EXECUTIVE SUMMARY

INFRASTRUCTURE AND EDUCATIONAL PROGRESSION

Marcelo Neri[1]
Rodrigo Moura[2]
Paulo Correa[3]

This report assesses the correlations between access to infrastructure and school performance indicators such as enrolment, attendance, grade repetition and academic proficiency. Infrastructure variables used are: access to public services (water, sanitation, electricity and telecommunications) and private infrastructure (number of people in the household, property rights). The research comprises the generation, description and analysis of several surveys. We have used a household survey (PNAD) to assess the correlations between conditions of access to domestic infrastructure and certain social variables. The social effects of infrastructure provision at home and also at schools were captured through information provided by the Ministry of Education, based on a performance evaluation of the SAEB.

Bi-variate Analysis
The objective of the bi-variate analysis is to draw a profile of the structure of correlations between infrastructure variables and those of school performance, analysing the role of each attribute isolately – i.e., we desconsidered possible interrelations of the performance variables with other explanatory variables.

Multi-variate Analysis
The multi-variate analysis aims to provide a better controlled experiment than the bi-variate analysis. Its goal is to capture the pattern of parcial correlations between **variables of interest** and explanatory variables. In the multi-variate analysis, we capture the correlations between variables of access to infrastructure and variables of social return, whilst maintaining the remaining constant variables. We have worked with two **variants** of the multi-variate regression model:
a) Regression in minimum ordinary **quadraticos** for continuous variables;
b) Binominal or multinominal logistic regression encompassing two or more categories of discrete variables as endogenous.

[1] Centro de Políticas Sociais do IBRE e da EPGE.
[2] Doutorando da EPGE
[3] Banco Mundial

Results of the Bi-variate Analysis

Exams Approval and Failure – SAEB

We present the dynamics of the infrastructure variables taken isolately in cases of success and failure in a maths exam. It is worth mentioning that the analysis of the "grade repetition" variable (referring to the question "have you ever repeated a year?") presents the following options: no, one and two or more years missed.The analysis that follows is classified by groups according to infrastructure variables: telecommunications (access to the Internet, computer, etc.) and access to public services (access to electricity, water, etc.).

Telecommunications

The consequences of access to computers at home are relatively unknown and the literature on the educational impacts of having a computer at home is, particularly, scarce. The importance of such analysis derives from the evidence found in the literature that the use of a computer at home may compensate for the lack of information technology at school. Our estimates show that students who have access to the Internet at home have had higher academic achievements in all school years, especially in comparison to those students without access to the Internet, in the most recent year. This impact is more remarkable as the grade increases. That is, the difference in achievement between who has and who does not have access to the Internet is greater for A-level students than for students from Year 7, which in turn is greater than for students from primary school. Performance remains better even for students who have a computer at home, even when they do not have access to the Internet.

It has also been observed that the higher the number of computers available for the students, the better their performance at school. We have used a set of statistics for the year 2001, which considers whether or not a student uses the computer to do their homework. We have concluded that there is a rather weak link between the use of a computer for homework and the student's proficiency. Only when the student does not use the computer, he or she will have a lower average mark than his/her peers who have used the computer, especially in Year 8 and the 3rd grade of primary school.

Those pupils and students who do not have a computer with access to the Internet experience an increased rate of grade repetition, confirming the result shown in the case of academic proficiency. The result is the same for all school years – greater access to computer has reduced the average grade repetition rates amongst the students, especially in 2003. Using a computer for homework may decrease grade repetition and even improve a student's performance, as long as he/she does not rely heavily on this strategy.

Public Services

The impact of granting access to infrastructure on human capital accumulation has not been the specific object of studies in the literature, but it has been often used as a control **variable** to assess the effect of certain factors in a given education-related variable (e.g. proficiency, enrolment, etc). As a result, it has become crucial to focus the analysis on aspects of access to infrastructure and academic performance. The question that naturally

follows, and which will be resolved in the multi-variate analysis, is how much influence access to electricity, water and so on, exerts on exams performance – when controlling for factors such as parents' education, State, number of people in the household, etc.

The use of electricity and water have a positive correlation with good academic achievements, but those without access to electricity are in greater disadvantage compared with those who lack access to water. Analysing other variables related to school infrastucture, we have observed that, for instance, well-maintained electrical and hydraulic systems within the school have a positive correlation with students performance. Lighting is another important aspect in the educational environment, as students in better-lit classrooms present improved performance.

Access to infrastructure also seems to reduce reproval rates, contrary to school achievement. Despite the lack of more robust data relating to school infrastructure, it has been observed that schools with more adequate facilities usually present lower repetition rates.

Other factors

A probable element that could explain this difference in performance could be the difference in access to infrastructure between rural and urban areas. A small progress has been noted from 2001 to 2003 in the results for these two types of school, namely 1.83% for rural and 1.2% for urban schools. Urban schools present lower repetition rates than its rural counterparts.

Brazil also presents differences between public (State) and private schools, as the figures on proficiency below reveal. The two networks have evolved little between 2001 and 2003 – around 0.8%. Repetition rate for all school years is far greater in public schools than in private ones, although repetitions have decreased sharply in state schools between 2001 and 2003.

Enrolment, Delay and Attendance - PNAD

Bi-variate analysis now lends itself to factors such as enrolment, **delay** and attendance. Infrastructure variables, such as found in PNAD, are different from the ones present in SAEB, which were analysed in the previous section. Here, we have restricted the concerned population to children and youth between the ages of 5 and 17 years old, who are in active school age.

Public Services

Access to water leads to greater enrolment rates, as access to electricity also encourages greater school attendance. Moreover, in a dynamic context, the enrolment rate has increased amongst those with access to water and electricity, throughout the years. The same rationale applies to delays in academic progression, that is, improved progression rates for those who access water and electricity. In terms of school attendance, results point in the opposite direction, i.e. students with access to such public services tend to have lower attendance rates.

Telecommunications

Pressing ahead with the analysis of the consequences of computer use in education, we have initially mentioned that the possession of this device at home encouraged school enrolment, but interestingly this differential rate (in comparison with pupils/students who lack a computer) has decreased in the past years, whilst remaining quite significant.
The same results apply to the use of the Internet. For the "**repetition**" variable, the average rate is greater for those who do not have either a computer at home or access to the Internet. The same is true for students without either a landline or a mobile phone. More specifically, students with computer, and Internet have increased rates of school attendance, whilst the opposite is true for the ones with mobile phones.

Private infrastructure and other factors
There is a growing relation between the number of rooms in a house and the enrolment, attendance and repetition rates. The spatial variable is specific for the place where individuals live, confirming the expected results, according to which people living in urban centers have greater enrolment, attendance rates and lower repetition rates than those who live in rural areas.

Multi-variate Analysis Results

Students and pupils performance can be influenced by the access to public services such as water, electricity, etc. Better infrastructure, both at home and at school, should improve students productivity increasing progression. It is worth asking, however, which other factors can also influence such variables of interest. Thus the assumed controls will help to isolate more accurately the effects of infrastructure on academic performance. Parents education, for instance, may positively affect their children's productivity, even when school infrastructure is somewhat precarious. Besides, the number of people living in a student's household can also influence his/her performance. We have noted that there is usually an optimal number of inhabitants which would help to improve the student's academic performance. We have also used sex, race and State as controls, since school aspects may be distinct among these groups. The models to be estimated are based on the following general equation:

$$y_I = F(x\beta)$$

Where:
- y_I is the dependent variable or the school performance indicators for which regressions will be calculated separately;
- x is the vector of controls, in such a way that x=(IND; FAM; ESC; INF; ENV). IND is the sub-vector that refers to the individual characteristics, FAM family characteristics, ESC school, INF household infrastructure variables, e ENV for externalities (e.g., living in a shanty town).

Proficiency
Regressions for the proficiency equation have been estimated by s**chool year**. Estimate model was linear for the co-variables. We have estimated the regressions by OLS but we have considered the hypothesis of **heterocedasticity**, we used **White (1980) estimation procedure to account for heterocedasticity.**

Based on these regressions, it is possible to infer that the lack of electricity has a harsher effect than the lack of water, in the majority of cases. The impact of the school infrastructure (maintenance of hydraulic and electrical systems and classroom lighting) also have positive effects on academic results. Access to computers may help students develop "academic aptitude", and achieve good grades as a result. Access to the Internet at home, while not showing a clear direction, it may cause a decrease in the student's productivity. Having a computer at home has a greater weight, in terms of magnitude, than having a computer available at school. This is probably explained by the other student background variables, as well as by the lower number of people competing to use a computer at home. Therefore, generally infrastructure may have a positive impact on academic proficiency, helping to reach an increase of up to 12 points on school grades (in cases where students have a computer at home, compared with those who do not have it). We have also observed an expected result concerning private infrastructure: a positive effect **on the level** and a negative effect in the quadratic **term**.

Grade repetition and attendance

For the independent variables of grade repetition and attendance, we have used a **multinominal logit model** where the basic value of comparison is *never having repeated a year* and *never having been absent from school*. The estimated model **by maximum likelihood** as follows:

$$P(y = j \mid x) = \exp(x\beta_j) / \left[1 + \sum_{h=1}^{J} \exp(x\beta_j) \right], j = 1,2, J = 2$$

Where:

- y is the variable of Grade Repetition
- the vector βj is the group of parametres for J=1 (has repeated a year); j=2 (has repeated twice). As the probabilities must add up to one, we must have:

$$P(y = 0 \mid x) = 1 / \left[1 + \sum_{h=1}^{J} \exp(x\beta_j) \right], j = 1,2, J = 2$$

Hence we have always compared it with the base (y=0) which refers to the student who has never repeated a year. We may simplify the **notacao** of the probability as follows:

$$p_j(x, \beta) = P(y = j \mid x)$$

$$p_0(x, \beta) = P(y = 0 \mid x)$$

Enrolment and ***delay***

For the variables of enrolment and atraso, we have estimated a logit **by maximum likelyhood**. The model estimated here is a particular case in point of the multinominal logit when j=2.

Grade Repetition

Interpreting the variable nres as of 1999, we have that one more inhabitant in the household increases the odds ratio by 0.141289 (comparing one repetition with no repetition), that is, it reduces the probability of not repeating a year and/or it increases the probability of repeating a year once. Most of the infrastructure variables are statiscally non-significant – with the exception of cpu, cpunet, water, nres and nres2 in some cases. Of all of them, only access to computer and to the Internet present a **negative signal** according to the expectations (with the exception of the year 1999) that is, who has computer and access to the Internet at home has a bigger chance of not repeating a year and a lower chance of repeating it. These interactions are mostly non-valid in statistic terms and their inclusion reduces the effect of others and/or making them not meaningful, especially in the 8th grade (secondary school) and 3rd grade primary school.

Enrolment

Infrastructure variables are meaningful and increase the probability of school attendance, in most regressions. Access to public services, such as water and electricity increases the probability of school enrolment. In terms of telecommunications, we observe the same effect, with telephone having the biggest impact in most cases. Private infrastructure creates important impacts. Bigger houses lead to higher attendance rates, due mostly to higher income levels which are generally associated with this variable. Restricting the analysis to 10-year-old (or older) inhabitants, the effect is similar to the previous multi-variate analysis, in which increases probability, but marginally this increase will be reduced. In terms of property rights, a person who owns the deed to a property (even when he/she is still paying for it) has a higher probability of being employed, whilst someone who pays rent has a lower probability of being employed. Considering the remaining factors, who lives in urban areas has a greater chance of being enrolled, compared to those living in gatherings **subnormais** (specifically for the years 2001 and 2004 where data is statiscally meaningful). Finally, it is important to mention that the "per capita income" variable is statiscally equal to zero, that is, enrolment probability differences are not explained by differences in income levels.

School Delay

Most infrastructure variables are significant when considering school delay. They also present an expected parcial correlation in terms of reducing the probability of access even for those who have access **to infrastructure**. Interactions in this case are meaningful.

Attendance

We have observed that the majority of infrastructure variables are not significant in the case of attendance, with the exception of the number of inhabitants per household (comparing low attendance with no absence at all). In the second regression, the variables become significant in greater numbers, when comparing low attendance with no absence cases. As expected, they present a robust economic signal. For instance, the greater the number of inhabitants or a case where a person owns both the house and the land (freehold), the greater the probability of high attendance, in relation to the inhabitants with different characteristics.

Simulators and extensions

The Centre for Social Policy from the FGV has a website with a user-friendly software called **Espelho (Mirror)** based on multi-variate regressions, where each user may insert a number of combined data, such as gender, age, location and others - in particular infrastructure variables – in order to create scenarios that can, for instance, check the probability of enrolment, grade repetition or good academic performance. We have also processed more restricted models which group all of the research, allowing the variable "year" to apply only to the linear coefficient in the regressions. Such models tipically present more consistent results in terms of simulation, reducing fluctuations in the estimates across years.

In this empirical analysis, we have thus studied net and parcial correlations between infrastructure and academic performance variables. Determing a causality direction between exogenous and endogenous variables remains the main issue in the design of policies. In this sense, a natural corollary is the analysis in **differences of differences**, based on the ocurrence of experiences where some communities are benefited by the increase on infrastructure provision, as opposed to other communities that aren't. A good candidate to this experiment could be the Favela-Bairro Programme in Rio de Janeiro, where some low-income communities have been benefited by massive local improvements in the public and private infrastructure (namely access to electriciy, water, sanitation, building improvements, road surfacing, etc.) – even though critics of this programme emphasize that aspects such as investment in human capital have not been directly implemented. The Favela-Bairro Programme, is a useful experiment to identifying the impact of various infrastructure investments on academic performance variables. This analysis would be carried out through a comparison of communities in periods before and after the programme, based on data from the 1991 and 2000 Census. These Census data combined with the possible sub-municipal data would allow a PNAD-like analysis, while also allowing a possible identification of causality between infrastructure provision and improvement in the education-related indicators.