$$
$$D = E\left[Y_i^T \middle| T\right] - E\left[Y_i^C \middle| C\right] \quad (1)$$

Adding and subtracting $E\left[Y_i^C \middle| T\right]$, expression that is translated as the supposed result for a subject in the treatment group if it had not been treated (an amount that cannot be observed but is logically well defined) is obtained:

$$D = E\left[Y_i^T \middle| T\right] - E\left[Y_i^C \middle| T\right] - E\left[Y_i^C \middle| C\right] + E\left[Y_i^C \middle| T\right]$$
$$D = E\left[Y_i^T - Y_i^C \middle| T\right] + E\left[Y_i^C \middle| T\right] - E[Y_i^C|C] \quad (2)$$

The above gives two important terms, the first is the effect of the treatment that is trying to isolate, that is, the effect of the treatment in the treatment group. That would answer the question: on average, in the treated groups, what difference did the treatment make? The second term represents the selection bias that captures the difference of untreated potential outcomes between the treatment and the control groups. That is, without a present treatment what difference of results there would be between the two groups (treaty and control). The groups may have had results of the variable of interest different on average even if they had not been treated. Then, systematic differences may arise between the groups treated and those that were not treated (Duflo et al., 2007).

It is easy to notice that $E\left[Y_i^C \middle| T\right]$ cannot be observed, which produces that the magnitude and the sign of the selection bias are impossible to quantify. So, there is no way to identify what portion of the difference between the treatment and selection group results are explained by this bias (Duflo et al., 2007). The first challenge is to pose a situation in which it can be assumed that the selection bias does not exist or that there is a way to correct it.

An option to eliminate selection bias is when individuals or groups of people are randomly assigned to the treatment and control groups respectively (Duflo et al., 2007). However, it can be randomized by observing variables. For this it is necessary to define a set of observable variables, with them strata are formed given by the intersections of the variables of said set. Then, within each stratum, it is randomized which individuals will belong to the treatment or control group. This scheme is called stratification. With this you can get to consider that the treatment is as good as a random assignment. After applying the conditionality of the set of theoretically observable variables, the selection bias disappears. Stratification ensures the improvement of accuracy only if the variables considered for this explain the variation in the treatment of interest (Cox and Reid, 2000; Duflo et al., 2007).

One way to analyze data that helps reduce selection bias is the difference-in-differences methodology. The differences in difference estimates use the pre-period differences in the results between the treatment and control group to control the pre-existing differences between the groups, when there are data both before and after treatment. If $Y_1^T (Y_1^C)$ is the possible result "if treated" ("if untreated") in period 1, after treatment, and $Y_0^T (Y_0^C)$ the possible "if it is" ("if not treated") result in period 0 before the treatment occurs. The individuals belong to the group T or group C. Group T t is treated in period 1 and is not treated in period 0. Group C is never treated. The difference in differences estimator is $\widehat{DD} = \left[\hat{E}[Y_1^T|T] - \hat{E}\left[Y_0^C \middle| T\right]\right] - [\hat{E}\left[Y_1^C \middle| C\right] - \hat{E}\left[Y_0^C \middle| C\right]]$ and provides an unbiased estimate of the effect of treatment under the assumption that $\left[\hat{E}\left(Y_1^C \middle| T\right) - \hat{E}\left(Y_0^C \middle| T\right)\right] = \left[\hat{E}\left(Y_1^C \middle| C\right) - \hat{E}\left(Y_0^C \middle| C\right)\right]$ that is, that in the absence of treatment, the results in both groups would have followed parallel trends.

3. Experimental design

After the methodology has been defined, the manner in which the experiment was designed and how the incentive was built for its subsequent application will be described in this section.

3.1 Sample size and measurement periods

To carry out the experiment, we worked with a group of drivers from one of the Transantiago operating companies. It was determined to work with three of the company's terminals. The terminals correspond to real estate intended for the arrival and departure of buses that provide services for the Public Transportation System. The number of bus drivers available to participate in the experiment is 96 for terminal 1, 90 for terminal 2 and 39 for the terminal 3. Since no collaborator can be forced to participate in an experiment and considering that everyone should be given the option to participate in this experiment due to company's policies, we did a face-to-face invitation to all the drivers of the three terminals. This resulted in the participation of a total of 134 drivers, 45 from the first terminal, 60 from the second and 29 from the third. This achieves 46.9%, 66.7% and 67.4% of participation per terminal, respectively.

The way in which the participating drivers were defined may produce a selection bias. However, to mitigate this bias, the volunteer drivers were randomly divided into two different groups for each terminal (group 1 and group 2). Also, considering the difference-in-differences methodology, for the first period of the experiment, group 1 takes the role of group exposed to treatment and group 2 is considered a control group with which the results will be compared. Then, in the second period of the experiment, these groups change roles. Thus, we hope to obtain comparable data between groups, since both are made up of drivers who wanted to participate in the experiment, but were encouraged to participate at different times.

To build the groups we used a stratification method to reduce variance and selection bias. The observable variables chosen to apply in the stratification are those presented in Table 1, which also explains the possible values of each of the variables. Then, for the continuous variables, age and seniority in the company, blocks were defined to reduce the number of strata and simplify the stratification.

To understand the variables chosen, it must be taken into account that in this company the drivers are divided into two groups, those who work during the morning (shift AM), those who work during the afternoon (shift PM). This shift is the same throughout the month, what is changing is the time of entry. Also, a worker has one or two free days a week, changing week by week. Therefore, as he must work 45 hours per week, there are weeks where he works at least 7.5 hours a day and there are weeks where he works at least 9 hours a day. Those who have signed the overtime arrangement work at least 10 hours a day. However, all drivers have the option of doing overtime whenever they want and it is necessary (Vega, 2019).

Table 1. Variables used in stratification and their possible values

Variables	**Possible values**
Gender	Feminine o masculine.
Age	Year categories: 20-29, 30-39, 40-49, 50-59, 60-69, 70-79.
Seniority	Year categories: 0-2, 3-5, 6-8, 9-11, 12-14, 15-17.
Type of Shift	AM, PM
Nationality	Chilean or foreign.
Overtime arrangement	Signed or not signed.

Finally, the field work lasted four weeks. In the first two weeks, from August 5th to 18th, 2019, group 1 of each terminal was exposed to the treatment and therefore group 2 had the role of control group. After August 19th to September 1st, the following two weeks, group 2 was exposed to treatment and group 1 was the control group. The period in which no group is treated can be considered as any period prior to the weeks in which the incentive was applied.

3.2 Choice of the variable for the incentive

According to the literature in case of applying a payment scheme for drivers this should be aligned or imitate the contractual structure that the operating company has with the transport agency (Gómez-Lobo & Briones, 2013). Then, to build the incentive, a variable was chosen that depended to a large extent on drivers and that was also related to the variables that are considered to make the payment for the service provision of the operating company. These variables are stipulated in the concession contracts with the transport agency and there are two types. The first relate to payment and the second relate to fines. The first group are: number of transactions entitled to payment, commercial kilometers traveled and the index of compliance with transport capacity (ICT), related to the size of the buses and when they are required. In the second group we have: frequency compliance index (ICF) and regularity compliance index (ICR) (Gómez-Lobo & Briones, 2013).

The variables ICT, ICF and commercial kilometers depend to a large extent on how the company executes the operation. Although the driver can affect the number of buses that are dispatched by the hour (ICF) or the commercial kilometers by refusing to carry out an expedition, the company may put another driver to make the expedition and thus meet the requirements. On the other hand, the type and size of the bus that is being dispatched (ICT) depends on the operation and not on the driver. Then, the variables that depend more on the driver are: the number of transactions or Bip validations, which in this case corresponds to the amount of fares paid by the users who boarded the bus and the ICR.

Thus, the variable chosen to apply the incentive is the number of Bip validations, since the ICR measures the time elapsed between the pass of one bus and the next. Therefore, the driver can affect the ICR, but the value of the index also depends on the driver in front of and behind him on the route. Therefore, the number of Bip validations is a more direct measure of performance than if the ICR were measured. It should be noted that all variables have factors external to the driver that may affect the driver's performance.

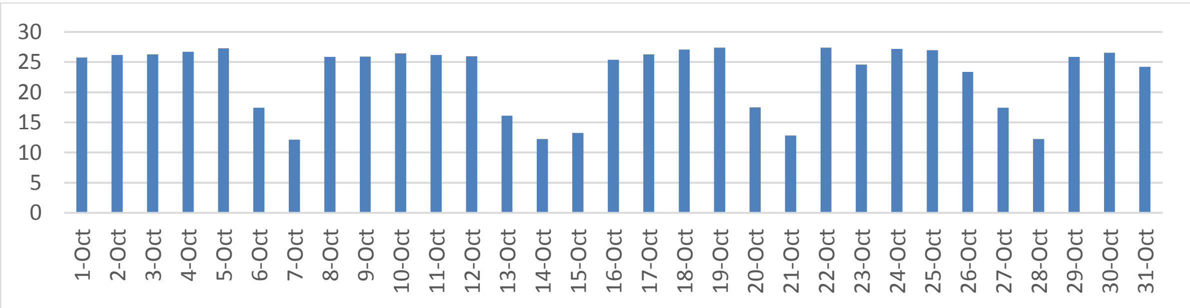
The driver may affect the number of transactions or Bip validations in the following ways. First, the driver can do an evasion check by requiring passengers to validate their fare when entering the bus. Second, he can control the speed in order to keep an eye on the road and wait a little longer at the bus stops in case more passengers arrive. Third and lastly, the driver can distance himself a lot from the driver in front of him and go very close to the driver who follows him, with the aim of capturing more transactions with a bigger interval between his bus and the one who follows him. In the latter case, the operator would not be obtaining more Bip transactions in general, but rather redistributing it among the drivers, which would not be aligned with the objective for which the incentive was applied.

3.3 Transaction analysis

We carried out an analysis of the bus driver that is comparable between them and allows us to identify his own behavior and not the operation's. Given this, we should understand the performance of Bip transactions and identify which aspects of the operation affect their volume. For this analysis, data from October 2018 were used and only the transactions of the 10 services contemplated in the three terminals selected were analyzed. The database contains information of all the bus expeditions made during that month and the respective Bip transactions obtained. An expedition is the trip made by the driver from the point of origin to the destination of a service (outbound), thus the return of this trip is a different expedition (inbound).

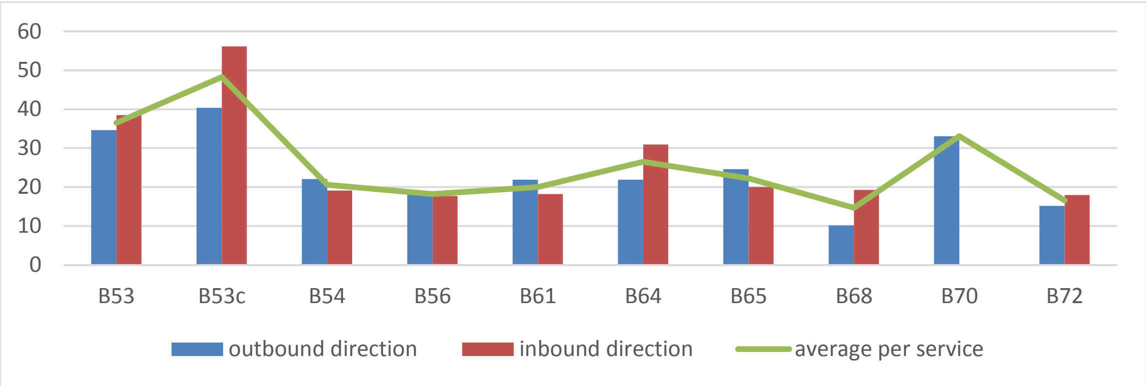
First, the general behavior of Bip transactions throughout the month must be analyzed. Figure 1 shows the average number of Bip transactions per expedition for each day of October considering the total number of transactions and expeditions for the 10 services studied. The first day of the month is a Monday, and we can distinguish business days and weekends since the latter have fewer transactions. It is possible to observe an expected behavior of decrease of transactions for the weekends and for the day of October 15th that was a holiday. However, for business days, differences are seen between weeks of the month and between days of the same week. This could be due to differences produced by the behavior of the bus drivers or by other aspects of the operation.

Figure 1: Number of average transactions per expedition for every day of October 2018.



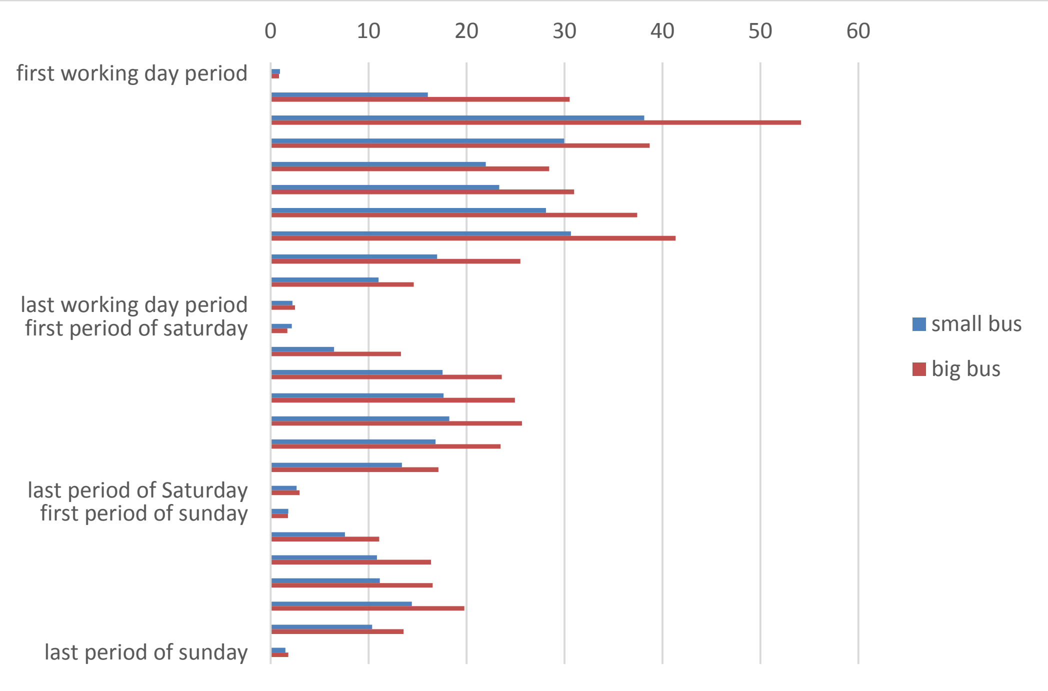
Secondly, it is necessary to know how these transactions are distributed according to each service, since different services are conducted by different groups of drivers and because the drivers during their working day make expeditions for more than one service from their respective terminals. Figure 2 shows the average of Bip transactions per expedition for the month of October for each service according to the inbound or outbound journey. Each service has different levels of performance per expedition, this may be due to the fact that the demand for passengers served by each one is different or may be due to some driver behavior. In the same way it can be observed that there are differences between an outbound or return expedition, even for some services this difference is greater than for others. This give us some evidence that service and direction (inbound or outbound) are important factors in the behavior of the volume of Bip transactions per expedition. Therefore, we should control expeditions by these factors as well.

Figure 2: Average transactions per expedition for October 2018 according to service and according to the round trip direction for each service.



Finally, it should be considered that the drivers work at different times of the day and the operation changes according to the different periods (peak or non-peak periods). In addition, the company works with different bus sizes. Small buses have a capacity of 49, 53 or 60 passengers and large buses have a capacity for 99, 100 102 and 105 passengers depending on the model. Figure 3 shows the ratio between the number of average Bip transactions per expedition for the month of October organized by periods and by bus size. In the periods of the day in which the demand for passengers is high, the size of the bus becomes relevant. This is shown for the morning peak hour, period 04, in which the average performance in the amount of Bip transactions of expeditions made on large buses exceeds by at least 7 transactions compared to an expedition in the same period performed on a smaller bus. So, both the size of the bus and the period of the day should be factors to consider in the construction of the incentive.

Figure 3: Average transactions per expeditions by bus size for each period for the month of October 2018.



3.4 Incentive scheme

Once the behavior of the transactions per issue has been studied, it is necessary to determine how the incentive will work. First, a minimum level of transactions was defined with which it must be met for each type of expedition. To determine this level of compliance, expeditions were ordered in descending order according to the number of transactions for each service-period-direction-bus size, then the threshold of transactions was assigned for the 20% highest number of transactions. This value corresponds to the minimum level of transactions that each expedition should have with the same characteristics. For example, if in a section there are 100 expeditions, when ordering these from highest to lowest in number of transactions, the expedition found in place 20 with, for example, 35 transactions, then the level of compliance will be defined as reaching at least 35 transactions for each expedition carried out in this service, in that period, in that direction and with that bus size.

However, for non-peak hours the level of compliance when the expedition is done in a large bus is less than in a small bus, therefore, when this occurs, the bus size is not considered and the same procedure is done. The same happens when the difference between the compliance level of an expedition on a large or small bus is less than 4. In addition, an approximation was made in such a way that the compliance level was rounded up to a multiple of 5. The latter is decided to facilitate the application of the incentive given that it will be easier for the driver to remember a closed value.

This incentive scheme seeks to equate the number of transactions that drivers get in each expedition. The theory is that the differences in the number of validations per expedition depend to a large extent on the driver's effort to attract passengers. Therefore, we must encourage to increase this amount and thus ensure that all expeditions reach the same average level.

Once the level of compliance is defined, it is necessary to determine the benefit obtained by the driver for complying with the incentive. This benefit corresponds to monetary payment for each extra passenger over the level of compliance. This scheme has been proposed to maximize change in behavior (Poblete & Spulber, 2012). For example, if the level of compliance for service 1 in the morning peak period for the outbound direction on a large bus is 26 and the driver achieves 36, then the equivalent value will be given to the payment for 10 extra transactions. This payment corresponds to 100 CLP [1]for each extra transaction, that is, 1,000 CLP. Thus, all the expeditions that the driver makes during the days of the experiment will be compared and a total amount of extra transactions will be obtained that will be translated into a payment at the end of the month, in bonus format. This payment should encourage drivers to seek as many passengers as possible.

With this incentive, drivers can receive a benefit proportional to their effort. Each driver has the possibility of earning a portion of money that does not depend on the performance of their peers and that depends on an aggregate performance over time, such as rewarding 30% of the drivers with the greatest number of transactions in the period, for example. Thus, each expedition is a new opportunity to earn an incentive. In addition, when making adjustments for the service factors (direction of the service, period of time and bus size), drivers should not prefer driving in times of greater demand or in services of greater demand or to drive only large buses.

4. Results

To evaluate the results of this experiment, an analysis of differences in differences was performed consisting of the execution of a linear regression such as that of equation 3. First, the variable Y corresponds to the variable of interest, for example, the number of Bip transactions per expedition or the average speed of the expedition. In addition, this form of linear regression is composed of three dummies variables, the first represented by T that takes the value 1 when the observation (expedition) corresponds to an individual in the treatment group, then the dummy variable P takes the value 1 when the observation is within the treatment period and, finally, the interaction $P * T$ takes the value 1 when the observation is in the period of treatment and corresponds to a treated individual. Given this, the parameter of interest corresponds to the coefficient δ_1, since its estimate corresponds to the difference of the treatment group between period 2 and 1 minus the difference of the control group between period 2 and 1, as explained in equation 4.

$$Y = \beta_0 + \beta_1 * T + \beta_2 * P + \delta_1 * (P * T) + u \quad (3)$$

$$\widehat{\delta_1} = \left(Y_{T,2} - Y_{T,1}\right) - \left(Y_{C,2} - Y_{C,1}\right) \quad (4)$$

[1] As reported by the Central Bank of Chile, the average value of the dollar for the month of August 2019 was 1 USD equal to 713.7 CLP and for the EUR of 794.5 CLP. Thus, the incentive of 100 CLP would correspond to 0.14 USD and 0.13 EUR.

The observation unit for the regression corresponds to an expedition, thus we analyzed all expeditions carried out by each of the drivers belonging to the three terminals studied for the days between May 1st and September 1st. Consequently, the estimated regressions for group 1 consider as a treatment period from August 5th to August 18th and as a treatment group only drivers belonging to group 1 independent of the terminal they worked. Then the regressions for group 2 consider as a treatment period from August 19th to September 1st and as a treatment group all drivers belonging to group 2 of all terminals.

The first variable of interest to our study is the level of transactions Bip obtained for each expedition. Then, Table 2 presents the result of estimating the factors presented in equation 3. Thus, the variable of interest (Y) represents the Bip transactions obtained in an expedition, then β_0 represents the value presented as "constant" in the table, β_1 represents the value presented as "treatment", β_2 represents the "period" and the coefficient of interest (δ_1) which represents the effect of the incentive on the Bip transactions of an expedition is presented as the value of “Interaction”. In addition, the first two columns correspond to a regression that considered as control group all the other bus drivers of the three terminals, the second two columns compare only the bus drivers who accepted to participate in the experiment.

Table 2. Differences in differences for Bip transactions (TRX) for group 1 and group 2.

	All drivers from all terminals		Drivers that agreed to participated as treatment	
TRX	Group 1	Group 2	Group 1	Group 2
Treatment	1.151**	0.751**	0.209	-0.298*
Period	-0.246	0.554**	-0.879**	0.393
Interaction	-0.282	0.366	0.350	0.527
Constant	19.684**	19.686**	20.626**	20.734**
R^2	0.00	0.00	0.00	0.00
N	186,532	186,532	107,898	107,898

Significance levels: * 5% and **1%

First, in table 2 it can be seen that the effect of the incentive on Bip transactions was not significant in any of the 4 regressions. For example, for the regression that considers only the bus drivers that participated, it is found that for group 1 the effect was 0.350 more Bip transactions over an average (“constant”) of 20.6 transactions, however, this result does not have statistical significance. In the same way, group 2 had an effect of 0.527 more Bip transactions, on an average of 20.7 average transactions, but again its statistical significance was low.

On the other hand, these results might say that the incentive did not have an effect on the behavior of drivers, and there could be many reasons for this: may be our incentive was not attractive enough and drivers did not change their efforts, may be the drivers did not understand our experiment. However, the possible differences in performance have not been considered according to the operating context of each service: maybe there is not much room for the driver to improve his performance.

Thus, the type of service in which the expedition was carried out was considered as an operating factor, since each service has a different user demand. Indeed, under a closer look on the services included on our experiment, most of them are feeder services with short routes, most of the load profiles show that there is only one stop during each route that receives a greater load of passengers (coming from a Metro station) and the rest of the stops are mainly used by passengers to get off the bus. Consequently, we will focus our analysis in the service with largest route located in terminal 1.

A linear regression was applied as presented in equation 3, but considering only the expeditions carried out by this service, and the rest of the non-participating drivers were also considered as a control group. Consistent results were found for both group 1 and group 2 (Table 3). There was a significant treatment effect and considerable in magnitude of 2,878 and 2,586, respectively. These increases correspond to 8.7% more Bip transactions compared to an average of almost 33 transactions per expedition with 99% significance for group 1 and an increase of 7.9% to 95% statistical significance for group 2.

Table 3. Differences in differences of Bip transactions (TRX) for longest service.

	Longest service	
TRX	Group 1	Group 2
Treatment	-0.975**	0.178
Period	-4.387**	-4.296**
Interaction	2.878**	2.586*
Constant	32.966**	32.603**
R^2	0.00	0.00
N	20,204	20,204
Significance levels: * 5% and **1%		

Therefore, that the experiment had a significant effect on drivers working on this service and their behavior was modified. At the same time, it could be said that the drivers working in terminal 1, which includes other four services did strive to increase the level of Bip transactions per expedition, but the only one that achieved a significant effect was the longest service. This may be due to intrinsic characteristics of the service and/or to the behavior of the demand, among other factors that allowed the change in the behavior of the drivers to be reflected.

Regressions were also performed considering the type of service along with the direction of the expedition (round trip) and the type of service along with the size of the bus. However, no clear trend was found in the results, since in some cases we found a significant effect, for example, for the large bus size, but not for the small bus size. We believe this is due to the fact that by making more specific analyzes the sample is reduced (we have fewer expeditions for each case).

Another important issue for this research is the presence of multitasking since drivers perform several tasks while driving. To demonstrate a possible presence of multitasking we selected a variable related to a possible change in Bip transactions: speed of the bus.

Moreover, an incentive for drivers to increase transactions may encourage to increase or decrease bus's speed and thus change the operation and safety of passengers.

We analyzed the average expedition's speeds for this service. The average speed is calculated as the time it took to carry out the expedition with respect to the kilometers traveled. The regression used is the same as the one used for the calculation of Bip transactions per service, but now the variable of interest is average speed. In this case, the rest of the terminal's drivers were also considered as a control group. Table 4 shows these results for groups 1 and 2.

Table 4. Differences in differences of average speed for the longest service.

	Longest service	
Average speed (km/h)	Group 1	Group 2
Treatment	0.506**	-0.209**
Period	0.049	-0.301*
Interaction	-0.610*	-0.739**
Constant	19.814**	20.079**
R^2	0.00	0.00
N	20,204	20,204
Significance levels: * 5% and **1%		

Table 4 shows a decrease in the average speed of 0.61 kilometres per hour with 95% significance for the case of group 1. For group 2, a decrease of 0.739 in speed is presented with 99% significance. Although it may seem that these magnitudes are not important, being the average speed of the expedition becomes relevant. In addition, since this service had an increase in transactions, it coincides operationally with a decrease in its average speed, since transporting more passengers might imply more stopping time at bus stops (for boarding and get off), alighting at more stops to carry more passengers, and/or asking more passengers to pay the bus fare.

5. Discussion

In this section we will proceed to discuss the results and possible limitations of this study. First, it should be considered that the incentive was defined with the objective of modifying the driver's behavior as much as possible. However, the actual effect that this may have on operation depends on more factors that only the willingness of the driver to increase his transactions and receive the monetary incentive. The characteristics of the service are also relevant, and this may explain why a significant effect was obtained only in one of the services studied. Probably, the drivers of said terminal have made an effort to increase their amount of Bip transactions, however, they could only achieve it with that service that had a longer route and a variable load profile. The load profile of this service has more than one stop of progressive loading and unloading, allowing the driver to have a better control of his effort's performance regarding the number of passengers transported.

A second explanation for finding the effect only in one service could be related to the amount that drivers could obtain for the incentive. Perhaps it was difficult for them to calculate how much money they could earn with this incentive, and/or perhaps the amount offered for each extra transaction above the threshold (100 CLP) could be unattractive considering how many extra transactions are needed to obtain a significant total amount, especially in shorter services. The difficulty of calculating their bonus could have affected the effort exerted by drivers to achieve the proposed goals.

Regarding the profit of the drivers, on average they obtained an incentive equal to 15,143 CLP and a maximum of 37,300 CLP, which corresponds to 3% and 7% respectively of the net base salary of a driver. From these results, it is important to highlight that said amount was obtained for the performance of only two weeks of the month and it should also be considered that since before 2007, drivers do not usually consider Bip transactions as an important factor in their work of conduction (Tiznado et al., 2014). Therefore, obtaining these amounts could be considered significant.

A third reason could be that the level of the proposed thresholds defined as 20% better may have been very demanding, which could have produced a lack of interest in drivers when thinking that these goals may not be attainable. A fourth reason could be that this incentive was only applied for 2 weeks for each group, which could be considered as insufficient time to see real behavior changes.

Another possible limitation in our study is that one might think that group 2 perhaps had an advantage since group 1 was exposed to the experiment first and therefore group 2 could have learned from the experience of the first group. However, the participating drivers did not receive feedback on their performance until at least 1 month after the end of the experiment. Therefore, we hope to have diminished the level of advantage of one group over the other.

It is worth mentioning that the transport system before Transantiago had safety issues regarding the speed of buses. Since drivers' salaries depended almost totally on the number of passengers they transported, drivers raced through the street to capture more passengers. Therefore, in this study was relevant to check if introducing an incentive that depends on the Bip transactions may present again these safety problems again. However, according to our results, the effect found was a reduction in speed. Perhaps, unlike the previous system, this reduction in speed may appear thanks to the implementation of a mixed salary, with a fixed part greater than a variable part related to the driver's performance with respect to indicators that measure the quality of the system.

Another reason that possibly explains why there was no increase in speed, is that in the context of this experiment each time only a group of drivers were our control. Thus, drivers did not know if the buses in front or behind were also participating in the experiment. Therefore, there may not have been a competition on the street to attract more passengers and, consequently, an increase in speeds. Moreover, bus drivers still have a fixed salary and our monetary incentive was designed to be attractive, but not that attractive as to replicate a *war for the penny* among participants.

Finally, it is worth mentioning that the analysis presented in the results section shows the comparative results of the performance obtained by the different groups of drivers, however, it would be interesting to analyze how the incentive affected the total number of transactions of the analyzed service. Given this, it was observed that for the month of August there was an increase of 8% more transactions compared to the month of July and 7% compared to the month of June. However, this increase may be due to other factors associated with the demand for transportation. Therefore, further analysis is needed to determine if there was an increase in the number of total transactions or if the effect shown in this work was only a regrouping of transactions between the participating and non-participating drivers of the experiment. At the moment, it is observed that this type of incentives produced a change in behavior on the part of the bus driver.

6. Conclusions

This paper shows the results on a monetary incentive for bus drivers applied in three bus terminals of Transantiago. Our results show that this type of incentive may have on effect on drivers' behavior in order to increase transactions per expedition, as long as there is a sufficiently long route allowing drivers to wield an effort. Thus, our research provides suggestive evidence that inefficiencies may be occurring in the operation because of the lack of adequate incentives for drivers.

Another important objective of this research was to test a possible existence of multitasking, and we studied the effect of this incentive on buses' average speed. Our results show that there was indeed a reduction in the speed of the service for the drivers that also increased the number of Bip transactions. However, this reduction in speed could occur for two reasons. Firstly, because carrying more passengers operationally decreases the cycle time. Secondly, due to a speed reduction behavior by the driver to attract more passengers. In this study we only found evidence of a possible presence of multitasking, but more analysis is needed to verify this phenomenon

Furthermore, our results suggests that when applying an incentive regarding Bip transactions may caused an effect on speed, so future incentives should consider and analyze these effects, since the quality of the service can be seen committed to diverting the attention of the bus driver between different aspects.

Although it was possible to suggest a multitasking effect by comparing Bip transactions with average speeds, we believe that there are other variables that may be affected but could not be analyzed. Further research should also focus on how other variables might be affected, such as regularity of the service, the quantity of overtime done by drivers, the percentage of driving with respect to the number of hours of work, punctuality in the time of admission, number of expeditions per working day, among others. If there were information on how these other variables are affected, an incentive could be constructed that was better aligned with the objectives for which it is created.

Finally, we wish to highlight that this study, on the one hand, could be a first step to evaluate how the behavior of Transantiago bus drivers changes through the application of an incentive. With the aim of producing an alignment of objectives between the operating company and the driver, which can be achieved by means of such incentive being related to operator performance indicators. On the other hand, this type of incentive (related to Bip transactions) could be a successful way to mitigate the non-payment of fees that the system currently suffers, but this scheme does not necessarily represent the incentive that works best in the quality of service of the system.

References

Avenali, A., Boitani, A., Catalano, G., D'Alfonso, T., Matteucci, G., 2016. Assessing standard costs in local public bus transport: evidence from Italy. Transport Pol. 52, 164–174.

Avenali, A., Boitani, A., Catalano, G., D'Alfonso, T., Matteucci, G., 2018. Assessing standard costs in local public bus transport: a hybrid cost model. Transport Pol. 62, 48–57.

Batarce, M., Galilea, P., 2018. Cost and Fare Estimation for the Bus Transit System of Santiago. Transport Policy (64), pp. 92-101.

Bolton, P., & Dewatripont, M. (2005). *Contract Theory* (The MIT Press (ed.)).

Börjesson, M., Fung, C. M., Proost, S., & Yan, Z. (2019). Do small cities need more public transport subsidies than big cities?. *Journal of Transport Economics and Policy (JTEP)*, *53*(4), 275-298.

Basso, L.J., Jara-Díaz, S.R., 2010. The case for subsidisation of urban public transport and the Mohring effect. J. Transport Econ. Pol. 44 (3), 365–372.

Dalen, D.M., Go�mez-Lobo, A., 2003. Yardsticks on the road: regulatory contracts and cost ef ciency in the Norwegian bus industry. Transportation 30 (4), 371–386.

De La Vega, J. M. (2018). *Análisis microeconómico de la relación entre operadores y choferes de buses en el transporte colectivo a través de la teoría de contratos*. Pontificia Universidad Católica de Chile.

Duflo, E., Glennerster, R., & Kremer, M. (2007). *USING RANDOMIZATION IN DEVELOPMENT ECONOMICS RESEARCH: A TOOLKIT* (No. 6059; Issue 6059).

Gagnepain, P., Ivaldi, M., 2002. Incentive regulatory policies: the case of public transit systems in France. *Rand J. Econ*. 605–629.

Gómez-Lobo, A., & Briones, J. (2013). Incentive structure in transit concession contracts: The case of Santiago, Chile, and London, England. In *The Clean Air Institute*. http://www.cleanairinstitute.org/caifiles/file/Incentive-structure-in-transit-concession-contracts.pdf

Hartman, R. J., Kurtz, E. M., & Moser, E. K. (1994). *TCRP Synthesis 3 Incentive Programs to Improve Transit Employee Performance CHAIRMAN The City College of New York*. http://onlinepubs.trb.org/onlinepubs/tcrp/tsyn03.pdf

Hensher, D.A., Houghton, E., 2004. Performance-based quality contracts for the bus sector: delivering social and commercial value for money. *Transp. Res. Part B Methodol*. 38 (2), 123–146.

Hensher, D.A., Stanley, J., 2003. Performance-based quality contracts in bus service provision. *Transport. Res. Pol. Pract.* 37 (6), 519–538.

Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *The Economic Nature of the Firm: A Reader, Third Edition*, *7*(January 1991), 232–244. https://doi.org/10.1017/CBO9780511817410.018

Poblete, J., & Spulber, D. (2012). The form of incentive contracts : agency with moral hazard , risk neutrality , and limited. *RAND Journal of Economics*, *43*(2), 215–234.

Puga, D. (2017). *Análisis de incentivos contractuales y propuestas para el rediseño de Transantiago*.

SECTRA. (2016). *Análisis de la Demanda de Conductores del Sistema de Transporte Público Remunerado de Pasajeros mediante Buses en Santiago*.

Sepúlveda, J.P. & Galilea, P. (2020). How do different payment schemes to operators affect public transport concessions? A microeconomic model. *Transport Policy*, (93), pp. 27-35

Tiznado, I., Galilea, P., Delgado, F., & Niehaus, M. (2014). Research in Transportation Economics Incentive schemes for bus drivers : The case of the public transit system in Santiago , Chile. *Research in Transportation Economics*, *48*, 77–83.

https://doi.org/10.1016/j.retrec.2014.09.034
Rojo, M., dell'Olio, L., Gonzalo-Orden, H., Ibeas, Á., 2015. Inclusion of quality criteria in public bus service contracts in metropolitan areas. Transport Policy. 42, 52–63.
Van de Velde, D., Veeneman, W., Schipholt, L.L., 2008. Competitive tendering in The Netherlands: central planning vs. functional specifications. Transport. Res. Pol. Pract. 42 (9), 1152–1162.
Vigren, A., & Pyddoke, R. (2020). The impact on bus ridership of passenger incentive contracts in public transport. In *Transportation Research Part A: Policy and Practice* (Vol. 135). https://doi.org/10.1016/j.tra.2020.03.003
Vega, C. (23 de mayo de 2019). Condiciones de operación y laborales RedBus Urbano. (A. Chaparro, interviewer) Santiago, Región Metropolitana, Chile.

Acknowledgements

Patricia Galilea and Alejandra Chaparro (Project administration) would like to thank the support of the Center for Sustainable Urban Development (CEDEUS, CONICYT/FONDAP 15110020) and the BRT + Centre of Excellence funded by the Volvo Research and Educational Foundations.

Joaquin Poblete (Funding acquisition) acknowledges financial support from ANID PIA/APOYO AFB180003 and Institute for Research in Market Imperfections and Public Policy, MIPP, ICM IS130002.

Juan Carlos Muñoz y Alejandra Chaparro (Funding acquisition) acknowledges support from FONDECYT (1150657).