Consultant's Report

August 2012

Impact Evaluation Study of
Tbilisi Metro Extension Project (Georgia)
Evaluation Design and Baseline Survey Report

Prepared by Nina Blöndal
International Impact Evaluation Expert
for Asian Development Bank (ADB)

Asian Development Bank

Table of Contents

EXECUTIVE SUMMARY

This report describes the evaluation design and baseline survey results for an impact evaluation of the Tbilisi metro extension project: a component of the Georgia MFF Sustainable Urban Transport Investment Program, Tranche 1. This work was supported under Regional Technical Assistance 7680, Implementing Impact Evaluation at ADB, which aimed for impact evaluation to be initiated and better integrated in ADB-supported operations and DMCs through the piloting of impact evaluation in each regional department and related DMCs.

The aim of this impact evaluation is to provide evidence of the outcomes of urban transport development for users of the infrastructure and other affected persons. For this purpose the evaluation examines the effect on students, households and businesses located within the project affected area. The impact evaluation aims to be "rigorous", understood here to mean that the impact of the project is estimated controlling for other factors, both observed and unobserved.

In order to provide such rigorous evidence it is necessary to construct a "counterfactual", an estimate of what would have happened in the absence of the project. A single-difference – before and after – approach constructs a counterfactual by comparing levels of key indicators for the project-affected group before and after project completion. But if external factors affect the indicators we are interested in, results may not provide a reliable indication of project impact. An alternative is the difference-in-differences approach. This method compares differences in indicators before and after the project for the project-affected group as well as a control group. It thus controls for any unobserved differences in key characteristics between the groups. It is valid so long as the differences are only in levels, not in the growth rates of these characteristics.

This study plans to use a difference-in-differences approach, examining two different cohorts of students studying close to the project site. It will measure changes over time in indicators such as student expenditure, attendance and grade point average (GPA). This approach is chosen because there appears to be no reason to expect different growth rates between cohorts of students studying at the same university, other than those caused by the project. A single difference approach is applied in the case of households and businesses.

This report carefully lays out the project and the evaluation- and survey-design, including some of the methodological challenges and considerations that led to this particular choice of design. It also summarizes key data from the recently completed baseline survey, laying the groundwork for completing a follow-up survey and impact analysis upon project completion in 2014.

I. INTRODUCTION

1. Recent years have seen a growing focus on the importance of measuring the final welfare outcomes of development projects, programs and policies through rigorous impact evaluation.[1] Whilst such evaluations have been increasing in numbers, the share of infrastructure projects remains low. This is partly because the conduct of quantitative impact evaluations of large-scale infrastructure projects poses specific methodological challenges. These will be discussed below. However, as support for DMCs infrastructure development is one of ADB's five core specializations defined in *Strategy 2020*, and infrastructure constitutes a large share of ADB projects, measuring the impact of such interventions whenever possible becomes increasingly important.

2. The majority of existing impact evaluations of transport infrastructure focus on rural infrastructure, roads in particular, as discussed in Section I.B and Appendix 1. Few rigorous impact evaluations of urban projects have been carried out. The metro extension project has therefore been carefully selected for this impact evaluation in order to contribute new evidence of the user-level outcomes of urban transport development. Because the metro system is already in operation and the component to be evaluated will merely extend it, benefits will accrue largely to households and businesses in the area where the extension is built, as well as to students who attend the university adjacent to the new station, making it easier to estimate the impacts.

3. It is anticipated that the results of the impact evaluation will be useful for the design and implementation of ADB's future urban transport projects in Georgia, other countries in Central and West Asia region, and possibly DMCs in other regions. Many countries in the region have insufficient, neglected, dilapidated, inappropriate or outdated urban infrastructure. The results of this impact evaluation should help assess the contribution that projects to improve this infrastructure can make to the country's development objectives, weighing them up against benefits of other potential investments. The results will also contribute to strengthening design and implementation arrangements of future projects to bring about the strongest possible benefits for households and businesses.

4. This report is structured as follows. Section I describes key impact evaluation concepts and reviews relevant literature; Section II provides an overview of the Tbilisi metro system and project area, describes the project, and lays out the logical framework; Section III discusses theory based impact evaluation, outlines the evaluation questions, and provides an overview of evaluation components and evaluation methodology; and Section IV summarizes key statistics from the baseline surveys. Section V concludes.

[1]The term "rigorous" is understood here to mean that the impact of the project is estimated controlling for other factors, both observed and unobserved.

A. Key Concepts

Attribution

5. A rigorous impact evaluation quantitatively estimates the final welfare outcomes of a project, program or policy. The challenge in doing so is to credibly establish causality, attributing any changes in outcome indicators to the project, in isolation from other ongoing developments in the project area. As such, this type of evaluation might more appropriately be named "attribution analysis".

Counterfactual

6. Usually, the evaluator will be able to observe or measure the factual situation, i.e. what actually happens to the development of any outcome and impact indicators for the project affected population – the so-called treatment group. But in order to examine attribution or causality, we would like also to observe the counterfactual situation, i.e. that which would have happened in the absence of the project.

Control Group

7. Simply put, the project impact is the difference between the factual and the counterfactual. By definition however, we cannot observe the counterfactual. Rather, rigorous impact evaluation approaches attempt to 'construct' a counterfactual by identifying a control or comparison group, which is as similar as possible to the project affected, or "treated", population.

Selection bias

8. A key challenge to constructing a credible counterfactual is selection bias, sometimes referred to as endogenous program placement. This refers to a statistical difference between the project affected population and a given control or comparison group, often stemming from the fact that the project area or population has not been randomly selected. Frequently a project targets either those groups that need it the most, or those with the highest potential for success. For example, it might be directed to a needy group, or to an area with higher than average potential for growth. In certain types of projects there is additional probability of selection bias stemming from self-selection when people, businesses or communities have to sign up or apply to become beneficiaries of a given project, or move close to a new project.

9. In some cases selection bias can be observed or measured – e.g. when related to the level of economic or agricultural development of a particular geographical location or population. More challenging from an impact evaluation perspective however, are those cases where bias is unobservable: for instance stemming from politically motivated selection of areas, or from self-selection by people who are 'entrepreneurial' or more or less risk adverse.

Experimental and quasi-experimental methods

10. Constructing a credible counterfactual and dealing successfully with selection bias to avoid biased impact estimates, is at the core of impact evaluation design. There are a number of methods that can be applied to approach this challenge –which in turn can be divided into experimental or quasi-experimental methods.

11. Broadly speaking experimental methods involve some degree of randomized allocation of the project or program. With random allocation and a sufficiently large sample size, those randomly chosen not to receive the program constitute a perfect counterfactual. In this case project impact can simply be measured as the difference between the two groups following project completion.

12. Quasi-experimental methods, on the other hand, are used when randomization is not feasible. Instead, these methods seek to identify a credible counterfactual using a variety of econometric approaches. A number of these are discussed in Section III.E.[2]

B. Evaluating the Impact of Infrastructure Projects

13. Impact evaluations of large infrastructure projects are still few in number, although several are ongoing. This is so because such evaluations pose specific methodological challenges, and it is particularly hard to make them "rigorous". One challenge is that infrastructure is most often placed in specific locations for political, economic or strategic reasons – thus leading to selection bias as described above. Another, which holds especially true for large-scale infrastructure, is that there is often only one road, bridge or railway being built, not many across different areas. And finally, particularly in the case of urban infrastructure projects, there are often large spillover effects throughout the city. These factors make it quite a challenge, and often impossible, to identify a suitable control group.

14. However, some studies and impact evaluations of transport infrastructure projects have been conducted, described in greater detail in Appendix 1. In addition to traditional ex-ante Cost-Benefit Analyses, some studies have used experimental or quasi-experimental methods, aiming at rigorous counterfactual analysis. The studies show impacts of infrastructure on economic growth, household consumption and poverty rates, employment opportunities, agricultural productivity, local market development, and even school enrolment. Studies of metro development have shown impacts on land use, property prices, employment, and population density. Results indicate that impacts are greater closer to the station than farther away and that they are greater the longer the stations have been in operation. However, there are indications that the effect is not necessarily linear - some find a negative effect on housing directly adjacent to the stations.

15. In Georgia, Lokshin & Yemtsov (2003) examine the impact of rural roads. Based on data from the Rural Community Infrastructure Survey (RCIS) and the Survey of Georgian Households (SGHH) they use a difference-in-differences approach to find that road and bridge rehabilitation projects generate clear economic benefits at the community level. Whilst the impact of on labor market conditions is positive, it is insignificant. On the other hand, they find that the number of small and medium enterprises increased while the importance of barter trade fell as a result of the road projects. Moreover, at the household level, access to emergency medical assistance was found to improve unambiguously.

[2] Other approaches have been discussed in the literature. For example, Baker et al (2000), Duflo et al (2008), Khandker et al (2010); Leeuw & Vaessen (2009); White, H. (2011), ADB (2006).

II. THE PROJECT

A. Project Overview

16. This Project is one component of the first of three tranches in a Multitranche Financing Facility (MFF). All three tranches will develop urban transport infrastructure in Tbilisi and secondary cities. Pre-identified components include construction or improvement of bridges and roads, and traffic management improvements, as well as institutional capacity building. The first tranche has four components: (i) completion of a 1.5 kilometer (km) Tbilisi metro extension on Vazha Pshavela Av. to the university district, serving 150,000 inhabitants; (ii) redevelopment of Gorgasali embankment in Tbilisi, including pedestrianized access to the river from the old city; (iii) upgrading of 20 km of the Mestia urban area road network; and (iv) urban renewal of main avenues in Kutaisi through the introduction of a 26-km cycle network. Non-physical outputs include strengthening of institutional and management capacity of the implementing agency and relevant municipalities and the establishment of skilled and experienced units to oversee management and implementation of the Project. The impact evaluation study will focus on component (i): The Tbilisi metro extension.

B. The Tbilisi metro system and project area

Background

17. The Tbilisi public transport system accounts for 49 percent of daily travel within the city as illustrated in Figure 1. The Tbilisi metro has two lines and with an estimated 260,000 users daily (ADB, 2010: Annex 1). It constitutes 14 percent of daily travel on public transport, according to a recent large scale survey of Tbilisi transport use (see Figure 2).

Figure 1 – Transport by source

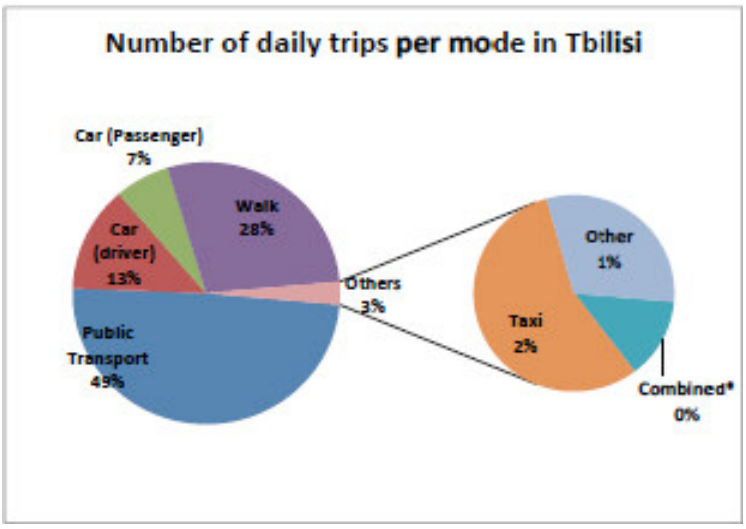
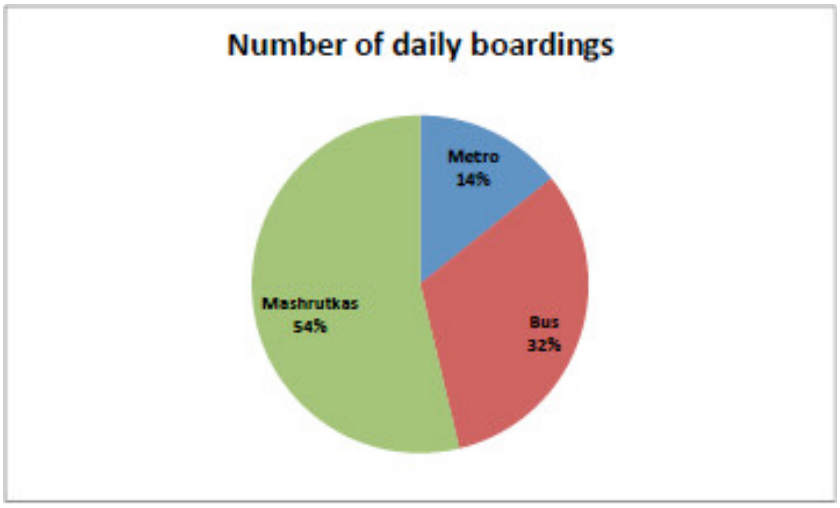


Figure. 2 – Public transport by source



Source: Joly et al (2012).

The Tbilisi metro

18. Line 2 of the Tbilisi Metro, also called Saburtalo Line, opened in 1979. At that time, the line ran from Vagzlis Moedani (Tbilisi Central Railway Station) to Delisi (at Saburtalo District), with a total length of 5.5 km. Plans were then developed for an extension of Line 2 to the Tbilisi University area, with 2 additional stations: Vaja-Pshavela and "University". Construction began in 1985, but ceased in 1993, due to financial and technical problems. Construction recommenced in 1998 and in 2000 the section between Delisi and Vaja-Pshavela was opened, adding 1.2km but only one tunnel. Consequently trains can move only in one direction at a time.

19. The Saburtalo Metro Line is connected to line 1 at Vagzlis Moedani and currently ends at "Vaja Pshavela" station with limited operations between "Delisi" and "Vaja Pshavela". Most passengers traveling to the western parts of Saburtalo (especially the Tbilisi University buildings) therefore have to transfer at "Delisi" station to buses and microbus-taxis. Largely as a result of this, Delisi is one of the busiest parts of Saburtalo District, with well developed commercial and transport infrastructure operating in the vicinity of the Metro station (ADB, 2010, Annex 11:20).

20. The metro generally provides a relatively good service. In 2009, total traffic was about 76.6 million people. Four metro stations have been recently refurbished and extensive maintenance operations are in progress for the rolling stock. The completion of the 1.5 km extension to the university is expected to add 4.4 million passengers per year or nearly 18,000 new passengers per day. (Ibid, Annex 1: 3)

Project site

21. The Project site is located in a part of Tbilisi urban area, which was developed in the 1960s. To the north, east and south the site is surrounded by residential blocks of various densities with 5- to 12-storey buildings. The area located west of the site is made up of industrial and building complexes housing scientific-research institutes, offices, laboratories, mechanical workshops, technical headquarters of cellular communication providers, and others.

22. A number of universities are located in the project area: Tbilisi I. Javakhishvili State University and Tbilisi State Medical University (both of these branches will henceforth be referred to as “TSU”) and the European School of Management. Development of the new metro station connecting the network with the University area will benefit the University students, providing them with additional choice of transport modes. Metro is more expensive than bus transport, but cheaper and more reliable than minibus and taxi. In addition seven primary and secondary schools are located within the project area.

23. The Tbilisi Municipality estimates that the completion of the metro extension to the University station, in addition to improving access to Tbilisi University facilities, Jikia street businesses, and the relatively remote Nutsubidze Plateau 1st District, will also extend residential and commercial development in the area, with creation of additional jobs (ADB, 2010, Annex 11: 35-7).

C. Expected project outcomes

24. Project documents foresee that the Tbilisi metro extension will help lead to an improved urban environment, an expanded local economy and better urban living conditions, as well as to better health, economic development and poverty reduction. It is expected to do so through increasing access to employment, lowering transport cost and travel times, for women in particular, and reducing pollution. In particular, the project design predicts that the poor and socially excluded will benefit from having a new low-cost option to travel to the city center. The following subsections outline these benefits in more detail.

Health and pollution

25. The Tbilisi transport network includes metro, buses and microbuses. The construction site is crossed by 9 bus routes and 9 fixed-run microbus-taxis. Most of the buses and microbuses run on diesel fuel and as most are in a poor technical condition the pollution from emissions exceeds the maximum permissible concentration. Traffic congestion, secondhand cars, the age of the car fleet, and the use of low-quality gasoline generate 80 percent of the air pollution in Tbilisi. The concentration of carbonic acid in the air already exceeds the norm by 1.5–2.5 times and increases health problems.

26. Project documents state that the opening of the metro station is expected to reduce emissions of harmful substances into the atmosphere by reducing the numbers of buses and minibus-taxis in the Delisi area and from a reduction in travel by car (Ibid, Annexes 8 and 11).

27. Whilst this is the most likely outcome as students and residents journey to and from the area by metro rather than bus or minibus, it will be worth observing whether a reduction in traffic at Delisi may be replaced by increased traffic around the new station, potentially reducing the net effect of the metro on air quality improvement. At the same time it should be considered that the minibus fleet is currently being upgraded with newer, more environmentally friendly vehicles, to be completed by May 2012. Disregarding the metro this move may itself affect air quality, particularly in areas with dense minibus traffic.

Transport cost and travel times

28. Public transport fares are generally low and public transport is still the main means of transport for 55 percent of the population in Tbilisi. Of the public transport users, 75 percent do not own a car. Subsidies have kept prices affordable and the municipality still subsidizes fares for around 200,000 citizens below the poverty line. However, unregulated microbuses, which now carry the bulk of passengers, do not accept social and concessionary fares. Besides, transfers between different modes of transport are not favored, as each trip has to be paid for independently. There is also no integration with microbus lines. This limits mobility, which has important consequences: (i) many people prefer to spend more time using a door-to-door bus line rather than to transfer and use the metro; and (ii) people who transfer between a bus line and the metro often do not pay for the use of the bus line, resulting in a fraud rate on the bus network estimated at 30–40 percent.

29. Therefore, the investment program will finance the development of an integrated ticketing system allowing transfers between different modes. During implementation, tariff and fare reform will be explored and a household survey will assess sensitivity to tariff increases in exchange for improvements to services, comfort, and safety (ADB, 2010, Annexes 8 and 11).

Employment

30. The investment program is expected to contribute to poverty reduction by stimulating economic development, according to project documents. In particular, it will increase the mobility of about 1.34 million people living in Tbilisi. It will generate an estimated 1,000–1,500 jobs during implementation for all components – less for the metro expansion alone – and is expected to promote socioeconomic development in the participating municipalities.

31. Whilst the construction phase will see an increase in employment of construction workers it is less certain to what extent the new metro station will lead to long term employment generation. If this is to occur it is likely to come from new and increased business activity in the project area. To explore this, the evaluation will seek to measure any changes in business activity, including employment and analyze whether any such change can be attributed to the metro station.

Benefits for women

32. The project will improve access to public transport for women, and planning, design and construction will address specific needs of women and men riders (e.g., adults traveling with children). These may include: improved security measures such as increased visibility of security personnel; lighting in all transport facilities, including stations, stops, waiting areas, toilets, and interchanges; information on what to do and where to complain in cases of violations of personal safety; and assignment of priority seats in trains for women. Reductions in time burdens are expected to benefit women in particular. In addition women, children and the elderly usually use schools, clinics and hospitals more than other groups. They are therefore predicted to benefit most from improved access.

University students

33. The metro station will be located approximately 500 meters from the Tbilisi State University at Maglivi, making students a main population expected to be affected by the project, facilitating easier and faster travel, time savings and travel cost savings.

D. Project theory and logical framework

34. The project log-frame and causal chain are outlined in Table 1. Besides providing an overview of the project logic they are used as a foundation for determining the evaluation questions, laying out more detailed causal chains for each project affected population, and guiding questionnaire design as discussed in section III. Note however, that whilst the log frame is the framework for analysis, it is not the analysis itself. Rather the evaluation has to interrogate the assumptions underlying the causal chain implicit in the log frame (White, 2011: 132)

Table 1. Logical framework[3]

Component	Inputs	Output	Intermediate Outcome	Final Outcome	Impact
A: Completion of a 1.5 kilometer (km) Tbilisi metro extension on Vazha Pshavela Av. to the university district, serving 150,000 inhabitants	USD 34M* for civil works - construction of: (i) The University station complex and facilities (above and below ground); (ii) Blind alleys behind University station (railway dead-end siding with crossover, forchanging track and train direction) (iii) Remaining part of passage tunnel from Vazha-Pshavela to University station; (iv) railway superstructure (arranged under/for the rail track)	1.5 kilometer metro extension on Vazha Pshavela Av. to the university district completed and in use.	1. A significant decrease in travel time to access the western Saburtalo area from the city centre and the left bank of the river using the Metro. 2. Improved access to Tbilisi University facilities, Jikia street businesses, and the relatively remote Nutsubidze Plateau 1st District 4. Reduced numbers of buses and minibus-taxis in the Delisi area as many passengers will no longer need to use these modes to complete or begin their journeys 5. Employment generation during the implementation phase 6. 2% increase in the number of women using municipal urban transport compared to baseline in 2008	***Positive:*** 1. Increased business activity around the metro extension 2. Increased property prices 3. Improved access to employment 4. Reduced spending on transport 5. Reduced time burdens 6. Reductions in pollution (including noise pollution) ***Potentially negative:*** 7. Decreased business for bus and minibus drivers/companies 8. Drop in property prices for housing immediately adjacent to metro 9. Decreased business in area around Delisi metro station.	Improved urban environment, an expanded local economy and better urban living conditions. Improved health, economic development and poverty reduction.

[3] Expected impacts as reported in DMF for the MFF. Inputs, outcomes, and outputs as reported in ADB 2010.

III. ANALYTICAL FRAMEWORK AND METHODOLOGY

A. Theory Based Impact Evaluation

35. The evaluation will take a theory-based approach. A theory-based impact evaluation looks first at the project or program theory and examines the entire project causal chain. That means measuring not only impacts but all the way from inputs to outputs to outcomes and then to impacts. This makes it possible to analyze which components or which stages of the project work well, and which ones do not. Identification of weak links in the causal chain makes it possible to pinpoint design features that need adjustment for future phases of the project or for similar projects. More specifically theory based impact evaluation aims to measure along the causal chain, i.e. quantitatively measure outputs, outcomes and impacts (see e.g. White, 2009).

36. For this purpose, using the project log-frame in Table 1 as a starting point, separate causal chains were developed for university students, households, and businesses in the project-affected area, respectively. Each of these is outlined below, providing an overview of the individual causal chains.

Figure 3 – Student causal chain

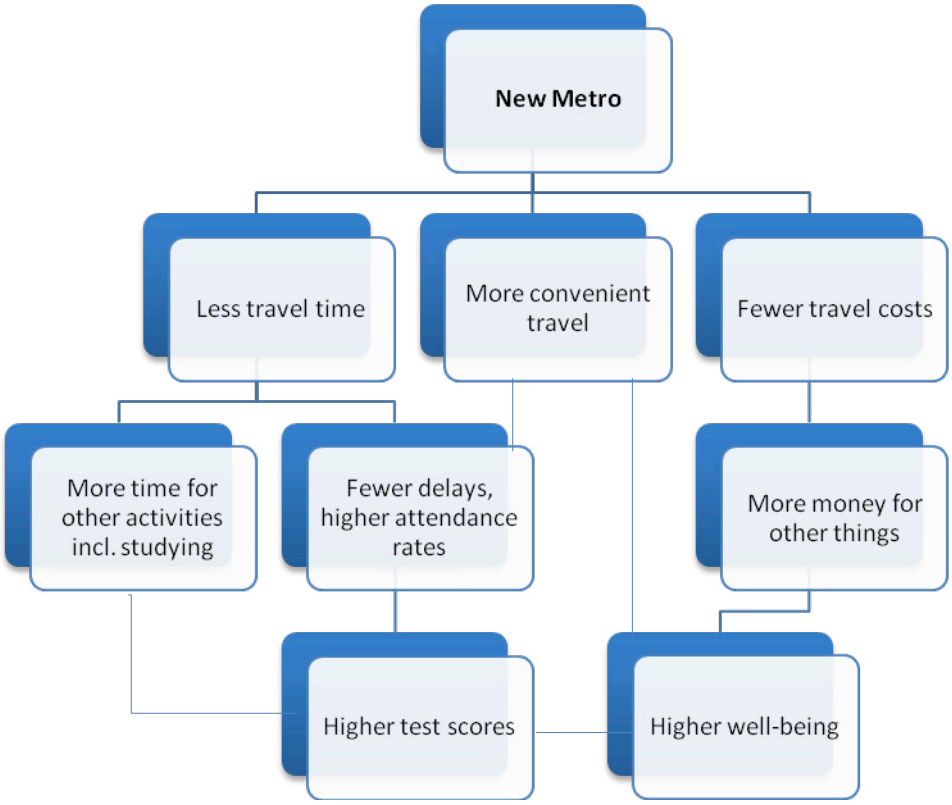


Figure 4 – Business causal chain

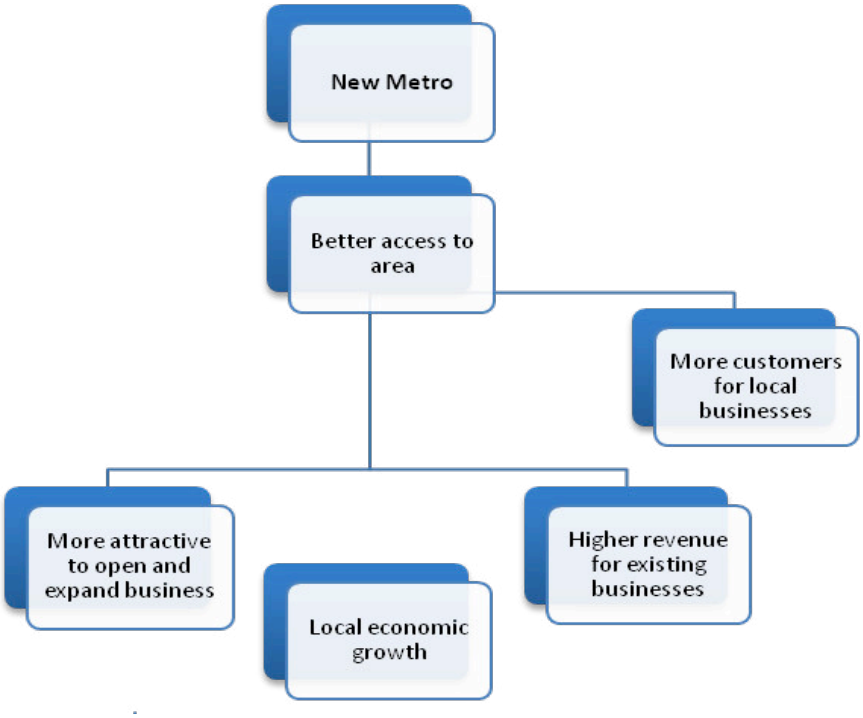
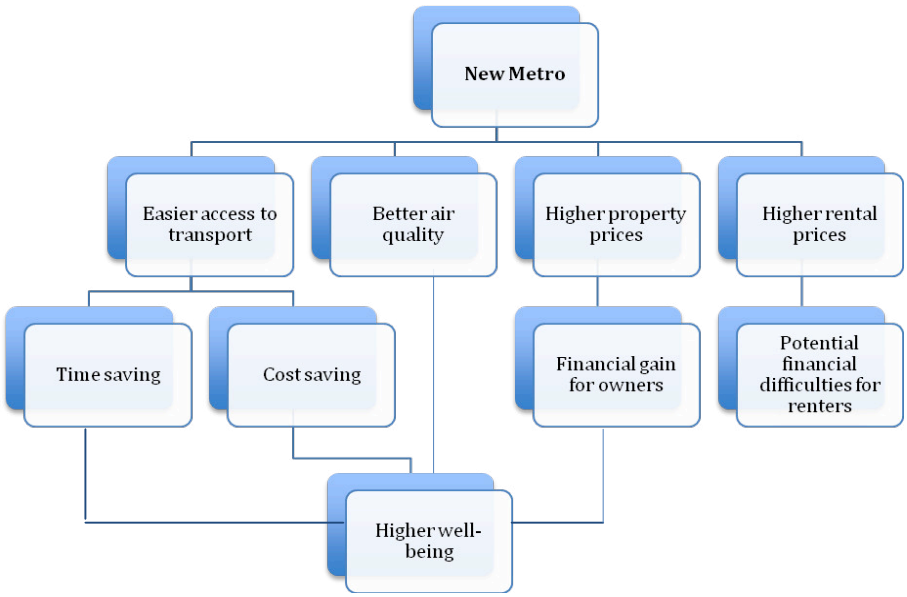


Figure 5 – Household causal chain



37. The figures above also provide guidance for development of questionnaires for each survey, and help evaluators detect any potential "weak links" in the causal chain – i.e. to determine why the project may or may not result in the intended impacts.

B. Evaluation questions

1. How can improvements in urban public transport affect local business activity and population groups dependent on such transport?

More specifically:

2. To what extent will the metro extension lead to local economic development, including increased business activity, revenue, and employment generation?

3. To what extent will the metro station affect university students commuting to and from Tbilisi State University? In particular, how will it affect their time use, expenditure patterns, attendance rates, and test scores?

4. To what extent will the metro extension contribute to improved air quality/ reduction in pollution?

38. All of these will be explored by measuring quantitatively along the project causal chain as outlined in Figures 3-5. Hence, we will explore whether the expected outcomes were achieved and to which extent, and if so whether these led to the expected, or other unexpected, impacts.

C. Evaluation components

39. The evaluation applies a mixed-methods approach to impact evaluation, combining quantitative and qualitative data. The quantitative component is of a quasi-experimental nature, combining a difference-in-differences approach in the case of students, and a single-difference approach in the case of households and businesses as described below.

40. The qualitative component at the baseline consisted of Focus Group Discussions, focusing mainly on transportation use and behavior of households and businesses. The following section describes each component and approach.

Student survey

41. As described in Section II.C, students are one of the main groups to be affected by the project as they travel to and from the university on a daily basis and often use public transport. Therefore the evaluation was designed to examine outputs and impacts along the entire causal chain as outlined in Figure 3, including time use, transport modes, travel costs and related impacts on consumption patterns, attendance rates and test scores for university students.

42. The study applies a combination of the difference-in-differences methodology with a so-called cohort approach. This includes:

i. Surveying a sample of three cohorts of students at Tbilisi State University to obtain information about demographics, income, expenditure, travel costs, time use and other relevant welfare indicators. This covers students enrolled in subject matters

belonging mainly to the Maglivi Branch, located in the project affected area, as well as students taking courses downtown, who should be less affected by the opening of the new metro station.

ii. A pilot survey of 900 students was completed in May 2012, the results of which are discussed in Section IV.A.

iii. Additional surveys of students enrolling in 2013 and 2014 will be carried out. Because the metro extension is expected to finish in December 2014 the early cohorts will function as a comparison group and the late cohort will constitute the treatment or project affected group. The treatment group will be surveyed in September 2014 and September 2015. For details see Section IV.

43. In addition to the control group provided by the cohort approach, the evaluation covers students enrolled at Ilia State University, the foreign languages university, as an additional control group to be surveyed at the same time as the groups above.

Household survey

44. A single difference – before and after – survey will be completed for a sample of 300 households in the project affected area. Originally, the intention was to conduct a household survey using a difference-in-differences approach. The comparison group would be identified using existing household level data such as census data or the Georgian Integrated Household Survey to match comparable sub-districts. However, reviewing the situation on the ground it turned out, not only that such data were unavailable, but also that it would be impossible to identify a comparison area for which the parallel trend assumption (described in Section III.E.1) would be valid. This is the case for the following reasons:

- A wide range of city development interventions are ongoing or planned and will target different city districts at different points during project implementation, in turn affecting factors such as growth potential, employment opportunities, and transportation availability.

- Tbilisi's districts, and sub-districts, differ significantly in terms of infrastructure, services, geographical location and population characteristics. It is highly likely that some of these differences are unobservable. Whilst the difference-in-differences approach eliminates selection bias in levels, the extent of diversity makes it likely that there will also be differences in trends in the absence of any intervention.

- The small size of the city and the lack of district and sub-district-level data makes it particularly difficult to identify suitable control groups.

Survey of business activity

45. The evaluation includes a single difference census of businesses in the project affected area, intending to capture changes over time to the level of business activity. This will be measured by the number of businesses as well as revenue, profit and customer base for businesses in the project affected area. The intention was to conduct a difference-in-differences approach for businesses. However, similar issues apply to this case as to the households, making it probable that the parallel trend assumption would be violated. This was the basis for following a smaller scale single-difference approach.

46. The possibility of measuring project impact on business registrations and business revenue using business registry data from the National Statistics Office, Geostat was also explored. With prospects of obtaining annual data at the district or sub-district level for the years 2008–2015 we hoped it would be possible to apply a difference-in-differences approach and potentially analyzing developments at several points in time both prior to project start and following project completion. So far, sufficiently disaggregated data for this effort do not appear to be available.

Qualitative component

47. In addition to the quantitative analysis of the data listed above, the evaluation includes a qualitative survey to explore household impacts and experiences. At the baseline this consisted of Focus Group Discussions for local female and male residents and business owners.

D. Other components considered

48. A number of other potential components and methodological approaches were explored but were excluded based on the assessment that it would not be feasible to complete these with sufficient methodological rigor.

Real estate prices

49. The new metro station is expected to lead to changes in land and real estate prices, as well as to rental prices. As discussed in the literature review, studies elsewhere have shown a significant effect of a new metro on real estate prices for housing near the new station. Some have also found that prices rose in response to the announcement of the new metro, i.e. prior to completion of the construction work.

50. In order to examine such effects, the evaluation team hoped to use data on real estate prices from the Revenue Office to be obtained for all city districts for each year between 2008 and 2015 allowing for an examination of price changes for all city districts and possibly sub-districts both prior to construction start, during construction and following project completion. Unfortunately, the data provided by the Revenue Office was insufficiently disaggregated for a thorough analysis.

51. Instead questions on rental and property prices were included in the business survey to explore developments over time for at least part of the project affected population.

Measurements of air quality

52. The evaluation team considered detecting environmental impacts by measuring air quality at the project site and a number of other selected sites throughout the city. This could be done by the Environmental Pollution Department at the National Environmental Agency and would include selecting suitable control areas based on a list of criteria including topography, meteorological conditions, size of roads, and distance from the river.

53. Such a survey would test air pollution at the University station as well as in the Delisi area and control areas (as identified above). This would be done through daily measurements over the course of a month each time. However, as was the case with households and

businesses the difficulties involved in obtaining a suitable control led to the decision not to carry out this component.

E. Evaluation Methodology

1. The difference-in-differences approach

54. Difference-in-differences is a quasi-experimental approach to impact evaluation. The difference-in-difference methodology estimates the Average Treatment Effect (ATE) by comparing treatment and comparison groups (first difference) over time (second difference). This is be done by carrying out identical household surveys just prior to project start (baseline) and after project end (endline). In doing so, this technique eliminates selection-bias since what matters is not whether treatment and comparison groups start out exactly at the same level of say, income, but rather the size of the change in this indicator over time for the two groups (see also Figure 6 below).

55. Mathematically, the difference-in-differences approach can be described as follows:

$$DD = E[Y_1^T - Y_0^T | T_1=1] - E[Y_1^C - Y_0^C | T_1=0]$$

Or in an OLS regression framework as follows, where T2 is a time dummy variable, D the treatment dummy, and the coefficient of interest is β_2:

$$y = \beta_0 + \beta_1 D + \beta_2 T2 + \beta_3 D*T2 + u$$

The parallel trend assumption

56. The impact estimates obtained from this approach are only valid however, if the so-called parallel trend assumption holds true. This is the assumption that the two groups would have developed at the same speed in the absence of the project. This may not be the case if change over time is determined by unobserved characteristics of either group. An example would be if particularly profit-oriented businesses are located, or open up, in the project affected area. It may also not be true if non-project related development plans and trajectories for the two areas are entirely different.

57. Figure 6 illustrates the difference-in-differences approach and shows how selection bias is eliminated if the parallel trend assumption holds true: What matters for an unbiased impact estimate is not the starting point, as long as the two groups grow at a the same trend. This is because our impact estimate is the difference between the treatment group (indicated by a red dot) and the comparison group (blue dot), over time.

58. Figure 7 on the other hand, illustrates how failure of the parallel trend assumption may lead to an under- or over estimate of impact. If the comparison group would have grown at a faster rate than the treatment group in the absence of the project, we will underestimate impact, and may even obtain negative impact estimates. Similarly, if the comparison group would have grown at a slower speed than the treatment group in the absence of the project, we will overestimate impact.

Figure 6 – Difference in difference with valid parallel trend assumption

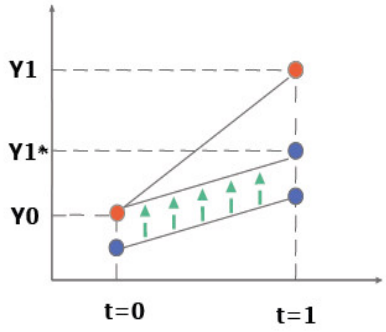
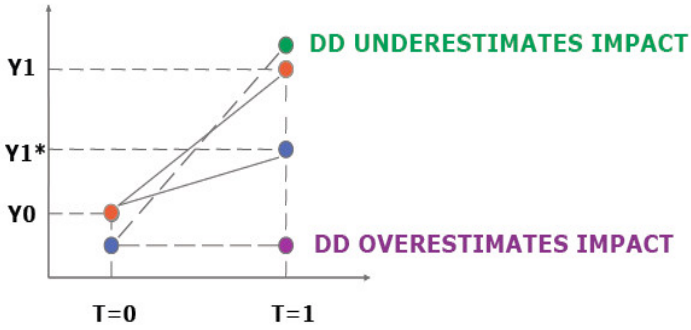


Figure 7 – Difference in difference with invalid parallel trend assumption



Selecting comparison groups at the baseline

59. In order to minimize the likelihood of selection bias and maximize the likelihood that the parallel trend assumption will hold true, the control group should be carefully chosen to be as similar as possible to the treatment group.

60. Often, as a first step, administrative data can be used to identify an appropriate geographical location (or locations) to survey as a "control area". As a second step, sometimes, propensity score matching can be used to select, from within the population sampled in the control area, the businesses, households, villages, etc, most similar to the treatment group (see Heckman et al, 1997).

61. Furthermore, in order to ensure that the parallel trend assumption holds true, evaluators can seek to match households and businesses not only on the levels of certain observable characteristics before the intervention, but also on the pre-intervention trend in these outcomes. Doing so will require availability of time-series household data however.

2. The single difference approach

62. A single difference approach can take two forms: The first is a so-called 'before and after comparison' – which measures differences in outcomes over time for the project affected group. The other measures differences in outcomes between the project affected group and a control

group – but only does so after the project has been implemented, i.e. without collecting data at the baseline.

63. The problem with these approaches is that we usually cannot reliably conclude that any changes measured – whether over time or between groups – can be attributed to the project itself as discussed in Section I.A. This is because changes in outcome indicators over time could be caused by other simultaneous developments amongst the project affected population, or could reflect selection bias between treatment and control group, also discussed in Section I.A.

64. Nevertheless in situations where more rigorous approaches are not feasible a single difference approach can be useful. In this study we consider this to be the case. For businesses, carrying out a census of businesses in the project affected area allows us not only to obtain a complete overview of business development in the entire area over time, it also enables us to detect any significant changes to business activity in terms of new businesses opening up. These developments in turn can be compared to developments in other city districts and city-wide, providing a good overview of any notable differences between areas.

65. Moreover, in cases where the project or intervention is of such a nature that certain outcomes are unlikely to have been caused by other factors, a single difference approach can be highly useful. In our study this is the case when looking at household travel times and travel costs before and after project completion. In the absence of any simultaneous introduction of non-project related price changes or infrastructure developments the single difference survey allows us to draw fairly reliable conclusions on the project's effect on these indicators.

F. Estimation strategy and sampling

Student survey

66. Following the OLS framework outlined in Section E.1, we will conduct a difference in difference estimation for the student survey as follows:

$$y = \beta_0 + \beta_1 D + \beta_2 T2 + \beta_3 D*T2 + u$$

where y is the outcome variable, D is the treatment dummy and T2 is a time dummy. The coefficient β_2 is the treatment effect.

67. The appropriate sample size is estimated using a power analysis approach, undertaking a two sample t-test based on the level of desired power (80%), statistical significance (5%), desired minimal detectable effect (between 5-10% change in test scores), and the standard deviation of the outcome variable from previous years.

68. We use test scores for 4,823 first year students in the first semester of 2012. The students have a mean GPA of 2.05, with a standard deviation of 0.96. In order to detect a percentage change in mean GPA at a 5% significance level and with a power of 80, we need a sample size of 2,750 students. This provides an upper bound on the student sample size. If we want to detect a 10% change in mean GPA with the same power and significance level, we need a sample of 690 students, which provides us with a lower bound. Since the 2012 survey is a "pre-baseline" survey, a sample size closer to the lower bound was selected, with a sample of 900 students split into three groups: Maglivi based students, downtown based students and Ilia University students.

Business survey

69. The business survey was designed to measure business activity, productivity, revenue and profit. Lacking a credible control group for businesses, this survey will cover businesses within the project affected only, enabling us to obtain a good indication of the effect of the metro on the listed outcome variables, but not allowing us to draw statistically valid conclusions on causality.

70. In the absence of a control group a panel survey will be completed for the treatment group only, and a first difference estimation will be applied. Since first differencing requires variation in our treatment variable over time as well as between businesses, and because a simple dummy variable for treatment would take the same value for all businesses in the sample, we choose distance to nearest metro as the treatment variable, D. We will conduct the following equation where y is business revenue/profit, D is the distance to metro variable and θ_1 the effect of D on y, which is expected to be negative coefficient; **x** is a vector of additional predictors of business revenue and profit, and u an error term.

$$\Delta y = \theta_1 \Delta D + \boldsymbol{\beta} \boldsymbol{\Delta x} + \Delta u$$

71. Other considerations include adding D^2 or divide the project-affected area into strata by distance to the metro so as to create a discrete treatment variable rather than a continuous one.

72. As for the student survey, estimation of sample size will depend on the desired level of power, significance level, and the desired minimal detectable effect. In addition it will depend on the number of predictors, the variance of D, and the variance of the error term, u. These would be based on estimates using similar data from previous surveys. Hence to arrive at a sample estimate using this approach we would need to run the regression on existing data. In this case unfortunately, we do not have access to such data and therefore the sample size was based on experience with similar surveys.

Household survey

73. We will follow the same process for households as that explained for businesses. Outcomes will include transport expenditure, non-transport consumption/expenditure, time use, and perceptions of pollution and well-being.

G. Other methods considered

Regression discontinuity design

74. In addition to the planned-for difference-in-differences approach, the evaluation team explored the option of applying a regression discontinuity approach. This approach can be used when there is a threshold, which defines a treatment vs. a control group. This can, for instance, be in the form of project eligibility criteria such as land size for participation in microcredit schemes, age for participation in pension schemes or test scores for scholarship programs. Evaluators can exploit this threshold to compare outcomes for people just below and above the threshold, based on the assumption that people above and below the given threshold have otherwise similar characteristics.

75. Regression discontinuity designs can be so-called “sharp” or “fuzzy”. In the sharp design the probability that D = 1 changes from zero to one as the running variable crosses the threshold, c. In the fuzzy design, the probability of treatment jumps discontinuously as X crosses c, but it does not jump by 100 percentage points. In other words, either some people with $X < c$ get treated, or some people with $X \geq c$ do not get treated, or (most likely) both.

76. The RD approach gives us an unbiased treatment effect at the threshold (discontinuity) but this is what we call a “Local average treatment effect” (LATE), which may not be generalizable to the population at large, i.e. those further from the threshold.

77. In the case of the metro extension, the evaluation would exploit the fact that the metro will benefit residents within a limited distance from the metro station only. People living further away from the station are not expected to benefit significantly from the project, which would allow for a fuzzy regression discontinuity design.

78. However, given the location of the metro on the outskirts of town this approach did not turn out to be appropriate, because beyond a short distance from the metro there are no houses to be sampled. The sample of households for the pilot survey was drawn from amongst all households living within a one kilometer radius from the University metro station, and 95% of the households were located 300, 400 or 500 meters away. If we were to use an RD approach we should have good reason to believe that the effect of the metro on travel time, travel use and travel costs would diminish statistically beyond 300 or 400 meters, which we do not have.

Instrumental Variables

79. The evaluation also considered whether an instrumental variable approach would be feasible. An instrumental variable approach to impact evaluation would rely on identifying an “exogenous” variable which determines whether or not a household or business benefits from the project. To be exogenous, this variable would need to be unrelated to any unobservable characteristics of the households or businesses. But for the approach to work, this variable would need to play a strong role in determining whether or not they benefit from the project. The use of this variable to “instrument” being affected by the project would remove any bias caused by unobserved differences between the treatment and control groups. For example, imagine that homes in different parts of the city were allocated to households on the basis of the first letter of the family’s surname, and there was no reason to believe that this characteristic is related to any unobserved household characteristics. Even if households subsequently moved within the city, there would remain a link between the surname and the location for some years. This variable could therefore be used as an instrument for benefitting from the project. This would remove the bias caused by endogenous placement- for example if more entrepreneurial households deliberately choose to move close to the new metro to take advantage of business opportunities. In this case, however, no suitable instrumental variable for access to the metro was identified.

IV. SUMMARY OF BASELINE SURVEY RESULTS

80. Student, household- and business surveys were completed during May and June 2012. The following sections summarize main variables of interest.

A. The student survey

81. As discussed in sections III.A and III.C the student survey was designed to measure changes in travel time and travel cost, as well as the effect on attendance, grades, time use and consumption patterns.

Time use and travel

82. Table 2 summarizes student time use. Of particular interest is the time spent traveling during the week preceding the survey – a full nine hours, a substantial amount of time compared to the average of 16 hours of lectures they attend.

Table 2 – Student time use

Hours spent on	Obs	Mean	Median	Std. Dev.	Min	Max
Lectures	898	16.35	16	8.11	0	42
Other studying	898	13.92	12	10.23	0	70
Going out	900	2.52	0	4.19	0	35
At home with family	900	15.96	14	12.16	0	84
Sports	899	1.36	0	3.69	0	28
Watching TV	900	5.15	2	6.63	0	50
Sleeping	900	50.97	49	11.07	7	92
Internet	900	14.61	14	10.67	0	70
Traveling	890	9.03	8	4.87	0.83	28

83. If we look at each of the three campuses separately as in Table 3, we see that Maglivi based students spend most time traveling with an average of ten weekly hours whilst downtown based TSU students and Ilia State students spend 8.9 and 8.2 hours, respectively.

Table 3 – Student travel time by campus (hours)

Campus	Obs	Mean	Median	Std. Dev.	Min	Max
TSU Maglivi	298	10	10	5.3	0.8	28
TSU Downtown	295	8.9	8	4.9	1	28
Ilia State	297	8.2	7	4.2	1	21

84. The data also show that nearly all students (98 percent) most frequently travel to university from home and vice versa, and they do so, on average, on 20 days each month. As shown in Table 4, in the case of students studying primarily at Maglivi this travel mainly takes place using bus (59 percent), minibus (40 percent), or metro (27 percent). For those studying downtown or at Ilia State University 72 percent use bus, 21 percent use minibus and 26 percent use metro. The assumption is that a proportion of the 40 percent using minibus to reach Maglivi will be replaced by metro-users.

Table 4 – Means of public transportation used by students

Transport	Percent using, Maglivi students	Percent using, Downtown/ Ilia students
Bus	59	72.8
Minibus	40.3	20.7
Metro	26.6	25.9

Expenditure

85. Overall, as illustrated in Table 5, students pay few of their expenses themselves. The mean amount paid in tuition fees was 1532, which in 98 percent of the cases was paid by the parents or other family. Very few students had a student loan. Close to one fifth pay something for accommodation (on average 273 GEL/166 USD per month), but only four students contribute towards this payment themselves.

86. In most cases (85–90 percent) students have daily expenses paid for by parents or other household members. The most frequently consumed items during the week prior to the survey were transportation, which 94 percent of students used, and cell phone, which 92 percent of students used. The vast majority of students had all related expenses paid for by others. This is also true for printing and photocopy expenditures – and for those who consumed items such as sandwiches and snacks, or who went to the cinema, café or bars, 90 percent of the expenses were paid by parents or other household members.

Table 5 – Selected student expenditure

Item	Consumed (percent of all students)	Student paid or contributed towards cost (percent of those who consumed)	Number of students who paid or contributed towards cost	Total amount paid, mean	Paid by student, mean	Paid by student, median
Tuition	100	2,2	15	1531,5	1070	1125,0
Accommodation	21,3	2,5	4	273	3,7	0,0
Sandwich/snacks	22,8	16,5	34	NA	8,8	5,0
Cafe/bars	26,4	10,1	24	NA	58,4	20,0
Cinema	14,8	10,5	14	NA	13,1	10,0
Accessories/bags	25,8	9,5	22	NA	24,3	14,5
Presents/personal treats	38,6	10,1	3	NA	42,9	20,0
Personal care	41,3	11,6	43	NA	12	6,0
Cell phone	91,6	16,4	135	NA	19,8	10,0
Print/copy	77,6	13,8	96	NA	14,7	4,0
Stationary etc	57,3	11,2	58	NA	5,6	3,0
Transport	93,9	13,7	123	11,9	1,8	0,0

87. In the case of transport costs a greater number of students paid or contributed towards costs, but the amounts were relatively small: The average cost of traveling during the week preceding the survey was 12 GEL (USD 7.3) of which students paid just 1.8 themselves.

88. Examining travel-related costs by campus (Table 6) we see that Ilia State students spend most, Maglivi based students come second, and downtown TSU students spend the least.

Table 6 – Travel cost by campus.

TSU Maglivi	Obs	Mean	Median	Std. Dev.	Min	Max
Transport cost	305	11.07	7.8	12.83	0	100
Of which paid by student	305	2.17	0	7.11	0	60
TSU downtown						
Transport cost	295	9.89	6.2	12.31	0	100
Of which paid by student	295	0.6	0	4.13	0	40.6
Ilia State						
Transport cost	300	14.73	9	21.37	0	190
Of which paid by student	300	2.56	0	9.08	0	120

89. Nine percent, 83 students, had a job during the last year, working on average 109 hours a month in the case of those we have data on. Whilst the data reveal that parents pay the majority of student expenditures, only 112 students responded to a question on parental income during the last year – and with a mean of less than 12000 GEL (7300 USD) it seems to be negatively biased considering the population group, though it is higher than the 2011 average household income in urban Georgia at GEL 9149 (5564 USD).

Attendance

90. Close to a third of students - 276 of them – did not miss a class in the month preceding the survey. For the remaining students the average number of classes missed was 3.8. The main reason for doing so, for 40 percent of students, was that they were "busy with other activities". A lower number, 11 percent, reported that missing class was due to delays in public transportation. This of course is a number worth watching in subsequent survey rounds – even though it is lower than the number of students missing class "because they were not in the mood" (15 percent).

Table 7 – Reasons for missing class

Reason	N	Percent
Too little time between Maglivi and downtown classes	6	1
Delays in public transport	26	4.3
Busy studying	29	4.7
Busy with paid work	28	4.6
Busy with other activities	169	27.6
Not in the mood	77	12.6
Not a useful class	5	0.8
Illness	216	35.3
Other	56	9.2
Total	612	100

91. When asked about arriving late to class however, public transport and traffic jams played a bigger part as shown in Table 8. Within the last month students arrived late on average 2.7 times. For this, 70 percent of students reported delays to public transport to be a main culprit, whilst nearly one third pointed out traffic jams as a main reason.

Table 8 –Reasons for arriving late

Reason	N	Percent
Too little time between Maglivi and downtown classes	1	0.24
Delays in public transport	168	41.08
Traffic jam	119	29.1
Woke up late	80	19.56
Other	41	10.02
Total	409	100

B. The Household Survey

92. The household data will enable us to track changes in travel time, travel costs and travel means for residents in the project affected area.

93. At the baseline almost 84 percent of households had a travel card for public transport and spent on average 12.6 GEL (7.6 USD) during the week preceding the survey, as shown in Table 9. Minibus consumed the largest part of that amount, at 7.8 GEL (4.7 USD). The average amount spent on transport including cost of petrol and taxi fares was 26 GEL (15.8 USD).

94. In total, 35 percent of households own a car and traveled on average 6742 km during the last year. But households which own cars also use public transport: 20 percent of these also have travel cards. Indeed, car owners spend nearly as much on public transport as do those without a car.

Table 9 – Household transport costs*

Transport	Obs	Mean*	Median*	Std. Dev.	Min	Max
Bus	286	4.8 (2.9)	3 (1.8)	5.9	0	30
Minibus	288	7.8 (4.7)	5 (3)	11.2	0	80
Metro	291	1.2 (0.7)	0	2.7	0	18
All public	**305**	**12.6 (7.6)**	**9 (5.4)**	**14.7**	**0**	**110**
Petrol	271	14.0 (8.5)	0	32.2	0	200
Taxi	295	3.0 (1.8)	0	8.8	0	70
All	**305**	**26.4 (16)**	**12 (7.2)**	**35.5**	**0**	**203**

* GEL with USD in brackets. Std.dev, min and max values in GEL.

95. On average households spent 5.5 hours traveling on public transportation during the last week, with a mean waiting time of an hour and 12 minutes (Table 10).

96. Just over half of the household members not in school and between the age of 16 and 67 are employed, this being true for 56 percent of men and 47 percent of women. The majority – 55 percent – of these people work in Saburtalo or Vake districts, and hence within a relatively short distance of their homes. To travel to work, less than five percent spend more than an hour. Nearly half spend 25 minutes or less. Currently, just over ten percent of those surveyed use the metro to travel to work.

Table 10 – Time spent on public transport last week

Hours spent	Obs	Mean	Median	Std. Dev.	Min	Max
Traveling by public transport	277	5.5	4	5.1	0	24
Waiting for public transport	277	1.2	1	1.5	0	10

97. When asked the main reasons for using the metro, 60 percent of respondents say they don't have a car, whilst cost is a reason for 25 percent, who state that it is cheaper than driving. Major reasons for not using, or not wanting to use, the metro include a dislike of crowds and infrequent departures.

Table 11– Reasons for using metro

Reason for using public transportation	Freq.	Percent
I would not use public transportation	1	0.33
I don't own a car	182	59.67
I don't like using my car in town	2	0.66
It is cheaper than driving	75	24.59
I can read/write/work/etc while travel	2	0.66
It is fast&efficient	14	4.59
No parking problems	7	2.3
I feel more secure	3	0.98
Less stress	4	1.31
Works well enough	7	2.3
I have no other choice	1	0.33
No answer	7	2.3
Total	305	100

Table 12 – Reasons for *not* using metro

Reason for using public transportation	Freq.	Percent
Departures not frequent enough	110	36.07
Don't like crowds	119	39.02
Too slow	11	3.61
Transfers/changes not efficient	8	2.62
Prefer comfort of own car	13	4.26
Metro/bus/minibus stop too far away	3	0.98
Often need to carry heavy loads	1	0.33
Often come back late	4	1.31
Too expensive	19	6.23
No reason	2	0.66
No answer	15	4.92
Total	305	100

C. The Business Survey

98. For businesses the aim was to be able to measure any changes in business activity and business revenue in the project-affected area as discussed in Section III.

99. A major challenge to reaching this objective however, was a large number of businesses refusing to participate in the survey, or, for those who did participate, a large proportion of non-responses to questions relating to revenue and costs. Indeed 178 – 38 percent – of the businesses visited refused to take part in the survey, 167 due to the 'content of the questionnaire' whilst the remaining 11 were due to absence of a manager or owner.

100. For those who did participate there was a 30 percent non-response rate to questions relating to revenue. To limit the sensitivity of these questions an option was introduced to respond not with exact numbers, but within a given range, which one quarter of those responding made use of.

101. According to the survey firm, ACT, non-response rates of up to 40 percent are common in Georgia and information relating to finances are regarded as being particularly sensitive. Unfortunately, this limits the prospects of obtaining an unbiased overview of business activity and development in the project affected area. However, the data may still provide an understanding of business development for those who did choose to both participate and respond to questions related to finance.

Revenue

102. As discussed, a number of business owners chose not to answer detailed questions about revenue and costs. Over 200 however, did provide either a number or a range for last year's income. For the 153 providing a number (Table 13a), the mean revenue was GEL 5983, with a median of 2000. For the 56 businesses providing a range (Table 13b), 41 percent were below 500 GEL, and another 15 between 500 and 2000 GEL. Provided that these businesses

report on revenue in the follow up survey we will get an indication of changes, though the sample is rather small for the purposes of ensuring statistical validity.

Table 13a – Revenue (range)

Range - GEL	Range - USD	Freq.	Percent
100 - 500	60-302	23	41.1
501-1000	303-605	7	12.5
1001-2000	605-1209	8	14.3
2001-3000	1210-1814	1	1.8
3001-4000	1814-2418	3	5.4
4001-5000	2419-3023	4	7.1
5001-6000	3024-3628	1	1.8
6001-10000	3628-6046	3	5.4
10001-20000	6047-12092	3	5.4
40001-50000	24185-30230	3	5.4
Total		**56**	**100**

Table 13b – Revenue (number)

Revenue	Obs	Mean	Median	Std. Dev.	Min	Max
GEL	153	5983.2	2000	13269.9	0.6	100000
USD		3617	1209		0.4	60460

Premises and clientele

103. Ninety-eight percent report that they serve clients at the premises and, hence, will be affected by any changes to the volume of people frequenting the area. Indeed, 65 percent use the premises exclusively for trade and 25 percent uses them exclusively for service provision.

104. Of 218 businesses that are renting the premises, 176 reported their monthly rent, ranging from 50-8100 GEL with a mean of 577 (349 USD), providing some scope for detecting changes to rental prices following the opening of the metro.

Table 14 – Monthly rent for businesses

Monthly rent	Obs	Mean	Median	Std. Dev.	Min	Max
GEL	176	576.8	300	936.7	50	8100
USD		348.7	181		30	4897

105. Close to half of businesses – 45 percent – are sole proprietorships with only one person working at the firm. Another 25 percent have two employees. Only seven percent have more than 10 people employed. Almost all employees are paid, as illustrated in Table 15.

Table 15 – Business employees

	Obs	Mean	Median	Std. Dev.	Min	Max
Employees	297	4.7	2	17.2	1	260
Of which paid	297	4.6	2	17.3	0	260

106. The vast majority of customers are local residents but university students, faculty and local workers are also reported to be frequent customers. During the last week, which most respondents reported as 'typical', the average number of customers was 198, with a median of 70 (Table 16).

Table 16 – Clients/customers last week

	Obs	Mean	Median	Std. Dev.	Min	Max
Number of clients	252	198.3	70	325	0	2100

107. Completing a new census of businesses following project completion, will allow us to detect any changes to the number of businesses operating in the area, number closed and opened up, as well as provide us with some overview of changes to rental prices, number of employees, customers and reported revenue for a subset of businesses – whilst keeping in mind that those choosing to report may well be statistically different from those who do report on this.

D. The Qualitative Study

108. The qualitative study consisted of Focus Group Discussions. Three groups were interviewed: Male residents in the project-affected area, women residents in the project affected area, and business owners. The discussions revolved around participants' transport use, opinions about various public transportation means and accessibility, and air pollution. For business owners the discussion also covered demographics of their clientele and expected effects of the metro on business activity.

109. Most participants were frequent users of public transportation and reported spending 2-3 hours daily on public transport. They most commonly use the bus, due to the cheaper fare, but all looked favorably upon the opening of the new metro station. For residents a main benefit of this will be late-night access to the area by public transport, currently not available. Additional expected benefits were a boost in apartment prices in the area, improved infrastructure and, for business owners, increased quantity of clients and competition.

V. CONCLUSION AND NEXT STEPS

110. The design of the evaluation and surveys described in this report took place between December 2011 and August of 2012. During this period a wide range of potential evaluation designs and approaches were examined and considered in order to ensure the highest possible level of rigor. The particular nature of the project and the data available meant that there were a number of limitations to the choice of methodology. Ultimately this led to the a design combining

a cohort difference-in-differences approach to measure impact on student expenditure, time use, attendance and GPA; and a single difference approach to obtain a less rigorous, yet useful, overview of developments for households and businesses. This was supplemented by qualitative Focus Group Discussions. The survey completed in June 2012 covered 900 students, 300 households and 300 businesses.

111. To complete the impact evaluation, follow-up surveys will be required. For households and businesses, the next survey should take place approximately six months after project completion – most likely in the fall of 2015.

112. As discussed in Section III.C, in the case of students the recently completed survey should be considered a pilot, or pre-baseline survey, with the actual baseline to take place in the fall of 2013. A number of follow-up surveys should be carried out subsequently: A baseline survey for students in the fall of 2013 with a follow-up survey of the same students to take place in the fall of 2014. At the same time a baseline should be completed for the second cohort, followed by an endline survey of the same group of students in the fall of 2015. An overview of all prospective surveys is provided in table 17.

Table 17 – Forthcoming surveys

	Year	Activity 1	Activity 2	Sample size
Students	Fall 2013	Baseline cohort1		TBD
Students	Fall 2014	Endline cohort1	Baseline cohort 2	TBD
Students	Fall 2015		Endline cohort 2	TBD
Households	Fall 2015	Qantitative endline		300
Businesses	Fall 2015	Qantitative endline		300
All	Fall 2015	Qualitative endline		TBD

113. All questionnaires used for this first round of surveys were designed based on the causal chains and expected outputs, outcomes and impacts described in this report. The same questionnaires should be used during the forthcoming surveys, which should target the same households and businesses in order to obtain a panel data set. The student surveys should follow the approach outlined above, and suggested estimation strategies are outlined in Section III.F allowing mid- and endline evaluators to complete the surveys as well as the analysis as smoothly and efficiently as possible.

Detailed Literature Review of Impact Evaluations of Infrastructure Projects

1. Impact evaluations of large infrastructure projects are still few in number, although several are ongoing. This is so because such evaluations pose specific methodological challenges, and it is particularly hard to make them "rigorous". One challenge is that infrastructure is most often placed in specific locations for political, economic or strategic reasons – thus leading to selection bias as described above. Another, which holds especially true for large-scale infrastructure, is that there is often only one road, bridge or railway being built, not many across different areas. And finally, particularly in the case of urban infrastructure projects, there are often large spillover effects throughout the city. These factors make it quite a challenge, and often impossible, to identify a suitable control group. However, some studies and impact evaluations of transport infrastructure projects have been conducted, as described below.

2. In the case of metro rail projects, most studies and discussions of impact take the form of traditional ex-ante Cost-Benefit Analyses, or the slightly more encompassing Multi-Criteria Analyses (Ustaoglu, 2009). In addition, a number of ex-post studies have been carried out, some of which look simply at changes in traffic-volumes (Vuk & Ildensborg-Hansen, 2006), whilst others look at the impact of new metro stations on land use, property prices, employment generation and population density. Vinha (2005) for example, examines the impact of the Washington DC metro on the distribution of employment and population using propensity score matching. She asks whether employment and residential construction increased more rapidly near metro stations than in other parts of the metropolitan area. She also examines the impact on the socio-demographic composition of population near metro stations and finds that there were significant impacts on employment and overall development density from proximity to a metro station. Her results also indicate that impacts on development are greater closer to the station than farther away and that they are greater the longer the stations have been in operation.

3. A number of studies cited in Estache (2010) have conducted before and after comparisons focusing on property prices as proxies for one impact of intra-urban train transit systems. A number of such studies, according to Vinha (2005), find positive effects from being near a transit station – although there are indications that the effect isn't necessarily linear as some find a negative effect on housing directly adjacent to the stations. In addition, some studies find that prices respond to the announcement of a new station, stressing the importance of timing when analyzing the impacts of a transit network improvement (e.g. McMillen & McDonald, 2004).

4. There is a slightly larger number of more rigorous impact evaluation studies of regular railways and roads. A few examine macroeconomic effects of railways. Banerjee, Duflo and Qian (2009) estimate the effect of access to transportation networks on regional demographic and economic outcomes across counties in China during 1986-2003. Applying an IV approach (using distance to a straight line between cities as an IV for distance to railway) they find that proximity to transportation networks has a large positive causal effect on per capita GDP growth rates across sectors. These benefits appear to reflect increases in aggregate production rather than displacement of productive firms to be near transportation networks. Donaldson (2010) uses data from colonial India to investigate the impact of India's railroad network on trade. Using a general equilibrium trade model, he finds that railroads decreased trade costs and interregional price gaps, increased interregional and international trade, and increased real income levels.

5. To our knowledge there are no existing studies which quantitatively examine the impact of metro or railways on community or household welfare and business activity. A few studies looking at household welfare have however been conducted for road projects. One such study –the first randomized impact evaluation of urban infrastructure placement – looks at the impact of a street pavement (road surface) project in the city of Acayucan, Mexico (Navarro-Gonzalez & Quintana-Domeque, 2010). The project randomly allocated 28 street pavements amongst 56 pre-approved streets and the evaluation surveyed communities and 1,231 households. The study found substantial increases in property values and land values as well as positive effects on credit and durable-goods consumption, which can be attributed to the project.

6. A few impact evaluations have been conducted for rural roads, although none were randomized. These generally indicate a positive impact on welfare through increased employment opportunities, local market development, and access to services. For example, Escobal & Ponce (2002) use propensity score matching to evaluate the impact of rural roads on key welfare indicators such as income or consumption for residents in some of the poorest districts of Peru. Results show that rehabilitated road accessibility can be related to changes in income sources, as the rehabilitated road enhances non-agricultural income opportunities, especially from wage-employment sources. The study also finds that income expansion has not been matched by an equivalent consumption increase; apparently because the additional income is allocated to savings, through increments in livestock, most likely they argue, because road quality improvement is perceived as transitory. Rand (2011) applies a matched double difference approach to analyze the employment generating impact of a tertiary road project in Nicaragua. He finds an employment gain attributable to the intervention of around 6.6 to 8.2 percentage points – all stemming from an increase in jobs generated locally.

7. In Georgia, Lokshin and Yemtsov (2003) examine the impact of rural roads. Based on data from the Rural Community Infrastructure Survey (RCIS) and the Survey of Georgian Households (SGHH) they use a difference-in-differences approach to find that road and bridge rehabilitation projects generate clear economic benefits at the community level. Whilst the impact of on labor market conditions is positive, it is insignificant. On the other hand, they find that the number of small and medium enterprises increased while the importance of barter trade fell as a result of the road projects. Moreover, at the household level, access to emergency medical assistance was found to improve unambiguously.

8. Finally a number of studies find direct poverty reducing effects. Mu and van de Walle (2011, 2007) combine matching with a difference-in-differences approach to examine the impacts of rural road rehabilitation on market development at the commune level in rural Vietnam. They find significant average impacts on the development of local markets and find evidence of considerable impact heterogeneity with a tendency for poorer communes to have higher impacts. In Bangladesh, Khandker et al (2009) use household panel data and apply a fixed effects analysis to examine the impacts of rural road-paving projects. They find that rural road investments led to higher secondary schooling enrollment and reduce poverty significantly through higher agricultural production, lower input and transportation costs, and higher agricultural output prices at local village markets.

Map of the Project Site[4]

[4] The project metro station is marked as ”State University” metro on the left handside of the map. TSU Maglivi is located on University Street at the bottom of the map.

References

ADB (2011) *A Review of Recent Developments in Impact Evaluation:* http://www.adb.org/documents/reports/developments-impact-evaluation/default.asp

ADB (2010) *Report and Recommendation of the President to the Board of Directors: Proposed Multitranche Financing Facility Georgia: Sustainable Urban Transport Investment Program.* Project Number 42414.

Baker, J. L. (2000) *Evaluating the Impact of Development Projects on Poverty A Handbook for Practitioners*, The World Bank, Washington D.C.

Duflo, E., Glennerster, R. & Kremer, M. (2008) 'Using Randomization in Development Economics Research: A Toolkit,', in Schultz, T. P. & Strauss, J. A. (eds*.) Handbook of Development Economics*, vol.
4, pp. 3895-3962, Elsevier (Also available as downloadable working paper).

Escobal, J. & Ponce, C. (2002) 'The benefits of rural roads: enhancing income opportunities for the rural poor', GRADE Working paper, 40-I.

Estache, A. (2010) 'A Survey of Impact Evaluations of Infrastructure Projects, Programs and Policies', ECARES working paper 2010-005, Université Libre de Bruxelles.

Heckman, J., Ichimura, H and Todd, P., 1997, *Matching as an econometric evaluation estimator: Evidence from evaluating a job training program*, Review of Economic Studies, 64, 605-654.

Joly, M., Lebo, E., Pipm, G. & Giely, M. (2012) *Survey Report: Feasibility Study Of the LRT Project in Tbilisi, Georgia*, Systra.

Khandker, S. R; Koolwal, G. B. & Samad H. A (2010) *Handbook on impact evaluation: quantitative methods and practices*, The World Bank, Washington D.C.

Khandker, S., Bakht, Z. & Koolwal, G. (2009) 'The Poverty Impact of Rural Roads: Evidence from Bangladesh', *Economic Development and Cultural Change* 57:4, 685-722.

Leeuw , F. & Vaessen, J. (2009). *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation,* The Network of Networks on Impact Evaluation.

Lokshin, M & Yemtsov, R. (2003) 'Evaluating the impact of infrastructure Rehabilitation projects on household welfare in rural Georgia', World Bank Policy Research Working Paper 3155.

McMillen, D.P. & McDonald, J.F. (2004). Reaction of house prices to a new rapid transit line: Chicago's Midway Line, *Real Estate Economics*, 32 (3), 463-486.

Mu, R. & van de Walle, D. (2011) 'Rural Roads and Local Market Development in Vietnam', *Journal of Development Studies*, 47:5, 709-734.

Mu, R. & van de Walle, D. (2007) 'Rural Roads and Local Market Development in Vietnam', Policy

Research Working Paper 4340, The World Bank.

Navarro-Gonzalez & Quintana-Domeque (2010): 'Urban Infrastructure and Economic Development: Experimental Evidence from Street Pavement', *IZA DP* No. 5346.

Ustaoglu, E. (2009) '*Economic Evaluation Of The Impacts Of New Rapid Rail Investments And Consequential Forms Of Urban Development Within The Greater Dublin Area*' Urban Institute Ireland-School of Geography, Planning, and Environmental Policy, University College Dublin.

Vinha, K. P. (2005) *The Impact of the Washington Metro on Development Patterns*. Dissertation submitted to the Faculty of the Graduate School of the University of Maryland, College Park in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Vuk, G & Ildensborg-Hansen , J. (2006) *Trafikale effekter af Metroen: en før- og efteranalyse*, Danmarks Transport Forskning, Lyngby.

White, H. (2011) 'Achieving high-quality impact evaluation design through
mixed methods: the case of infrastructure', *Journal of Development Effectiveness*, 3:1, 131-144.

White, H. (2009). *Theory-Based Impact Evaluation: Principles and Practice*. 3ie Working Paper No. 3.