

GPOBA EVALUATION NOTES

NOTE N. 1

APRIL, 2012

SHOULD WE RE-EVALUATE WHAT TO EVALUATE (AND HOW)?

LESSONS LEARNED FROM THE GPOBA PILOT PROJECT “KAMPALA WATER CONNECTIONS FOR THE POOR” AND METHODOLOGICAL NOTES ON THE EVALUATION OF OBA PILOTS

By Luisa M. Mimmi

Keywords: Output-Based Aid (OBA), evaluation, water access, slums

Luisa Mimmi is a consultant at the World Bank, 1818 H Street NW, Washington, DC 20433, USA.
Telephone: +1-202-458-8224. E-mail address: luisamimmi@gmail.com

1 Introduction

The Global Partnership on Output-Based Aid (GPOBA)¹ is implementing a small portfolio of impact evaluations (IEs) and qualitative assessments of pilot projects to verify whether Output-Based Aid (OBA) has achieved its declared goal of “increasing access to basic services—like infrastructure, health-care, and education—for the poor in developing countries”. In addition, the ongoing IEs attempt to investigate any differences in performance between the OBA approach and other subsidy design choices, consistent with the OBA principle of making aid more targeted and transparent by linking it to the delivery of specific “outputs.”

The existing portfolio of studies spans across various sectors and countries in which the OBA approach has been piloted, namely improved access to water or electricity (Uganda and India), solar home systems for off-grid areas (Bangladesh), and maternal health vouchers (Uganda).

As the first IEs start producing results, it seems timely to reflect on some lessons suggested by the initial empirical evidence collected. One study in particular, focused on an OBA water pilot in Kampala, Uganda, illustrates what a challenging endeavor impact evaluation can be. Its preliminary findings point our attention to the arduous task of estimating the nature and magnitude of the effects of an innovative water provision scheme, especially in the context of urban slums, when an array of expected and unexpected shocks occur. Drawing on this pioneer case-study, some questions arise and some ideas can be offered to inform future evaluations of OBA pilots.

2 The Case Study

In 2008, GPOBA and the public utility, National Water and Sewerage Corporation (NWSC), initiated a four-year, US\$4 million cofunded project as part of a larger pro-poor policy adopted by NWSC since 2004 to comply with the Government of Uganda’s Poverty Eradication Action Plan (PEAP). This output-based aid (OBA) pilot was intended to provide access to piped water to 400,000 of Kampala’s urban poor through the installation of 19,000 new yard taps and 800 public standposts, called public water points (PWP), half of which were traditional and half of which used prepaid metering. At the outset, the agreement favored yard-tap connections (each serving three families) in areas where land ownership was clearly established. The PWP (each serving 30 families) would be installed in poorer areas where land tenure was uncertain. Half (400) of these PWP were to be randomly selected to use prepaid meters, an innovation that arguably increases quality and affordability for the users, and can reduce the utility’s share of nonrevenue water.²

The objective of the project was to accelerate, through a geographically targeted, performance-based financial incentive, NWSC’s efforts to provide clean-water connections to Kampala’s growing urban and peri-urban poor residents – a problematic customer segment due to weak property rights, high residential mobility, and other agency problems (e.g. erratic willingness of land owners

¹ The Global Partnership on Output-Based Aid (GPOBA) is a partnership established in 2003 by the UK (DFID) and the World Bank. The program’s other donors are the IFC, the Netherlands (DGIS), Australia (AusAID), and Sweden (Sida). See <http://www.gpoba.org/> for detailed information about GPOBA and the OBA approach.

² The prepaid meter innovation was interesting for its implicit control on prices. In fact, the prices faced by households obtaining water from traditional public water points can vary for a number of reasons. First, public water points are managed by a member of the community who has power to price discriminate across households. Second, since public water points are local public goods, long outages may reflect collective action problems. The outages may be caused by the public water point manager, who might fail to pay the water utility company or report service interruptions.

to allow tap installations, risk of pricing abuse on the part of the public standposts' operators, or unclear division of responsibilities in operation and maintenance of the equipment).

3 Improving Access to Water for the Urban Poor

Extensive literature has discussed the positive effects of access to clean water on health and general well-being (Galiani et al. (2005), Kremer et. al. (2010), Fewtrell et. al. (2005)). Recent evidence shows that even when better access to clean water does not improve health, there are still considerable gains in time and possibly a diminishing of social conflict (Devoto et. al. 2011). "Improved" water access, such as household connections or public standpipes, offer various degrees of effectiveness depending on certain accompanying circumstances (e.g., sanitation infrastructure), behavioral change (adequate hygiene and education levels), and the socioeconomic environment, which can determine the successful adoption and sustainability of interventions. For GPOBA, the relevant question was: How is the provision of clean water best offered?

More specifically: Which is the best service-delivery option in the context of slums and peri-urban poverty: shared public standposts or semi-individual yard-tap connections? The OBA Kampala pilot assumed that relatively better-off beneficiaries with stable land tenure would self-select for the semi-individual yard-tap connection. Despite an additional service charged by NWSC, the yard-tap connection allowed to skip the arbitrary price determination from PWP operators who exercise de facto monopoly power and may even discriminate across households³. The yard-tap connection is also easier to manage because of the immediate relations between the tap owners and their close neighbors.

However, at the time of the project's midterm review in 2010, the rolling-out of yard-tap connections had been very slow.⁴ The densely populated target area, with its mobile housing patterns and lack of land ownership, was not conducive to the more expensive yard-tap connections. Thus, the targeted number of households for yard-tap connections was lowered to 6,000 and the target for PWPs increased to 200 conventional and 2,333 prepaid, while maintaining the same subsidy grant.⁵

The midline results of the impact evaluation found a high rate of sample attrition between the baseline (2008) and follow-up (2011) surveys: only about 55 percent of the original 1,900 households sampled at baseline were found during the follow-up (the sample was evenly split between home owners and renters). This high mobility was driven by both household characteristics (with poorer and younger households more likely to attrite) and location. Attrition was systematically related to poor neighborhood sanitary conditions, though not to access to improved water supply, or rental status.

An important policy lesson must be acknowledged: Contrary to the project design assumptions, in the targeted poor urban settlements, shared connections to improved water appear to be the optimal choice, despite some risks of abuse in the community management arrangement. In the

³ NWSC sources reported that, as of 2011, whereas the NWSC subsidized tariff was set at 17 USh per jerrican, the kiosk operators (usually the person who donated the land for the PWP installation or his/her agent) charged 50-100 USh per jerrican, depending on availability, seasonal variations, and presence of alternative water sources.

⁴ Back-to- Office Report (BTOR) World Bank/GPOBA Mission to Uganda, August 30 – September 3, 2010.

⁵ Office memorandum, "Kampala Water Connections for the Poor Project (Grant No. TF091510): Amendment of GPOBA-funded Grant Agreement," November 9, 2010.

context of high poverty, high mobility, and poor sanitary conditions, people refrained from making a higher investment (even subsidized) in the yard-tap, opting for the lower commitment of a shared water kiosk. As for the impact of prepaid meters on reducing the “middle-man” mark-up on prices and improving service reliability, this aspect could not be observed by the midline survey because too few prepaid PWP were functioning at the time.

4 The Challenge of Evaluating an Experimental Pilot Project

This case clearly illustrates that evaluating an infrastructure pilot project in a difficult context is far from the ideal “laboratory-like” conditions for a rigorous impact evaluation, in which the working assumptions hold and the project smoothly follow the plans. Unfortunately, the standard quasi-experimental design adopted for the evaluation (matched difference-in-difference estimation with partial randomization of the type of intervention)⁶ had limited tolerance for unexpected “shocks.” Besides the above-mentioned revision in target outputs, the pilot project suffered a delay in procuring and distributing the tokens necessary to operate the prepaid PWP. In addition, some contextual water rationing, service discontinuity⁷, and contamination issues⁸ occurred during implementation that may have undermined the functionality of some of the newly built access points.

A resulting delay in program implementation during the three-year interval between surveys meant that the degree of beneficiary exposure to the OBA intervention was insufficient to observe precise program effects. Verified installed water points at midline were only 10–20 percent instead of the planned 50–60 percent (see Figure 1). Thus, estimates of the project’s impacts on the beneficiaries’ well-being at this survey round were inconclusive.⁹ Furthermore, the anticipated comparison of traditional and prepaid-meter PWP (in terms of service quality and affordability) could not be observed. This comparison and other aspects of the evaluation can still be carried out through an additional survey, but the high attrition of participants is an obvious drawback to waiting another year or two to collect more information.

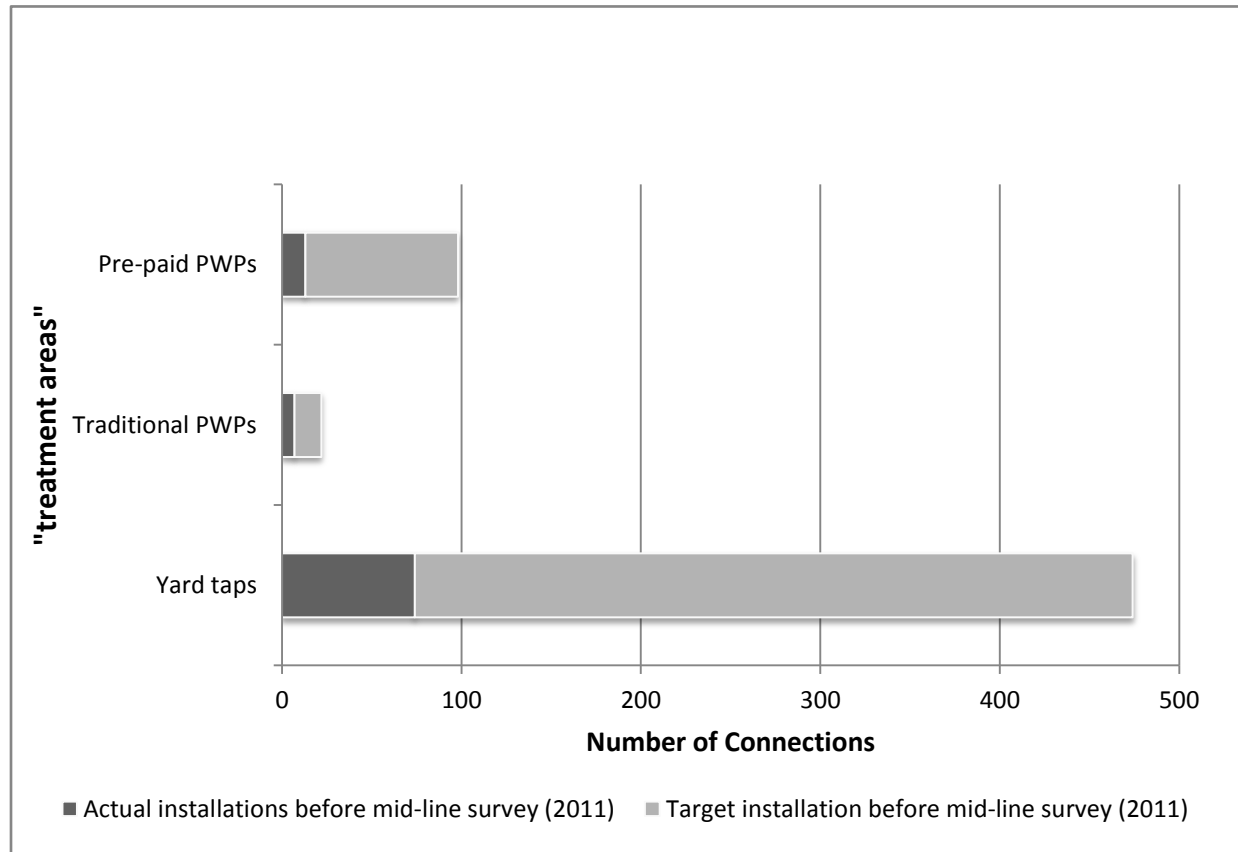
⁶ Although randomized controlled trials are the “gold standard” for impact evaluation design, many infrastructure interventions are not suitable for randomization because the technical nature of the intervention often prevents it due to engineering design or utility regulation constraints that determine the areas selected for installation of the infrastructure.

⁷ “Field surveys undertaken shortly after completion (Kampala City Council with support from the French Development Agency (AFD) provided safe water through public taps and protected springs between 2002 and 2006, constructing 307 PWP) showed that a number of kiosks and PWP had become dysfunctional due to poor operation and maintenance by their operators. The main reason was found to be the backlog of huge arrears accumulating from non-payment of PWP/kiosk operators to NWSC. This meant that NWSC had to disconnect those non-paying PWP and kiosks (in total between 5% and 10% of these were disconnected). In addition field surveys undertaken in 2011 showed that the so-called protected catchment upstream from the spring tap had been constructed upon – sometimes with latrines, rendering these so-called protected springs unsafe” (AFTUW, 2012, p. 27).

⁸ At the time, OOSKANews Inc.(2008, p. 4-5) reported: “Ill-planned urbanization and changing rainfall patterns are being blamed, and experts have warned that residents run the risk of consuming contaminated water from the piped water supply. Piped water in Kampala comes mainly from the state-run National Water and Sewerage Corporation and from natural springs found in many parts in and around the city. Both sources are being contaminated, and there is a looming problem of fecal matter contaminating water supplies.”

⁹ Initial evidence suggests a statistically significant reduction of distance and time allocated for fetching water, although little variation in quantity and expenditure. The health effects look fuzzy and mixed, and no mayor effect is found in time devoted to school or productive activities for adults (Habyarimana 2012).

Figure 1 Planned vs. Actual Water Points Installed between Baseline and Midline Surveys, 2008–11



5 A Word of Warning to Impact Evaluators

This example poses an important methodological question for impact evaluators of development interventions, and specifically for OBA practitioners. In principle, it is widely accepted that impact evaluations (IEs) should be conducted on pilot projects with an innovative and/or controversial feature (larger, more mature projects should have already been tested). Nonetheless, the experimental nature of these pilots often means they are prone to delays or design failures and consequent restructuring, whereas an IE requires a stable project design and certainty of assumptions. Therefore, evaluations face serious challenges if the project scope changes or there are severe delays, especially when the evaluation strategy is based on differential timing of construction and statistical matching of the planned “treatment” and “control” areas.

We could easily conclude that, with the experimental nature of GPOBA’s pilots, an appropriate degree of flexibility should be built into the design of future IEs. But how would flexibility work in practice? One option is to increase the interval between the baseline and follow-up survey to allow for changes in the program, or to multiply the rounds of surveys (David McKenzie recently published an articulate discussion on the determination of the optimal number of follow-up data collections).¹⁰

¹⁰ See David McKenzie (2011).

If time is a limiting factor, another option is abandoning the popular paradigm of baseline plus follow-up surveys, and opt instead for ex-post assessment(s) only. Less robust but equally popular impact evaluation designs use cross-sectional surveys (data collection is conducted once, possibly on a large population) and techniques, such as matching observable characteristics, to extrapolate program effects on the participants. For instance, a GPOBA-funded evaluation on its Bangladesh Solar Home System (SHS) program is using an ex-post design, an option made viable thanks to the massive amount of nationwide data collected by the local implementing agency, Infrastructure Development Company Limited (IDCOL). Nevertheless, estimating impacts from a baseline remains preferable, at least in principle, in the case of one-off interventions such as subsidized connections to water, which are expected to have multiple and heterogeneous effects that cannot simply be captured through comparisons or through an ex-post subjective recall exercise.

In other words, although the pre-post comparison of “treatment” and “control” areas remains the best method to capture size and significance of the program effects, evaluators must appreciate the risks of evaluating innovative infrastructure schemes, especially where project implementation issues affect the data collected by surveys.

A strategic rethinking could be explored for future evaluation of complex infrastructure pilot projects. In fact, there are simplified, rapid ways to assess project impacts that provide meaningful insight while minimizing data collection requirements.¹¹ The key is to narrow the scope on a case-by-case basis to the vital dimensions of interest in each application of the OBA concept, which could possibly shift the emphasis to more qualitative aspects of evaluation. The case study suggests a few specific questions that are still relatively unexplored, and could be studied with short and cost-effective evaluation designs.

First, while the vast majority of IEs on development programs focus on measurable benefits to the targeted beneficiaries, it could be argued that, if the goal is to justify the adoption of an OBA mechanism, the critical stakeholders affected are the implementing agencies and the suppliers/utilities. There is still very little evidence that verifies if and to what extent carefully sized result-based subsidies actually accelerate their involvement in the promotion of universal coverage of basic services in a pro-poor and equitable way.

Second, like most public policy interventions, OBA pilots are voluntary rather than coercive (legally required). They require participation and, therefore, behavior change from the involved agents (e.g. customers or utilities). Considering these dynamics, it is clear that intangible political economy and institutional considerations may heavily affect the project’s effectiveness. This influence is why an increasing number of evaluation experts is calling for a more nuanced approach to evaluation, such as the behavioral design approach¹², to combine quantitative and qualitative data to map the reasons that could drive compliance to assess likely levels of program participation and whether and how the intervention may be replicated in other settings.¹³

A third idea for rethinking the evaluation design is that when dealing with an experimental approach (like OBA) and its application in challenging contexts, it may be worth shifting attention from the prevailing evaluation of the project’s final impacts to ex-ante evaluations at project

¹¹ For a comprehensive discussion, see chapter 7 of Valadez and Bamberger (1994).

¹² See Datta and Mullainathan (2012).

¹³ In a recent article, Howard White advocates more widespread application of a theory-based approach (which is necessarily a mixed-methods evaluation) to infrastructure projects’ impact evaluation as a means to go beyond just reporting a counterfactual-based measure of the impact on selected beneficiaries and delve deeper into a contextual analysis of the intervention, including the reasons and channels through which the impact is expected to occur. See White (2011).

preparation phase. Investing in evaluation efforts at this stage (or merely getting more closely involved with preprogram feasibility studies by the project teams) could help define the optimal program design and thus reduce the high costs of implementing projects that are later found to be ineffective or delayed due to erroneous assumptions or poor policy performance. Again, the emphasis would remain on theory-based models and on close collaboration with local counterparts for in-depth understanding of the context.

In conclusion, evaluating the OBA approach is just as complex as adopting it, and subject to a trial-and-error cycle. Thus, it is essential to further the discussion of strategic goals and explore new possibilities to fine tune GPOBA's future evaluation strategy in a creative and flexible way.

Acknowledgements

The Evaluation Notes series is intended to summarize lessons learned from evaluations and key policy findings on OBA-related topics. This note builds on an evaluation conducted on the GPOBA pilot project "Kampala Water Connections for the Poor" (evaluation reports are available from the author upon request). The evaluation was undertaken during 2008–11 with financial support from the Global Partnership on Output-Based Aid (GPOBA) and the Spanish Impact Evaluation Trust Fund (SIEF). The views expressed in the note are entirely those of the author and should not be construed to imply endorsement by GPOBA, the World Bank, SIEF or any other project stakeholder.

Bibliography

- AFTUW (2012). *Uganda, Impact of Social Connections on the Poor in Kampala - for external review*. Washington: The World Bank.
- Datta, S., & Mullainathan, S. (2012). *Behavioral Design A New Approach to Development Policy*. Washington DC: Center for Global Development.
- Devoto, F., Duflo, E., Dupas, P., Pons, V. and P. William (April 1, 2011). Happiness on Tap: Piped Water Adoption in Urban Morocco. *MIT Department of Economics Working Paper No. 11-05*.
- Fewtrell, L., Kaufmann, R. B., Kay, D., Enanoria, W., Haller, L., and J. M. Colford (2005). Water, Sanitation, and Hygiene Interventions to Reduce Diarrhoea in Less Developed Countries: A Systematic Review and Meta-Analysis. *Lancet Infectious Diseases* 5 (1): 42–52.
- Galiani, S., Gertler, P. and E. Schargrotsky (2005). Water for Life: The Impact of the Privatization of Water Services on Child Mortality. *Journal of Political Economy* 113 (1): 83–120.
- Habyarimana, J. (2009). *GPOBA-NWSC Project - Baseline Survey Report*. Working Paper.
- Habyarimana, J. (2012). *Connecting the Urban Poor: Mid-term Report on the Impact and Management of Piped Water Services from Uganda*. Working Paper.
- Kremer, M., J. Leino, E. Miguel, and A. Zwane (2010). Spring Cleaning: Results from a Randomized Impact Evaluation of a Source Water Quality Improvement. Forthcoming *Quarterly Journal of Economics*
- McKenzie, D. (2011). Beyond Baseline and Follow-up: The Case for More t in Experiments. *Policy Research Working Paper*.
- OOSKANews Inc. (2008). Kampala Facing Serious Water Problems. *Weekly Water Report* 3 (no.31), pp. 4-5.
- Valadez, J. and M. Bamberger (1994). *Monitoring and Evaluating Social Programs in Developing Countries - A Guide for Policymakers, Planners, and Managers*. Washington, DC: The World Bank.
- White, H. (2011). Achieving High-Quality Impact Evaluation Design through Mixed Methods: The Case of Infrastructure. *Journal of Development Effectiveness*, 3(1), 131–144.