

TECHNICAL COMMENT ON “BUILDING STATE CAPACITY: WHAT IS THE IMPACT OF DEVELOPMENT PROJECTS”, OCTOBER 2021

**Urban, Resilience and Land Global Practice
November 2022**

1. Overview

The Development Impact Evaluation Group of the World Bank (DIME) has published a study, *Building State Capacity: What is the Impact of Development Projects* (the DIME study), in October 2021 which states that it is an evaluation of the impact of the World Bank-supported Urban Local Development Support Program (ULGSP) project in Tanzania which was executed from 2013 to 2020. The evaluation claims to provide “evidence on whether institutional outcomes targeted by the capacity building project in Tanzania are different across local governments that received the project and those that did not” (p. 2) and, on the basis of this evidence, makes various statements about the project’s impacts, specifically that “the financing incentives provided by external partners did not add significant value beyond process of change already unfolding in the country” (p. 5), and “the project’s value-added to change in state capacity is estimated to be zero” (Abstract).

In fact, the DIME study does not constitute a valid impact evaluation of the ULGSP project intervention: most of the evidence it generates is not relevant to understanding the intended impacts of the project, and its claims and conclusions about these impacts do not have a valid empirical base. It follows that its inferences about projects of similar design in other countries are equally questionable.

There are three chief reasons for this: (i) the study focuses on and measures changes in a range of “governance outcomes” which the project was neither intended nor designed to target, ignores core objectives the project did in fact seek to achieve, and uses flawed measurement indicators; (ii) it employs a subjective method of measurement which is incapable of reliably capturing the actual, de facto impacts of the project; (iii) it measures too short an implementation period to draw valid inferences about the project’s impacts. The study’s findings are particularly questionable in the face of other recent evidence, using much more robust methodologies, which show directly contrary results i.e. that a performance grant operation with a similar incentive structure to that of the ULGSP was successful in enhancing local government performance in targeted versus non-targeted local governments (*Leveling Up: Impacts of Performance-Based Grants on Municipal Revenue Collection in Mozambique*, 2021).

The most that can be said of the DIME study is that it comprises an interesting description of changing perceptions of local officials and communities over part of the period during which the ULGSP project was being implemented. But as an impact evaluation – which is its central claim – it does not have value.

2. Scope and indicators measured

The ULGSP was not designed to address most of the “governance outcomes” on which the study focuses and which it erroneously claims the operation seeks to achieve. While the measurement of such governance factors might be useful for other purposes, it cannot validly be used to evaluate the impact of an initiative which did not target them in the first place. Put differently, there is a significant disconnect between the scope and nature of the factors the study measures and the objectives and design of the

operation: most of the indicators focus on factors which the ULGSP is not designed to address, and/or are not within the control of local governments (LGs) in Tanzania, and/or are so vague and indeterminate that they cannot convey any useful information about the impact of the operation.

To list just a few examples:

The Government Official survey examines:

- Timely disbursements to LGs - which are not controlled by LGs or directly targeted by the operation;
- Delays in budget guidelines provided by central government – these are also not controlled by LGs or targeted by ULGSP;
- Whether the budget is executed in accordance with “good results” – which is vague and unmeasurable: what is a “good result”?
- Five items relating to managements’ efforts to attract, retain and promote staff – in Tanzania local government management has almost no control over the factors (such as emoluments, appointments and transfers) which might affect this (for all but junior staff these are central government powers)
- Whether officials personally handled grievances – the “personal handling of grievances” by this or that government official is not a useful indicator of the functionality of a grievance-handling mechanism and positive or negative trends in “personal handling” do not necessarily indicate anything about the performance of that system.

The Citizen survey examines:

- Whether they have seen “problems with local governments”, which is extremely vague – the ULGSP does not/cannot address all “problems with local governments”: moreover, because this is so broad, how can one say whether any reported increases or decreases in problems are in areas which the operation targets?
- Whether councilors are corrupt – which ULGSP does not address;
- Whether policeman are paid bribes – LGs do not employ or control police officers and the ULGSP does not address this;
- Whether health standards are maintained – ULGSP does not address health and significant aspects of health delivery are controlled by central government;
- Whether there is easy access to primary school placements – ULGSP does not address education
- Perceived quality of land management – which LGs do not control.

Some 26 of the 44 indicators in waves I and II of the Citizen survey and some 25 of the 54 indicators in waves I and II of the Government Official survey are either not targeted by ULGSP, and/or are not controlled by LGs in Tanzania, and/or are so broad and vague that they cannot provide useful information about the institutional performance factors that ULGSP was designed to impact or material changes in such performance. Nine out of 10 of the wave III indicators of the Citizen survey fall into this category.

The disconnect between the indicators that the DIME study measured and the factors that the ULGSP targeted is well demonstrated with reference to property tax. At the ULGSP’s inception in 2013, property tax administration was a local government responsibility and a lot of emphasis was placed in the design of the incentive grant - hence in the Annual Performance Assessments through which LG performance is objectively measured (on the basis of verifiable factual data) - on improvement in property

tax performance. However, in 2016 property tax was recentralized away from local government to national government and thereafter was no longer a municipal responsibility. As a result of this, property tax administration became irrelevant to the institutional performance or strengthening of local governments in Tanzania and obviously had to be dropped from the ULGSP APA. Notwithstanding, four measures in the DIME waves II survey focused on property tax for both the 2016 and 2018 rounds and were retained as indicators of LG institutional performance. Moreover, the survey questions ask the respondent (the government official) whether they, themselves, pay property tax on various forms of real estate asset. How a measurement of the personal payment of a national tax by a local government employee is meant to be an indication of the quality of a local government's tax administration systems is unclear. Put more directly, these indicators are simply not relevant to what the ULGSP was trying to achieve and cannot measure anything useful about the impact of the project. A careful study of both the Official and Household surveys across all three waves in the context of the functional responsibilities and powers of local governments in Tanzania and the actual – and necessarily limited – objectives and design of ULGSP reveals a great many such examples.

In addition, key areas of institutional strengthening which the ULGSP did, in fact, target are omitted. For example, the ULGSP placed substantial emphasis on improving procurement systems, financial management systems (much more widely than internal audit systems) and the execution of infrastructure funding. These are all ignored in the DIME study.

Finally, many of the indicators which arguably have some relevance to the ULGSP are only diffusely or superficially related to project impacts. For example, the Officials survey contains a cluster of measures of "Staff Capacity" (p. 16). These include perceptions on matters such as the share of staff who can create PPT presentations or write a memo. The surveys containing such questions are justified on the basis that they provide "broader measures of governance quality" (p. 8). Putting aside the obvious issues surrounding the enormous latitude for subjectivity and the resultant problems with response comparability intrinsic to this type of question (any PPT? Four slides of basic text or 25 slides with animation? Any memo? Three lines requisitioning stationery or 10 pages motivating a change in a municipal by-law?), it is difficult to understand what aspect of "governance quality" or state capacity such indicators can effectively measure. Certainly, they are greatly inferior to measures such as the effective functioning of the local government procurement system which was objectively measured through the APA but which the DIME survey omitted.

Evaluating a project intervention on the basis of things that the intervention was not designed to address and has no means of addressing violates basic and widely-recognized principles of impact evaluation. The DIME study does indeed measure some things but those things have little to do with the project or the impacts the project sought to achieve.

3. Survey methodology

The DIME study relies almost entirely on an opinion survey methodology which captures changes in the perceptions and beliefs of local government officials and households on a range of indicators in two groups of local governments – the ULGSP target group and a comparator group (of the 108 indicators in waves I, II and III of the survey it appears that in only two cases was any objective, factual data gathered). Unlike the project Annual Performance Assessment it thus contains no objective or verifiable information on the actual changes in the local government institutional performance that formed the core focus of ULGSP. The most that can be said of the DIME exercise is that it establishes some findings about trends in the perceptions (whether "self-reported" or entirely subjective) that government officials and citizens

have of the indicators measured: whether or not these perceptions have any relationship to what actually happened is not established. In other words, the study does not comprise an assessment of the *de facto* relative impacts of the ULGSP in respect of the specific and well-defined institutional performance objectives which the operation was designed to achieve. It is purely a measurement of the subjective beliefs of the officials and citizens of a wide range of “governance” variables, many of which have little or nothing to do with the operation. It is thus not possible to know how actual conditions changed in the ULGSP group of local governments relative to the comparator group. Without this knowledge it is not possible to say anything about the relative impact of the operation, which is the basic “difference-in-differences” peg on which the study wishes to hang its hat.

Interestingly, the DIME study concedes that “One obvious way to measure (the) outcomes in non-project local governments would be to engage the same firms that are scoring the project local governments on the APA to undertake the same process of data gathering and scoring for the non-project LGAs.” (p. 8-9). This would indeed have provided a robust means of measuring the relative changes in the target vs the comparator group. To justify this decision, the report offers two reasons: the capacity constraints of the firms undertaking the APA and the costs. (“However, due to the intensity of data gathering required for the APAs under the project, the same auditing firm could not administer APAs in comparison LGAs under the same timeline as project LGAs. Given both this and the additional cost, the evaluation draws on the survey data...” p. 9).

This holds no water. First, the annual cost of the APAs was approximately \$388,700. The total cost of the two survey rounds of the DIME study was \$823,288 giving an annual average of \$411,644. Particularly given scale economies, it would have thus been cheaper – certainly no more costly – to extend the APAs to the comparator group than to undertake the opinion surveys. Second, the APAs were undertaken by UPIMAC, a well-established high-capacity firm which has undertaken multiple such exercises in countries such as Uganda, Ethiopia and Tanzania – in some cases covering many more local governments than the ULGSP covered in Tanzania. There is no indication, and it strains credulity, to suggest that it could not extend the ULGSP APAs to include the comparator group – in fact, it would have been much easier to undertake this exercise than the one employed by the DIME study.

There is also the question of respondent bias to consider. Of the 11 officials surveyed in each council, seven were civil servants. Most of these officials are appointed by central government, chiefly the agency (PO-RALG) which commissioned the survey - which agency also is responsible for the posting (transfer) and promotion decisions affecting these officers. The report states that assurances or anonymity were given, but it is not clear how these could have been guaranteed or why any such assurances would have been believed, particularly in an environment such as that in Tanzania which is characterized by heavy central control of the local civil service through both state and party mechanisms. Under such circumstances, serious questions arise regarding the likely candor and veracity of the survey responses (the survey was conducted through in-person interviews). It is hardly surprising that the movement in reported perceptions amongst Government officials in both ULGSP and comparator LG groups was broadly and uniformly positive – but this is quite as likely to reflect the incentives of the responding officials as any actual, *de facto*, changes in material condition. The paper contains a sentence mentioning that this reality may create “some bias”, then immediately dismisses this problem because “this potential bias would affect both treatment and comparison groups”. But this is beside the point, which is rather that because a purely perception-based survey method does not reliably or verifiably capture the factual position regarding how local governments are actually performing on matters such as procurement, financial management, and so on, it matters not whether the indicators move in tandem or

apart across targeted and control groups - the information it generates is of little value in understanding actual impacts in these areas.

The choice of the opinion survey methodology remains unexplained. Whatever the reason, the basic point is that this methodology did not provide a reliable way of measuring actual, de facto changes in the target vs control groups of local governments and was incapable of generating data which can be relied on to evaluate the actual impact of the project.

3. Evaluation period

The period covered by the study is simply too short to allow for a valid impact evaluation. The ULGSP became effective in February 2013 and closed in December 2020, giving it an effective life of almost eight years. The assessment period covers the time from the baseline to the endline surveys i.e. between February 2016 and April 2018, a period over which \$117m, or 49 per cent of the ULGSP performance grant funding was disbursed to local governments. It is not clear how a study which covers a two-year slice of an almost eight-year project and less than half of its funding can be used to properly understand its overall impacts, particularly given that such (institutional) impacts typically accumulate fairly slowly and incrementally over time – one of the main reasons that the operation had an above-average timeframe.

The DIME study contains a fairly extended discussion about changes in the early stages of the project (before the study period) and the evolution of APA scores over time across the project period (Appendix B). This discussion is necessarily speculative and contains no clear conclusions. Its purpose appears to be to downplay the difficulties that such a limited examination of the project creates for generating findings about project impacts. In this it fails. The DIME study repeatedly emphasizes that it relies on a “difference-in-differences” methodology and the fact remains that it only generates difference-in-differences data for just over a quarter of the project’s life. It is not possible to come to reliable conclusions about overall project impacts with such a limited examination.

4. Conclusion

Given its mistaken focus, flawed methodologies and limited measurement timeframe, the DIME study does not comprise a valid evaluation of the impact of the ULGSP on its intended LG institutional performance strengthening objectives and cannot be understood as such. Its inferences and conclusions about the impact of the ULGSP project in Tanzania – and other similar projects in other countries – do not have a valid basis and cannot be sustained.