

Learning from the Experiments That Never Happened

Lessons from Trying to Conduct Randomized Evaluations
of Matching Grant Programs in Africa

Francisco Campos

Aidan Coville

Ana M. Fernandes

Markus Goldstein

David McKenzie

The World Bank
Development Research Group
Finance and Private Sector Development Team
December 2012



Abstract

Matching grants are one of the most common policy instruments used by developing country governments to try to foster technological upgrading, innovation, exports, use of business development services and other activities leading to firm growth. However, since they involve subsidizing firms, the risk is that they could crowd out private investment, subsidizing activities that firms were planning to undertake anyway, or lead to pure private gains, rather than generating the public gains that justify government intervention. As a result, rigorous evaluation of the effects of such programs is important. The authors attempted to implement randomized experiments to evaluate the impact of seven matching grant programs offered in six African countries, but in each case were unable to complete an experimental evaluation. One critique of randomized experiments is publication bias,

whereby only those experiments with “interesting” results get published. The hope is to mitigate this bias by learning from the experiments that never happened. This paper describes the three main proximate reasons for lack of implementation: continued project delays, politicians not willing to allow random assignment, and low program take-up; and then delves into the underlying causes of these occurring. Political economy, overly stringent eligibility criteria that do not take account of where value-added may be highest, a lack of attention to detail in “last mile” issues, incentives facing project implementation staff, and the way impact evaluations are funded, and all help explain the failure of randomization. Lessons are drawn from these experiences for both the implementation and the possible evaluation of future projects.

This paper is a product of the Finance and Private Sector Development Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at fcampos@worldbank.org, acoville@worldbank.org, afernandes@worldbank.org, mgoldstein@worldbank.org, and dmckenzie@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

**Learning from the experiments that never happened:
Lessons from trying to conduct randomized evaluations of matching grant programs in Africa[#]**

Francisco Campos, *World Bank*

Aidan Coville, *World Bank*

Ana M. Fernandes, *World Bank*

Markus Goldstein, *World Bank*

David McKenzie, *World Bank, BREAD, CEPR and IZA*

Keywords: Matching grants; Impact Evaluation; Randomization bias; Learning from failure.

JEL codes: O14, H25, D22, C93

[#] We are thankful for the support of DFID and the World Bank's Knowledge for Change Trust Fund for supporting our efforts to start impact evaluations in this area, as well as the World Bank operational staff and Government counterparts who helped and supported our attempts. We thank Alvaro González, Smita Kuriakose and Tom Wagstaff for comments. All views expressed in this paper should be considered those of the authors alone, and do not necessarily represent those of the World Bank.

1. Introduction

A typical matching grant consists of a partial subsidy - most commonly covering 50 percent of the cost, but ranging as high as 90 percent - provided by a government program to a private sector firm to help finance the costs of activities to promote exports, innovation, technological upgrading, the use of business development services, and, more broadly, firm growth. Matching grant programs are one of the most common policy tools used by developing country governments to actively facilitate micro, small, and medium enterprise competitiveness, and have been included in more than 60 World Bank projects totaling over US\$1.2 billion, funding over 100,000 micro, small and medium enterprises.¹ Add in funding provided by other development agencies and national governments, and it seems likely that at least two billion dollars has been spent on these projects over the last twenty years.

Yet despite all the resources spent on these projects, there is currently very little credible evidence as to whether or not these grants spur firms to undertake innovative activities that they otherwise would not have done, or merely subsidize firms for actions they would take anyway. From a social return perspective, the rationale for developing matching grant programs is usually also based on the assumption that there are positive externalities to workers, other firms, and to the country as a whole from having firms undertake these activities – workers will receive jobs and can use their upgraded skills in other parts of the economy, additional firms will learn from firms participating in the program, the market for business development services will be developed, the government will receive additional tax revenues, and society will benefit from broader economic growth. There is even less evidence to support these assumptions.

Several case study and non-experimental evaluations have attempted to provide some evidence on the impacts of matching grant programs (e.g. Biggs, 1999; Phillips, 2002; Castillo et al., 2011; Crespi et al., 2011; Gourdon et al., 2011, Lopez-Acevedo and Tan, 2011). However, these programs typically cater only to a tiny fraction of the firms in a country, and the firms that self-select or are selected for these programs are likely to differ in a host of both observable and unobservable ways from firms that do not receive the funding. This is likely to lead to an upward bias in non-experimental evaluations if more entrepreneurial firms with positive productivity shocks are the ones seeking out the program, and a negative bias if it is better politically connected but less productive firms that receive the funding.

¹ Data from a World Bank Latin American and the Caribbean overview available at <http://go.worldbank.org/OVDGTHSWY0>.

Randomized experiments do not suffer from selection bias and offer the potential to provide more credible estimates of the impacts of these programs. Moreover, matching grants would appear *ex ante* to be one of the types of private sector development activities most amenable to experimental evaluation.

We therefore set out to design randomized experiments to prospectively evaluate seven matching grant programs in six African countries. Five were to be supported through World Bank loans and technical support, while two stemmed from a direct engagement with the government focused on increasing the use of impact evaluation (IE) to evaluate national programs. Africa is the region where matching grants have been used the most, accounting for slightly more than half of all such projects supported by the World Bank. Thus, conducting evaluations of projects in this region offered the potential to both understand the impacts of the existing projects, as well as inform future work in the region. However, for a variety of different reasons, none of these experiments ended up being completed.²

The continued debate about the role of randomized experiments in development has included the discussion of there being selection into which projects are able to be implemented as experiments (e.g., Ravallion, 2009) and with which partners (e.g., Allcott and Mullainathan, 2011), as well as the need for a trial registry in ensuring that all studies undertaken end up being reported on (e.g., Rasmussen et al., 2011). Most of the attempted studies here would not have even made it to the trial registry stage, but we think there are still important lessons to be learned from discussing the attempted evaluations, their reasons for failure, and the insights for future work that can be drawn from this. This is particularly the case given that almost all existing experiments with firms are with microenterprises (McKenzie, 2010 provides a review), and that experiments working with small and medium enterprises (SMEs) will typically have to work with government agencies or ministries rather than being researcher- or NGO-led. As a result, the lessons from matching grant programs are likely to be informative for many attempts to evaluate other government SME development programs such as export promotion activities, and training and support-service programs.

² Concurrently a randomized evaluation of a matching grant program was successfully undertaken by Bruhn et al. (2012) in Mexico, and we will draw some lessons from their work in this paper as well.

The proximate causes for our inability to complete randomized experiments are quite simple – in some cases governments were unwilling to randomly select recipients of the grants³, in others the application rates to the programs were too low to enable the planned selection of a random sample of eligible applicants, and in others continued implementation delays prevented us from starting. We therefore investigate what the underlying causes were, and discuss the roles of (a) political economy; (b) program eligibility criteria; (c) “last mile” delivery of the program; (d) incentives facing project implementation staff, and (e) difficulties matching impact evaluation funding cycles to the realities of these projects. These findings have insights both for the design of future matching grant projects, as well as for their possible evaluation. Designers of projects should reconsider the eligibility criteria and selection processes, work more on delivery, and better incentivize project staff. Future attempts to carry out experiments in this area need to moderate their expectations about timelines, better align funding cycles with them, use evaluation techniques designed to get more out of small samples, and also consider what we call “little IE” - experiments that can be done within the context of the broader project that address questions on design and efficiency, using IE as an operational tool, rather than asking what the overall impact of the project is.

This paper is not a critique of matching grants *per se*, but rather a synthesis of challenges in designing impact evaluations to measure the causal impact of such programs, and an attempt to draw out the insights that the attempted evaluations have provided on ways in which the implementation of matching grants may be improved in the future. We acknowledge that a number of matching grant programs have been implemented without impact evaluations, and may have been successful in achieving their targets – although, unless these targets are simple targets like funding a target number of firms, it would be difficult to tell if they have worked without an impact evaluation.

The remainder of the paper is structured as follows: Section 2 provides a brief introduction to the theory and practice of matching grant programs, Section 3 discusses our planned prospective evaluations of seven projects, and Section 4 examines the reasons why these experiments failed. Section 5 concludes by drawing out the lessons from this experience for both policymakers and researchers.

³ This concern would generally be raised at the beginning of the engagement on an IE design, but in some cases it became pressing only after the program had been launched.

2. The Theory and Practice of Matching Grants

Matching grant programs have been seen as a market-based approach to encouraging firms to purchase specialized services such as training for employees, implementation of standards and quality certification, product development, trade promotion, marketing support, and support for technology upgrading. Instead of the government directly providing these services or subsidizing suppliers of these services, the hope has been that providing subsidies to the buyers of these services will be more efficient, since they can then choose the specific services that their business needs, and the increased demand may stimulate a competitive response from existing and new independent providers of business services (Biggs, 1999). One of the earliest such funds was set up in 1961 by the Irish Export Board as a marketing development fund, while the earliest World Bank support for such a scheme were projects in India (India Engineering Export Development Project) and Indonesia (Export Development Project) in 1986, both of which required 50 percent matching funds from export-oriented firms (Phillips, 2002).

Over the past twenty years matching grants have been a mainstay of World Bank projects to enhance private sector competitiveness, especially in Africa (33 projects) and Latin America and the Caribbean (16 projects). Their popularity continues, with a review of 36 recent World Bank projects in Financial and Private Sector Development identifying 40% that including a matching grant scheme as a component of the project. While initially many of the programs were directed explicitly at exporters, over time the focus has evolved to more general private sector development and competitiveness. Grants have ranged in size from as small as \$200 in some of the African projects to as high as \$500,000 in some of the export- or bio-technology-oriented projects, with a typical project offering grants in the \$5000-\$10,000 range. The most common match proportion is 50 percent⁴, meaning that a \$5,000 grant would go towards purchases of \$10,000 in services.

In most cases the funding is restricted to the purchase of soft services, and excludes capital equipment, wages, or other recurrent business expenses.⁵ Sometimes higher matching proportions (e.g., the government paying 70 or 75 percent) are given for smaller firms, and lower matching

⁴ The rationale for setting any particular matching proportion is to maximize the private investment and public gains induced from each dollar of public spending. However there is no empirical evidence to support any given proportion over another, making the 50% match somewhat *ad hoc*.

⁵ Of the seven projects we consider, only one in South Africa supports capital equipment purchases with a current cost-sharing of 50% (it began in 2010 with a cost-sharing ratio of 35%). This is an independent government intervention, which is not part of a World Bank lending program.

proportions (e.g., the government covering only 25 percent) are given when the project does allow some capital expenditure

2.1 The Economic Justification for Government Funding

The assumption underlying the use of a matching grant program is that firms are not investing enough in business development services currently, and that by lowering the effective price paid for such services, firms will purchase more of them. The question which then arises is “why are profit-maximizing firms not purchasing these services already if it is profitable for them to do so?”. Then, even if not profitable, are there enough positive spillovers associated with these investments to warrant incentivizing this investment? Finally, one could also ask whether matching grants are the best way to incentivize private investment to induce the public gains, or could other forms of funding or credit be more effective? Answering these questions is necessary to provide an economic rationale for government intervention, and for understanding the underlying problem that the matching grants are intended to fix. This in turn will guide the questions that need to be answered by any evaluation.

A first set of reasons why firms may not be undertaking profitable investments is due to *market failures*. First, firms might be credit-constrained, and so unable to undertake lumpy profitable investments in business services. While the first best solution would then be to fix the credit market, there has been limited success in encouraging banks to lend to SMEs through partial credit guarantees and other such schemes, and so matching grants might be a second-best solution to this credit market failure. While banks may finance equipment since it can be collateralized, they are less likely to finance consulting, training, or high-risk intangible activities such as those associated with startups and innovation, for which matching grants are often used. A matching grant may also improve the signal of quality of the business investment proposals (since they are, ostensibly, reviewed by the government for quality, collecting information through a process that would be too costly for a bank to implement), reducing risks for banks and increasing the likelihood of successful loan applications when firms need access to credit in order to fund their matched proportion.

Second, owners of small and medium firms might be risk-averse, and avoid making investments in business development services that have high expected return, but which involve risk, because of their inability to insure this risk. Here the first best solution would be an equity market or venture capital market which would enable firms to share risks with investors. In the absence of such markets, the matching grant effectively increases the expected return on the investment in services by lowering its

price, therefore inducing firms to take on more risky profitable projects. In fact, some variations of matching grant programs are designed similarly to an equity investment, where the government buys a stake in the business as a way of providing a match to their investment, with the expectation that this will be repaid if the firm generates a profit – effectively making it a loan in the case of success and a grant in the case of failure.

Third, the missing market may be on the *supply side*, with a country having a rather limited supply of business development service providers that can offer the service and credibly signal that their quality. The matching grant program may therefore help by increasing the demand for such services enough to generate a market for new service providers to enter.

A second set of reasons concern *information and decision-making constraints*. Firm owners may just not have sufficient information about the range of possibilities for making use of specialized business development services in their firm, or undervalue them (e.g., Bloom et al., 2012), and therefore be unaware that profitable service investment opportunities exist. The costs of gathering the information or the complexity of the information may be too great, in which case improved salience of the information or support in interpreting it could stimulate investment. The matching grant in this case may work by drawing a firm owner's attention to the range of potential services that their business could use, and getting the firm owner to think about which may be profitable (with technical support in some cases). Alternatively, firm owners may intend to make the investments, but keep putting this off for another day because of present-bias in their preferences. A matching grant scheme with an application deadline may then be the nudge that firm owners need to bear the initial mental, time, and monetary costs of making the business development services investment, whose returns will not be felt until the future.

Third, the above factors may combine with regulatory barriers and small markets to result in a *lack of competition*. In the absence of competition, firms will not feel the same pressure to innovate and increase productivity, and there will be less reallocation from less productive firms not using these services to more productive firms. The matching grant may increase competition by enabling productive firms to overcome credit constraints or other market failures and begin competing with less productive firms.

Alternatively, it may be that firms are already profit-maximizing, and that investing in specialized business development services is not a profitable investment in terms of the private returns

to the firm. Nevertheless, if there are *positive externalities* (not captured by the firm) from purchasing specialized services, the social returns may exceed the private returns, leading to a justification for the government to subsidize the price. For example, if firms train workers in information technology, but then many of these workers leave to work for other companies or to start their own firms, the training may not be worth it from the firm's point of view, but the increase in human capital may still have positive returns from the point of view of society. Likewise increases in employment may have positive externalities for other citizens if unemployment is an issue, and other firms may learn from the innovations undertaken by firms who find new products, processes, or markets as a result of using these specialized services.

World Bank-financed matching grant programs typically have enterprise growth as their main objective. They often justify their need by the lack of supply of service providers and firm-level information constraints (as in the Mozambique Competitiveness and Private Sector Development Project for example), but commonly refer to externalities as a critical economic justification for the program. For example, the Mauritius Manufacturing and Services Development and Competitiveness Project Appraisal Document (PAD) notes that while intellectual property right protections help firms that develop new products to gain private benefits from these products, there are no such protections in place when a firm first introduces an existing product into the country. As a consequence, firms may under-invest in searching for technologies in which a country may have an untested comparative advantage, since the entire risk of failure is borne by the firm, whereas if successful, the benefits will be shared with imitators. Secondly, the PAD argues that since Mauritius has a reasonably flexible labor market, firms may under-invest in worker training and skills upgrading for fear of their workers leaving to work for other firms or to start their own firms.

What if none of the above cases hold? That is, what if firms are currently optimally investing in specialized business development services in an unconstrained way given the current market price of these services, and that there are no externalities from such services? In this case, the matching grant does not correct any distortions but rather *may create distortions*. Lowering the price of the specialized services below its market price will have a price effect and an income effect on the firm. The price effect will tend to cause firms to over-consume specialized services, some of which (the non-effective ones) may help deplete firms' investment capacity, while the income effect will give them more resources which they may use to buy other inputs or take as returns to the business owner. If the specialized services are lumpy purchases (e.g., acquiring a quality certification), then it is also possible that the price

effect might be zero – that firms who would have purchased these services anyway continue to do so (and just get a transfer from the government to do what they would have otherwise done), while firms who would have not purchased these services continue to not purchase them when the grant is available. Phillips (2002) notes a worst-case scenario in which subsidies may allow inefficient and unprofitable firms to stay afloat.

Note that these factors make the optimal selection of firms to participate in a matching grant program quite different from the decision process facing a venture capitalist or outside investor (McKenzie, 2011a). A venture capitalist is interested in identifying which projects have the largest private returns, whether or not those returns are caused by their funding. In contrast, the government should not be trying to pick the “gazelles” (enterprises which will grow fastest), but rather the projects which have the greatest additional impact from receiving government funding (referred to as “additionality” in the matching grant literature), and which have the greatest positive spillovers on other firms. Both factors mean that impact evaluation relative to a counterfactual is required, rather than just observing what the before-after change is in participating firms.

2.2 Examples of Matching Grants in Practice

We conducted qualitative interviews and case studies of some previous African matching grant recipients and examined a sample of approved applications for the projects considered for the evaluations in order to provide concrete examples of the activities and business development services that firms seek these grants for in practice. To preserve anonymity, we withhold both the names of the firms and their countries in describing these cases.

Some examples of approved applications include: a clothing manufacturer who wanted to hire a designer to develop new designs and patterns; a firm making medical and surgical devices who wanted to hire a consultant to implement lean manufacturing techniques in their factory; a firm doing interior design who wanted to conduct a market study to decide whether or not there was a market for selling new window shades; a legal firm seeking technical assistance to help it set up an outsourcing service; a small beach-side hotel that was looking for support in developing its website; an agri-business that wanted to conduct a study-tour in China in search of new equipment for their business; a producer of mineral water that wanted to obtain an ISO quality certificate; an IT company that wanted to participate in an industry-specific international fair; a carpentry that was seeking support for training eight of their

employees in using a new machine. Table 1 summarizes the approved activities over a period of two years for one of the programs we wanted to evaluate.

Among firms which had previously received grants, one of the most successful cases was a wind-generated electricity producer. It used a matching grant of \$8,000 to develop the feasibility studies that allowed it to enter into a joint-venture with a European firm and obtain an investment of US\$1 million. It took four years after funding for the project to become commercially operational, and within the first year of operation the firm was already supplying 14 percent of all the electricity consumed in the two districts it services. This project serves as a possible example of multiple levels of impact, with the firm receiving greater profits, consumers receiving better quality electricity service (brown-outs and black-outs were reduced), and the environmental benefits through clean energy provision. Even in this case it is not possible to ascertain absolutely whether or not this would have happened without the matching grant, but it is certainly the type of case where multiple market failures and externalities seem apparent, and public benefit from the grant seems likely. Another two examples of former grant recipients concern investment in worker and firm training. A civil engineering firm used the grant to help fund quality certification training for its workers, while a communications firm used the grant to develop an employee training scheme. Both firms said that they had been unsure about the effectiveness of such training, but after seeing the results from the grant funding, they had subsequently invested more of their own funding in continuing this activity.

For other approved matching grants the additionality needed to justify the government support is not as clear, particularly taking into consideration the information available to firms on the benefits of these investments and their lack of constraints in accessing them. In fact, according to a survey commissioned by the government in one of the countries where we started an evaluation, approximately 25 percent of the firms receiving the matching grant stated that they would have pursued the activity fully anyway and an additional 49 percent confirmed that they would be pursuing the activity in the absence of the matching grant, albeit using a less expensive consultant (31 percent of total), at a later stage (10 percent), or at a smaller scale (8 percent). Examples of less successful uses of the matching grant, or at least of apparent limited public benefits, include providing business services to firms who have the skills, knowledge and capacity to conduct the activities and are expected to invest in those areas. These comprise, for instance, an IT company receiving a matching grant to develop a website or a travel agency wanting continuous support to participate in various international tourism fairs. Applications like these are attractive both to the business (little risk and they continue doing the

work they would do anyway) and to the government (the application is likely to be well-polished with clear links to the business's operations), but they reduce the incentive for the government to extend its reach to identify cases where the most impact can be generated rather than cases that are most likely to be successful.

As these examples illustrate, there is substantial heterogeneity in the types of business development services and activities funded through matching grants. Many of the services and activities funded seem likely to have mostly private benefits (plus potentially fiscal and employment benefits), while the extent to which the services and activities funded actually will lead to significant externalities and spillovers is less obvious.

3. The Planned Prospective Evaluations of Seven Matching Grant Programs

In order to attempt to prospectively build impact evaluations into a range of private sector activities being financed through the World Bank, we worked in early 2010 with the World Bank's Director of Financial and Private Sector Development (FPD) in the Africa Region and her team to identify a set of projects to evaluate. Matching grants were one of the most common elements of private sector development projects being planned. In February 2010, the DIME-FPD initiative of the World Bank organized a 4-day workshop on impact evaluations in Dakar, Senegal, which brought together researchers, World Bank operational teams, and the key government counterparts in order to explain what is meant by impact evaluation, and to begin the process of building impact evaluations around key components of these projects. Through the course of these activities, we identified six projects - five of which were World Bank-funded - in which matching grants would be used. The seventh project comes from an engagement that started in 2008 with the Department of Trade and Industry (DTI) in South Africa to identify critical projects that should undergo an impact evaluation. In this process, the South African government had shown interest in evaluating a matching grant program focused on micro and small firms, notably because of the public debate around its effectiveness and its significant use of public resources.

Table 2 details the name of the seven projects, the amount of resources allocated, the average grant size expected, and the targeted number of recipient firms. Typically the projects were designed to last for three to five years, with funding either being allocated on a rolling basis or through multiple funding windows. For the World Bank-funded projects (5 out of 7 of these projects), three dates are provided to give some sense of the length of the time from project design until official implementation

began: the World Bank Regional Operations Committee (ROC) decision meeting date, which is the formal corporate review deciding whether or not to proceed with a proposed operation; the Board approval date when the project is officially approved by the World Bank, and the project effectiveness date when the country has met all the necessary conditions needed to begin receiving the funds from the World Bank. The actual length of time from project conception to the first matching grants being given out is substantially longer, involving both preparation time before the ROC decision meeting, and time taken to launch the program and receive applications after project effectiveness. Given that the ideal time to begin an impact evaluation is early in the preparation process, and certainly before Board approval, it can easily take two years from the time when work begins on impact evaluation design to the time when the first grants are given out to businesses.

3.1. Why is Impact Evaluation Important for These Projects?

In discussing these projects with World Bank operational staff and with governments, there was in some cases a sense that these were tried and tested policies that were politically popular, so there was no need for rigorous evaluation.⁶ However, in other cases there was genuine skepticism as to whether these grants really did provide benefits beyond the firms that received them, and concern over the costs of implementation. Phillips (2002) provides figures on the implementation costs of earlier matching grant projects in Africa, noting that they ranged from 19 percent of the amounts given out in Mauritius, to 40 percent in Kenya and 54 percent in Uganda. Moreover, these figures excluded some key costs, such as the costs of setting up project committees to evaluate the applications to the matching grant program. Therefore the full costs are relatively high compared to the amount of matching grant funds given out. Given that the resources originate in a World Bank loan that eventually needs to be repaid, citizens of these countries could therefore be paying to subsidize relatively well-off business owners with little public benefit in return.

We also discussed the point that the purpose of the impact evaluation is not just to learn whether the program works or not, but to help improve the future operation and targeting of such programs if we can learn what types of businesses the program works for. As Biggs (1999) notes, ideally matching grant programs should select projects with large economic and social returns to the country that would not otherwise find private funding – but in his direct questions to recipients of an earlier

⁶ Note that all World Bank projects are required to include project indicators and some form of evaluation, but at most these typically measure the number of firms receiving grants and sometimes before-after changes for the recipients captured through surveys of beneficiaries.

matching grant program in Mauritius, 80 percent of recipients said they would have made the investments in technology transfer even without the matching grants. In many countries the mindset was still very much one of trying to select the “best” firms or the ones most likely to succeed, not necessarily those for whom the grant funding could make a difference between success and failure. We hoped that the process of engagement of World Bank operational staff and governments in the impact evaluation would also help improve project implementation in this respect.

Finally, given that these programs are pervasive throughout Africa, we emphasized that there is a global public good element to the knowledge produced by any impact evaluations – the results would not only influence and inform policy in the country being studied, but also help inform efforts in other countries. For these reasons we were able to raise funding from different sources to help finance the impact evaluations, so that, when approaching these countries to try to implement evaluations, we were coming with some funding in hand to supplement their data collection costs.

3.2 Why Use Randomized Experiments as an Evaluation Tool?

There are a number of private sector development policies in which a randomized evaluation is difficult or near impossible (e.g., changing a national regulation, building new infrastructure). Matching grants on the other hand satisfy a number of conditions that make randomization a possibility: i) they involve selection of individual firms; ii) the numbers of firms involved can be large enough to potentially generate enough statistical power for measuring impacts (provided that appropriate methods are used, see McKenzie, 2011b, 2012); and iii) data on key outcomes may be measured reasonably well through firm surveys.

Furthermore, matching grants are a type of program for which it is hard for non-experimental methods to convincingly deal with selection issues – firms self-select into whether or not to apply for the program, and then government implementation units select which firms receive the grants. An explicit element of both selection stages is likely to be forward-looking behavior on the part of both the government and the firm. Firms that receive support are more likely to be perceived (by the firm and by government) as having faster potential growth in the future than firms that do not apply or do not get selected, even when they are observably similar in terms of current size and recent growth history. As such, matching on observable characteristics and differencing out of firm fixed effects is still unlikely to control for the differences between the firms which participate in the program and those which do not. In some circumstances regression-discontinuity designs may offer an appropriate (local) estimate of the

program's effects provided that explicit scoring thresholds are used, but the number of firms close enough to the scoring threshold to enable such an approach is likely to be small in most cases.

Randomization is also likely to be a more cost-effective option than other quasi-experimental methods. Propensity score matching, for instance, would require firstly developing a sampling frame of businesses (which in itself is likely to be a challenge in many African countries) and then collecting enough information from a large enough sample to be able to select the most appropriate matching group of businesses to serve as a counterfactual – increasing the costs, time and logistics required to prepare a baseline survey. In contrast, a randomized experiment would not require a sampling frame (since all participating businesses in the study would come from the actual applicant pool) and baseline data for all of these businesses could potentially be collected cheaply through an application form that captures key outcome variables and indicators on which the randomization is likely to stratify on.

3.3 What are the Key Initial Issues to Consider for an Impact Evaluation of these Projects?

As outlined in McKenzie (2010), there are a number of typical challenges in conducting an impact evaluation of policy operating at a firm level. First, the number of SMEs in many African countries is relatively small. In our group of matching grant impact evaluations, the DTI's BBSDP project – a matching grant for black-owned SMEs⁷ in South Africa - is a clear example of this challenge. The government was interested in supporting firms of a certain size because it already had a number of incentives for micro enterprises and wanted to shift focus from survivalists to established firms with potential for growth. In that vein, when preparing the project, the government had selected a minimum annual turnover of Rand 1 million (approximately USD 125,000) for firms to be eligible to apply. However, based on a set of representative firm-level surveys in South Africa, it seemed that less than 3 percent of the population of firms would qualify to participate in the program. This led to adjustments in the program, which culminated in a revised version of the project with a minimum annual turnover of Rand 250,000, increasing the chances of finding eligible enterprises.

Second, there is considerable heterogeneity in firm characteristics and performance, which affects the statistical power to detect the impacts of the programs and thus the sample size requirements. This is an important consideration, especially when one of the objectives of the study is to measure impact heterogeneity to understand for which type of firms the matching grant is most

⁷ The eligibility criteria imposed among other aspects the need for 51% of the ownership to be black-owned (African, Indian or Colored under the South African definition).

effective. For example, considering the limited effects of traditional management training programs (McKenzie and Woodruff, 2012) and unrestricted grants (Fafchamps et al., 2011) for female-owned enterprises, it is important to estimate the gender-disaggregated added-value of increasing access to a range of business development services and (in the case of South Africa) new equipment. While traditionally governments and donors think of a matching grant as a mechanism to support existing firms that already use a minimum set of business tools - and this often leads to targeting specific sectors such as manufacturing - it is not clear that the highest impact is achieved by investing in these better-equipped enterprises, which already have access to information, financial mechanisms, and technical skills. Testing whether matching grants are more effective for firms with more limited access to networks and business tools (such as female-owned enterprises concentrated in a limited number of low productivity sectors) is hence very relevant from a policy perspective. The downside of this goal though is that it implies the need to work with significantly larger sample sizes, adding another layer to an already demanding initial setup.

We were aware of these two issues, as well as the typical concerns about attrition and integrity of the data collected in advance of developing these studies. What we will explore next are additional barriers to evaluating matching grant programs that go beyond these concerns.

3.4 Implementing the Randomization

Phillips (2001) notes that all grant programs face a rationing problem since services are effectively supplied at below the free-market price, and so by definition there is excess demand for them. Given that the government is effectively giving away free money to firms, one might expect significant demand for this funding, resulting in the need for projects to select which firms receive it. Since we believe there is substantial uncertainty over which firms would best benefit from receiving these funds, our suggestion was for randomized evaluation based on an *oversubscription design*. The idea here would be to make the matching grant programs open for all firms meeting certain basic eligibility criteria, and then randomly select which firms would be awarded the grants. In the event of more demand for the grants than the project could fund, this would provide a fair and equitable way of ensuring that all eligible firms received an equal chance of benefitting from these public funds, and might reduce concerns about political connectedness determining who receives the grants.

We planned to implement the oversubscription design as follows. First, the program would set an initial limit on the number of grants that could be given out in the first year of the program. Each year

there would be one or two funding rounds. This limit would be based on natural manpower constraints (there is a limit to how many grants can get processed, screened, verified, and site-visited), funding constraints (typically the program has funding allocations by year), and how many firms apply in the first year. Next, the program would issue an initial call for proposals coupled with an advertising/marketing campaign to launch the program. In certain instances, we would provide support to government teams in preparing outreach logistical plans, which could include establishing partnerships with provincial government offices, microfinance institutions, banks, sector associations, with the objective of promoting the program, organizing workshops, helping firms prepare applications, and even collecting applications. Firms would then be aware that the program has its first round window open with a defined deadline - with potential behavioral effects on the decision to apply on time - and all applications received by a given date would be considered for the first set of funding, with projects not selected eligible to re-apply again in future rounds.⁸ The applications received would then go through an initial screening to ensure that they meet the basic eligibility criteria in terms of firm size, sector, and planned use of the funds, and, in most cases, would be visited by one of the project specialists.⁹ A random set of firms whose applications meet these criteria would then be selected for funding.

These intensive marketing campaigns stood to benefit both the impact evaluation design and the project implementation. From the point of view of the evaluation, they could boost the number of applicants, improving the likelihood that an oversubscription randomization method could be justified. On the project side, these campaigns would increase the reach of the program, ensuring that businesses with less access to information through geographic or network marginalization could become aware of the program and submit applications if interested. Since lack of information is likely to be associated with other constraints that the matching grants have been designed to overcome, it is plausible that these firms could have the most to gain from the program. The time-bound windows for applications also have the advantage of allowing the project implementation team to dedicate time initially on marketing and use the rest of the year working on due-diligence and processing of the grants, reducing the risk of over-committing to multiple activities. Moreover, these funding rounds would also provide information on what the demand is for these matching grants, and enable the identification of certain groups of firms that the program would like to reach but which are not applying (e.g., female-owned

⁸ In the first window of applications for the Mozambique project applications were concentrated in the last few days.

⁹ In certain cases, the activities to be conducted would finally be agreed at this stage, because often the firm had an idea of what it wanted but through a needs assessment conducted by the project, it would be able to refine it.

firms, or firms located outside the capital city). The program could then adjust its marketing efforts for future rounds to ensure it reaches out to these groups. Each subsequent funding round would again follow the same procedure of randomly selecting the target number of recipients from among the eligible firms.

We also considered two alternative forms of randomization. The first was a *phase-in design*, in which some firms would receive the funding now, and others would receive it later. This did not seem practical for three reasons. First, not all firms apply at once. Second, if firms knew that they might receive the funding in the future, they might adjust their behavior today, and third, some of their funding requests might be time-sensitive.

The second alternative was a *randomized encouragement design*. Under this approach, the program is open for all firms to apply. A baseline survey of firms would be conducted to learn about the potential demand among firms for the program, and to provide data for monitoring purposes on firms that get the grants and firms that do not. During the process of conducting the baseline survey, firms will be randomly chosen to receive additional “encouragement” to apply for the matching grants. This encouragement would be in the form of different marketing and informational approaches to increase awareness about the program. Since resources are limited and it is costly to visit firms individually, this approach would clearly not be feasible to follow for all firms. Therefore, the program fairly gives all firms the same *ex-ante* chance of receiving this additional encouragement, and uses this to learn about the needs of firms for the program, the effectiveness of different marketing strategies, and the impact on firms of getting the matching grants. The danger with this approach is that the encouragement may fail to encourage many firms into applying, so larger samples are typically needed than with a randomized experiment of applicants. In our first attempt to evaluate a matching grant project, we tried unsuccessfully to use this type of encouragement design to measure the impact of the Small Enterprise Development Agency (SEDA) in South Africa. The main reason for trying this approach was that SEDA had been operating on a first-come first-serve basis since 2004. Additionally, the government was reluctant to limit eligible applicants from receiving the grant, making randomization difficult to implement in practice. Given the lack of success in this first attempt, we considered this encouragement option as a less desirable alternative to randomization in subsequent evaluations (Goldstein, 2011).

4. Why Did the Randomized Experiments Not Happen?

Out of the seven projects for which we discussed impact evaluations, five initially agreed to implement the projects with an oversubscription-based randomized experiment included, one (SEDA, South Africa) agreed to implement the encouragement design approach and another (Mauritius) had recently obtained World Bank board approval but was not yet at the project effectiveness stage. This project had already agreed with the government that the program was to be run on a first-come, first-serve continuous basis for applications, and since this had been approved by both the Ministry of Finance of the country and the World Bank board, it was not possible to renegotiate these terms. An encouragement design coupled with a potential matched difference-in-differences back-up plan for evaluation was then designed for this country, giving seven potential randomized experiments.

In what follows we discuss what happened to the seven cases where we had an initial agreement to at least start an impact evaluation. In order to mitigate potential political sensitivities, we will not identify for which projects particular issues were a problem, but rather we highlight the set of issues that affected one project or another. In none of the cases was a randomized experiment implemented in the end, although non-experimental evaluations are ongoing in a couple of cases.

4.1 Proximate Causes of Randomization Not Being Implemented

We can group the proximate causes for not being able to implement the randomization into three interrelated reasons. Table 3 indicates which reasons apply to each project.

The first reason was widespread implementation delays in the projects. This made it impossible to start the evaluations until the projects started and caused conflict with funding deadlines for being able to do the evaluations. In one case, the government decided to soft launch the program while many details of the intervention were still under discussion. The details of the program were modified along the way during the soft launch period, which lasted over one year. This obviously had adverse consequences for the marketing strategy and the team's focus on conducting an impact evaluation.

Second, during these implementation delays, there was often turnover of government staff, leading to the government changing its mind about participating in the randomized experiment, or even about the project itself. In one case where the World Bank project was funding a matching grant with a 50 percent match, the government launched a second matching grant program, which was royalty-based, whereby the government provided 90 percent of the funds and the firm only 10 percent, with

firms then repaying through royalties on incremental sales if the activities funded succeeded in increasing sales. The World Bank project was ultimately cancelled following low demand for the original matching grant.

Third, the most common reason for the inability to randomize was that program take-up was low: our randomization strategy was based on selecting among an excess supply of eligible applicants. Instead, all projects struggled to obtain enough eligible applicants. In one country where the project was half over when we started discussing the evaluation, the disbursement rate was only 23% and new efforts were under way to increase spending. World Bank operations have an incentive to ensure that the loans “disburse”, that is, that the money promised in the loan is actually spent, and similarly project implementation units and governments want to show that their project can spend the money allocated to it, to help support future requests for such money. Therefore, when the few applications were actually received, grants were given to all firms that passed the eligibility criteria making randomization impossible.

These problems are not unique to the matching grant programs that our studies were built around. Goldberg and Ortiz del Salto (2011) undertook a review of World Bank matching grant programs and found that of the 42 projects with an original closing date before 2012, 79 percent were extended at least one year, reflecting both delays in initial implementation, as well as in disbursing money. In a review of 15 completed projects, they report that on average only 70 percent of the allocated money was disbursed, although this average masks considerably heterogeneity across projects, ranging from only 6 percent being disbursed in one project in Zimbabwe, to 99 percent in a project in South Africa and 116 percent in a project in Mozambique.

4.2 Underlying Causes of a Randomized Evaluation Not Occurring

It is easy enough to blame the inability to conduct an experiment on low program take-up, repeated implementation delays, or government unwillingness to randomize. But in order to understand how researchers and project teams can design projects and evaluations better in the future, it is important to delve deeper into some of the fundamental causes underlying these outcomes. These concern the incentives facing governments, firms, researchers, and project implementers. Again these are to some degree interrelated, but we discuss each issue in turn.

a) *Political economy and capture reasons*: Matching grant programs consist of handing out subsidies to firms. As a result of this free money, there is a risk of capture, with those in charge of the

funding at various levels wanting to use it to advance their own interests. In one case, the local Chambers of Commerce were used as implementing partners for the program. This had the obvious advantage of using existing organizational links with firms and ensuring that the business community had some buy-in for the program. However, our view is that the major Chamber viewed these grants as something they could use to buy goodwill among their members. As a result, the Chamber lobbied to make the eligibility criteria for the program exclude many small firms (who were not Chamber members), appeared reluctant to engage in serious outreach to firms beyond its member base, and, as a result, far fewer applications were received at this major Chamber which served the capital city and other neighboring areas home to the majority of firms than were received from a second Chamber of Commerce serving other areas which seemed more willing to reach out to smaller firms.

In another country, the national government launched a credit line administered by local governments for (existing or planned) businesses at the same time as the matching grant program. From the entrepreneurs' perspective, this option would provide cash-in-hand under a loan, which would, in theory, have to be paid back with interest, but in practice had low enforcement. These two initiatives led to a number of government officials at the national and local levels conflating the projects (often asking in workshops about the interest rate on the matching grant), and politicizing the programs, promoting the credit line in detriment of the matching grant because it was providing more effective political gains (cash in hand versus "just" a business development service).

A second manifestation of political economy concerned election politics. In order for projects to become effective and for matching grant programs to be launched, the government in each country had to undertake a number of steps. In multiple countries these steps were delayed due to election cycles, and in the case of a change in government the interest in pursuing the original project could be reduced. In one case, funding rounds and randomization had been agreed with the project team and the highest ranks of the Ministry, but popular demonstrations in the streets led to a cabinet reshuffling and the replacement of the Minister. The new Minister decided to push for a revised industrial policy strategy with clear sectors of focus. With that goal in mind, the Minister decided to use this program as a mechanism for driving the government's industrial policy, and in that process, decided to stop any randomization.

As with all project evaluations, another major political economy issue is how much desire and pressure there is from top levels of government to know the impacts of policy. The case of the successful completion of a randomized experiment in Puebla, Mexico by Bruhn et al. (2012) offers a

sharp contrast to the seven attempted experiments in African contexts. In Puebla, the director of the program wanted to prove that it worked, and had heard of the MIT Poverty Action Lab. Impetus for the randomized evaluation therefore came directly from the head of the program himself, who directed program staff to follow all the suggestions made by the evaluation team in order to ensure that the program could be rigorously evaluated. While there appear to have been some incentives for him to have done this (proving his program worked would help in budget discussions), it also appears to in part reflect the research interests and educational background of this director, making it harder to replicate elsewhere.

Ownership of the evaluation at the level of the implementation team is helpful for a successful impact evaluation. In a couple of our matching grant studies, our initial discussions and workshops were with the management team, top government officials, and Monitoring & Evaluation (M&E) specialists to garner buy-in for the impact evaluation. Conversations with the day-to-day operations team that would implement the project happened only in a second stage - in one of these cases in particular, the project was based on a specific region of the country, which we visited only after the initial interactions. The top officials were interested in learning about their projects and using the evaluations to prove the programs' effectiveness and obtain further funding. Nonetheless, when we started working with the implementation teams on the ground, we faced significant blocks in implementing the evaluations. The political economy of headquarters versus regional offices came into play and the implementation teams' feared that we would be auditing their work and assessing their competencies, despite numerous discussions about the objectives of the study. For these two projects, the implementation teams saw the evaluation as risky (these were two existing projects with problems) and although the top levels officials considered the evaluation to be important, it was not important enough to justify potentially damaging already difficult relations with operational teams.

b) *Overly strict eligibility criteria:* The eligibility criteria in many cases were set in a way to exclude the vast majority of firms in most African countries. They tended to be based on a notion of picking the firms which would grow the fastest, not necessarily the ones which would benefit the most from the grant. Despite most African firms having fewer than 10 workers and not being fully formal (McKenzie, 2011b), most programs required firms to be fully formal. This could be fine if the existence of the program served as an incentive to bring firms into the formal registration and financial systems, but in some cases the programs required the firms to have several years of formal registration, including audited accounts and current tax records. In one case, we sat with the project implementation team to

assess the reasons for low-take up. In addition to traditional issues of firms' lack of working capital to fund the acquisition of the business development services upfront and other red tape issues, we noticed that one of the first questions in the application form was whether or not the firms were registered, although in theory this was not a criterion for selection into the program. Firms were also asked during application to attach their business registration certificate. This could naturally put-off a number of firms

AVO

After discussions with the project implementation team of another case, it was clear that the targeting was reaching fewer firms than initially anticipated, while a critical group of businesses in a lower turnover range were not being serviced because of the minimum turnover eligibility criterion. This led to a change in the criteria (reducing the turnover requirement from USD 60 thousand per year to USD 30 thousand and increasing the matching proportion by government) and a dramatic increase in applications (Figure 1).

c) *“Last mile” issues and red tape*: Much of the emphasis in matching grant programs concerns the design of the eligibility criteria, and the bureaucratic procedures involved in setting up a project implementation unit to run the program. Far less attention is given to what Mullainathan (2009) terms the “last mile” problem – how do you actually get firms to take up a program offered? Marketing is part of the need, but there are a number of issues related to psychology and behavioral economics that are also important.

Marketing efforts varied across our seven projects. In order to ensure a sufficient number of applications and make firms aware of the program, in one country impact evaluation funding was used to supplement the project's marketing budget and radio and television advertisements were produced and shown to get the word out to firms. In other countries the outreach was less intensive and slower – in one country there were only two branch offices for the program throughout the country, neither particularly noticeable, and the marketing materials produced were not visually very appealing.

Developing strong awareness campaigns can be a challenge. Despite one of the projects investing considerably in a large launch and communications campaign, a survey of 209 businesses conducted one year after the launch found that none of the sampled businesses had heard of the program. Of those contacted, 39% were interested, and ultimately only 6% were both interested and eligible to participate, highlighting the need for targeted awareness campaigns when eligibility criteria are restrictive (Figure 2).

Many projects create a number of roadblocks that prevent or inhibit even interested and potentially eligible firm owners from applying. For example, in designing one project, the default was to have each firm separately obtain a letter from the tax authority saying that they were not delinquent on any tax payments. This required firms to spend additional time and effort proving to a government program that the firms were current with another part the government! This requirement was changed after our impact evaluation team pointed out the problems, but still other bureaucratic steps which increase the cost of applying remained. In qualitative interviews, one firm remarked that “even to apply to these matching grants (intended to help firms get business development services) firms already need to use a consultancy to put together their candidacy folder”. While the use of online forms can in principle reduce some of this bureaucracy, in some African countries where business usage of the internet is still low, the need for applications to be online and correspondence to be via email potentially served as another barrier for small firms to access the program.

Other elements of how the program was to be implemented may have deterred some firms from applying in the first place. One big issue is how the funding is actually paid out. Governments want to try and ensure that the funds are used for the purpose firms said they would be used for, and ensure that firms are actually putting in their matching share. They also want to try and ensure that firms are not colluding with business service providers to pay overinflated prices and then splitting the subsidy without work actually getting done. The solution in some cases has been for the government to require firms to undertake a procurement process which involves getting multiple bids in writing, requiring the firms to pay 100 percent upfront, and then reimburse the firms for 50 percent of the amount upon presentation of receipts and proof that the service was actually provided. But if the reason for firms requiring the matching grant in the first place is in part difficulty obtaining financing, this may prevent some firms from being able to use the program. Moreover, when there are already issues with the extent to which firms trust the government, and with the speed at which the government processes reimbursements, some firms may be unwilling to take the perceived or actual risk of never getting reimbursed. A better alternative, which the World Bank has used in Lesotho, has been for the matching grant unit to directly pay the approved service provider the government’s share of the cost, so that firms do not need to prepay and then be reimbursed.

An obstacle for evaluations that get to this stage (none of ours did), is that even conditional on being awarded a grant, a non-trivial number of firms may choose not to take it up – in the successfully implemented randomized experiment in Puebla, Mexico, Bruhn et al. (2012) report that only 80 out of

150 successful applicants actually proceeded to take-up the grants even though these firms had all signed letters of interest and applied for the program, and that in their case firms were only expected to pay between 10 and 30 percent of the costs of consulting services. The result from this low take-up is lower statistical power, making it harder to detect the impact of such programs. Although we did not manage to implement the randomized experiments, the projects that we covered also suffered from low take-up. In one of our cases shown in Figure 3, out of the 165 firms that had activities approved (under two windows in mid-2011 and early 2012), only 51 firms completed the activities proposed as of September 2012. An effective take-up rate of 31% has large implications for the power of the study. Identifying a 20% increase in firm turnover assuming an average turnover of USD 30 thousand with a coefficient of variation of one would require a sample of 786 firms assuming full compliance.¹⁰ A 31% take-up rate, however, would increase the sample size required to 8168 in an intention-to-treat estimate – clearly an infeasible task even if the true impact is large. In another country, the government provided an incentive to accredited individual consultants to help firms in their application process. This led to a perverse incentive to artificially increase submissions, with a number of applications covering activities that the firm never had the intention of implementing. The government has since then revised the incentive scheme to pay a proportion upon completion of the activity.

d) *Incentives facing project staff:* Project implementation units are typically staffed by individuals on fixed wage contracts who do not have strong monetary or career incentives to generate an excess number of applications or increase screening mechanisms to improve the targeting and quality of the submissions. More applications mean simply more work for staff without any clear reward. Considering the high rotation in these types of “private-sector like” public jobs¹¹, they do not necessarily value the potential scaling up of the project if strong positive results are shown by the impact evaluation. Therefore, the effort is often kept relatively low, even in cases where the extra mile to generate the excess number of applications would just require organizing promotion better, involving the right partners in the awareness campaigns, and conducting more regular one-to-one meetings with firms to assess needs and explain the available services. In one case, the government was interested in reaching to firms outside of the capital city mostly for economic/political reasons (not impact evaluation-related), but the implementation unit blocked setting individual targets for performance

¹⁰ A power of 0.8 and a two-sided significance level of 5% were used in this calculation.

¹¹ Members of the Project Implementation Units are paid out of the project – hence it is government money – but they usually do not sit in the Ministry, having a separate structure and higher-paid jobs than traditional government officials.

(eventually these were set later once the Minister blamed the project for non-performance in certain regions of the country). In three other cases, it was clear - and communicated to us at various moments in time – that there was a conflict between different team members in terms of the appetite for reaching out to new applicants. This lack of incentives from project implementation units is sometimes also shared with World Bank operations teams, who in theory would like to see an impact evaluation of their program, but have few formal incentives for taking on, or requesting from the government team, the additional work required to implement a successful evaluation.

e) *Funding cycles for impact evaluation*: Funding for the impact evaluations came from monitoring and evaluation budgets included in the projects themselves, from World Bank trust funds the researchers applied for, and from external donors. The non-project funds typically were given for periods of two years, with any funding not spent during this period having to be returned. This created a severe mismatch with the cycle of the project that was being evaluated, especially once implementation delays occurred. Funders were in some cases willing to give a single extension of up to one-year, but were not willing to extend funding longer than this. Moreover, many funders were inflexible in terms of which countries they would fund the research on, with an extreme focus on Africa, so that it was not possible to take funding raised for one matching grant impact evaluation and transfer it to learn from a matching grant project taking place in another part of the world. As a result, in some cases we had to make decisions to cancel the evaluations - despite the effort on design, the researchers' involvement and the field coordinators having been hired - rather than wait and see whether applications would pick up in future rounds of the project.

5. Lessons for Future Evaluations and Future Matching Grant Projects

Many of these risks were not unforeseen, with one grant application for example noting that we anticipated three potential risks: timing of the project launch getting delayed; having the government and/or World Bank operational staff changing their minds and deciding not to proceed with randomization; and the risk of insufficient demand for the matching grants program. We took actions to mitigate these risks through working closely with operational task leaders and government units, building some slack for delays into our evaluation timelines, and active attempts (including helping to fund advertising campaigns) to try to ensure enough applications. Despite these actions we thought it was unlikely that all the evaluations would be able to proceed as randomized evaluations, but hoped that by taking a portfolio approach, we would at least enable a couple of cases to succeed. As noted, this was too optimistic, and so in this section we attempt to draw lessons for future attempts.

One difficult question facing researchers deciding to work on these evaluations is the extent to which they should actively try to change the design, and especially the implementation, of the project they are evaluating. On one hand, this can be part of the value of a prospective impact evaluation, with the questions raised by researchers improving project design and performance. This approach is the one often used in evaluations led by academics working with NGOs, with full-time field coordinators tasked with micro-managing the roll-out of the program being evaluated. However, this raises concerns about scalability and generalizability, and may result in a different set of firms participating in the program than would be the case without researcher involvement.¹² If the program has heterogeneous impacts on firms, the average treatment effect for the set of firms participating in a project which has had substantial researcher involvement in implementation may then be potentially quite different from what the average treatment effect would be in a standard program without researcher involvement. However, our view is that many of the recommendations for ways to design matching grant programs to make them more amenable to conducting randomized experiments are also likely to make the projects better for firms as well, so that in practice this conflict may not always arise.

Given this, we have the following recommendations:

1. *Have more realistic expectations about the time it takes to implement these projects:* this applies to project staff designing the timelines for projects, to researchers considering evaluating them, and to funders. In most cases it seems likely that it will take at least three years from the start of researcher involvement in these evaluations to see results, unless the government can be more efficient than usual. Grant agencies should consider funding portfolios of evaluations over longer time horizons, with the expectation that funding will shift across projects and to encompass potentially new projects within that general topic range. McKenzie (2012a) sets out some additional implications for funders.
2. *Change the mindset from picking winners to picking positive treatment effects:* The mindset of policymakers and operational staff is often that these projects should be trying to identify the firms that are gazelles (i.e., potentially high-growth). This raises the risk of little additionality, as the project simply ends up subsidizing firms that would grow anyway, and makes it more difficult to set eligibility criteria that will enable more firms to participate. In contrast, once the emphasis is shifted to funding the firms that would benefit most from the grant funding, it

¹² Just promoting the program to firms that would not typically apply to the program can have implications on external validity. If the researchers' effort creates new demand from different type of firms, it would ideally be important to identify upfront which these are to be able to disentangle the impacts by type of firm.

should become more immediately apparent to government staff that we have relatively little idea as to what identifies these firms, and so there is a role for expanding the program to serve a broad range of firms and then scientifically measuring impact to see who the program works best for.

3. *Focus more on eligibility criteria and on making it easier for firms to apply:* the most innovative companies may not be those which are already formal, large, and long-established. Market failures may apply more to younger and smaller companies, so ensuring that eligibility criteria do not rule these firms out seems a good idea. But just as important as the initial criteria is making it easy for firms that meet these criteria to find out about the funds, apply without undue burden, and receive the funding promptly. Incentives for both government and World Bank operational staff are often focused more on getting projects launched and ensuring money is spent, rather than on the efficient delivery of services once projects have been launched.
4. *Use evaluation techniques that can obtain more power out of relatively small samples:* despite all these efforts, in many African countries there are just not that many firms with more than one or two workers, and the projects will likely end up funding fewer than 100 firms a year. McKenzie (2011b, 2012b) discusses how one can obtain more power out of such sample sizes by restricting attention to a more homogeneous set of firms, and taking multiple measures on them. An extreme case is illustrated in Bloom et al. (2012), which conducted an experiment with only 20 textile firms in India, but as these firms were all in the same sector and weekly production data were collected, impacts of management consulting were still possible to measure. In practice, matching grant programs are unlikely to have the same level of homogeneity of firm production technologies and the same richness of data, but steps in this direction can still be made.
5. *Practice “little IE” as well as “big IE”:* our goal in these failed experiments was impact evaluation (IE) around the overall impact of these grants. We refer to this as “big IE”. While important, we have seen the difficulties in doing this. Given the large number of design issues that matching grant programs grapple with that do not have a strong evidence basis for decision-making, it seems possible to also conduct more “little IE”. By this we mean experiments embedded in the overall broader project that can help learn not whether or not matching grants have an impact overall, but instead about what the impacts of changing different design features in these grants could be. This shifts the question from “what is the impact of the matching grant scheme?” to “what is the most effective way to maximize the impact of the matching grant scheme?”. For

example, experiments could be done to test different ways of implementing reimbursements, different incentive schemes for project implementation staff to see which generates more and better applications, examining demand for the program when different match proportions are offered, testing alternative information campaigns, or overlaying complimentary interventions such as credit guarantees to address other identified constraints and test for interactions.

This offers value both to the researchers and to operations teams. On the one hand, thinking about using impact evaluations to tweak and test features of the program has the potential to use the impact evaluation as an operational tool to improve the speed and efficiency of the project disbursements (for instance, by identifying mechanisms that can increase the quantity and quality of business applications and claims). On the other hand, this offers the opportunity to test cross-cutting mechanisms that may be applicable to more than just matching grants (e.g., how do deadlines influence firm behavior?).

6. *How should we conduct impact evaluation if most outcomes are failures?* Hall and Woodward (2010) document that among venture-capital backed entrepreneurs almost three-quarters receive nothing at exit while a few receive over a billion dollars. Karlan et al. (2012) likewise make the point that in some cases smaller firms experimenting with training and capital investments may have negative expected returns in the short-run from such experimentation, but with some outliers succeeding. It seems plausible that the impacts to society from matching grant projects may follow a similar model – the wind-powered electricity production deal cases are likely to be much rarer than cases of firms undertaking actions that have mostly private payoffs or no realized benefit at all. Since much of the focus of randomized experiments has been on identifying mean impacts, more work needs to be done to consider how to best apply these tools when most of the expected action is at a tail.

Many of these lessons are likely to be applicable beyond the case of evaluation of matching grants. Other activist government programs to help SMEs are likely to run into many of the same issues of implementation delays, limited take-up, political constraints, a relatively small number of firms, and difficulties matching funding cycles to project cycles. It is hoped that sharing the lessons of our experiences with matching grants may help improve the likelihood of success of impact evaluations in those areas as well.

References

Allcott, Hunt and Sendhil Mullainathan (2011) "External validity and partner selection bias", Mimeo. Harvard University.

Biggs, Tyler (1999) "A microeconomic evaluation of the Mauritius Technology Diffusion Scheme (TDS)", Regional Program on Enterprise Development Discussion Paper no. 108.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts (2012) "Does management matter? Evidence from India", *Quarterly Journal of Economics*, forthcoming.

Bruhn, Miriam, Dean Karlan and Antoinette Schoar (2012) "The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in Mexico", Yale Economics Department Working Paper no. 100.

Castillo, Victoria, Alessandro Maffioli, Sofá Rojo and Rodolfo Stucchi (2011) "Innovation policy and employment: Evidence from an impact evaluation in Argentina", IDB Working Paper no. IDB-TN-341

Crespi, Gustavo, Alessandro Maffioli, and Marcela Melendez (2011) "Public support to innovation: The Colciencias experience", IDB Working Paper no. IDB-TN-264

Goldberg, Michael and Daniel Ortiz del Salto (2011) "Matching grants: A review of matching grants in projects promoting private sector development", World Bank internal Powerpoint presentation available at http://intresources.worldbank.org/INTLAC/Resources/257803-1226691316407/Matching_Grants_Promoting_Private_Sector_Development.pdf

Goldstein, Markus (2011) "A disappointment with encouragement", *Development Impact Blog Post* April 5, <http://blogs.worldbank.org/impactevaluations/node/524>

Gourdon, Julien, Jean Michel Marchat, Siddharth Sharma and Tara Vishwanath (2011) "Can matching grants promote exports? Evidence from Tunisia's FAMEX II program", pp. 81-106 in Olivier Cadot, Ana Fernandes, Julien Gourdon and Aaditya Mattoo (eds.) *Where to spend the next million? Applying impact evaluation to trade assistance*. World Bank: Washington, D.C.

Hall, Robert and Susan Woodward (2010) "The burden of the nondiversifiable risk of entrepreneurship", *American Economic Review* 100(3): 1163-94.

Fafchamps, Marcel, David McKenzie, Simon Quinn and Christopher Woodruff (2011). "When is capital enough to get female microenterprises growing? Evidence from a randomized experiment in Ghana". CSAE Working Paper WPS/2011-11.

Karlan, Dean, Ryan Knight and Christopher Udry (2012) "Hoping to win, expected to lose: Theory and lessons on microenterprise development", Mimeo. Yale.

Lopez-Acevedo, Gladys and Hong Tan (2011). *Impact evaluation of small and medium enterprise programs in Latin America and the Caribbean*. The World Bank.

McKenzie, David (2012a) "Improving funding of impact evaluations: end the fiscal year and other rules that have outlived their usefulness", *Development Impact blog post*, June 24, <http://blogs.worldbank.org/impactevaluations/node/829>

McKenzie, David (2012b) "Beyond baseline and follow-up: the case for more T in experiments", *Journal of Development Economics*, 99(2): 210-21.

McKenzie, David (2011a) "Should development organizations be hunting gazelles", *All about Finance blog post*, December 8, <http://blogs.worldbank.org/allaboutfinance/should-development-organizations-be-hunting-gazelles>.

McKenzie, David (2011b) "How can we learn whether firm policies are working in Africa? Challenges (and solutions?) for experiments and structural models", *Journal of African Economics*, 20(4): 600-25.

McKenzie, David (2010) "Impact Assessments in Finance and Private Sector Development: What have we learned and what should we learn?", *World Bank Research Observer*, 25(2): 209-33.

McKenzie, David and Christopher Woodruff (2012) "What are we learning from business training and entrepreneurship evaluations around the developing world", World Bank Policy Research Working Paper.

Mullainathan, Sendhil (2009) "Solving social problems with a nudge", TED India Talk, http://www.ted.com/talks/sendhil_mullainathan.html.

Phillips, David (2001) "Implementing the market-based approach to enterprise support: An evaluation of ten matching grant schemes", World Bank Policy Research Working Paper no. 2589.

Phillips, David (2002) "The market-based approach to enterprise assistance—an evaluation of the World Bank's market development grant funds", *Small Enterprise Development* 13(1): 26-37.

Rasmussen, Ole Dahl, Nikolaj Malchow-Møller, and Thomas Barnebeck Andersen (2011) "Walking the talk: the need for a trial registry for development interventions", *Journal of Development Effectiveness* 3:4, 502-519

Ravallion, Martin (2009) "Should the randomistas rule?", *The Economists' Voice*, www.bepress.com/ev, February 2009.

Table 1: Example of Matching Grant Activities in One of the WB-Funded Projects

Firms' Activities by Category	Number of Activities Approved	Proportion of total amount (in USD) of grants approved
Employee training	116	30%
Websites and e-commerce	58	15%
Quality certification	15	13%
Trade fair participation	33	10%
IT systems	30	8%
Design of promotional materials	52	6%
Improvement of production efficiency	9	5%
Business plan	12	5%
Market research	6	3%
Short-term management contracts	7	3%
Product development research	2	1%
M&A, partnerships, investors' search	7	1%
Distribution systems	1	1%
Packaging design	3	0%
Total	351	

Table 2: Planned Evaluations of Matching Grant Programs in Africa

Country	Project	Match (%)	Total Amount	Expected Grant Size	Anticipated number of grants	Key Project Dates		
						Decision Meeting	Board Approval	Project Effectiveness
Cape Verde	SME Support & Economic Governance Project	50-75%	\$860,000	\$2,000-10,000	280	Oct-09	Apr-10	Nov-10
Ethiopia	Sustainable Tourism Development Project	50%	\$3 million	\$50,000	60-100	Mar-09	Jun-09	Feb-10
Malawi	Business Environment Strengthening Technical Assistance	50%	\$2 million	\$5,000-10,000	200-400	Mar-07	May-07	Oct-07
Mauritius	Manufacturing & Services Development & Competitiveness Project	50%	\$8 million	\$12,000-20,000	500-670	Oct-08	Jan-10	Mar-10
Mozambique	Competitiveness and Private Sector Development Project	50-70%	\$4.16 million	\$7,500	500	Oct-08	Feb-09	Oct-09
South Africa	Black Business Supplier Development Programme (BBSDP)	50-80%	\$17 million	\$14,000	1,200	NA	NA	Sep-10
South Africa	Small Enterprise Development Agency (SEDA)*	70-90%	\$1.6 million*	\$1,250	1300*	ongoing since 2004		

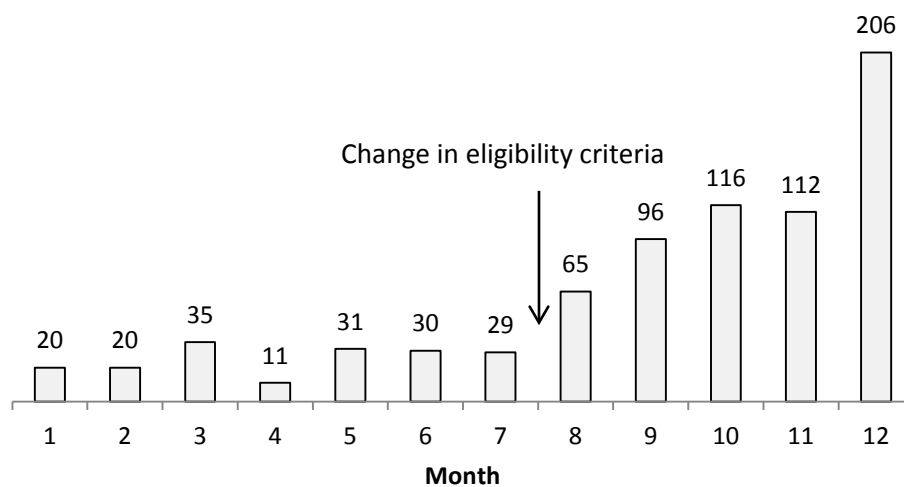
Notes: * denotes per year. The South African projects were funded by the Department of Trade and Industry, while the other country projects were all funded through World Bank loans.

Table 3: Summary of Causes Behind Inability to Conduct Randomized Experiments

Project	Was Project Ultimately Cancelled?	Proximate Causes			Underlying Problems				
		Implementation Delay	Govt. decided not to randomize	Low Take-up	Political Economy or Capture	Funding Cycle Inflexibility	Eligibility Criteria too strict	Project implementation unit incentives	"Last mile" delivery
A	Yes		X	X	X	X			
B	No		X	X				X	X
C	No	X		X	X	X	X		X
D	No		X	X			X	X	X
E	No	X	X		X			X	X
F	No		X	X	X		X		
G	No	X		X		X	X	X	X

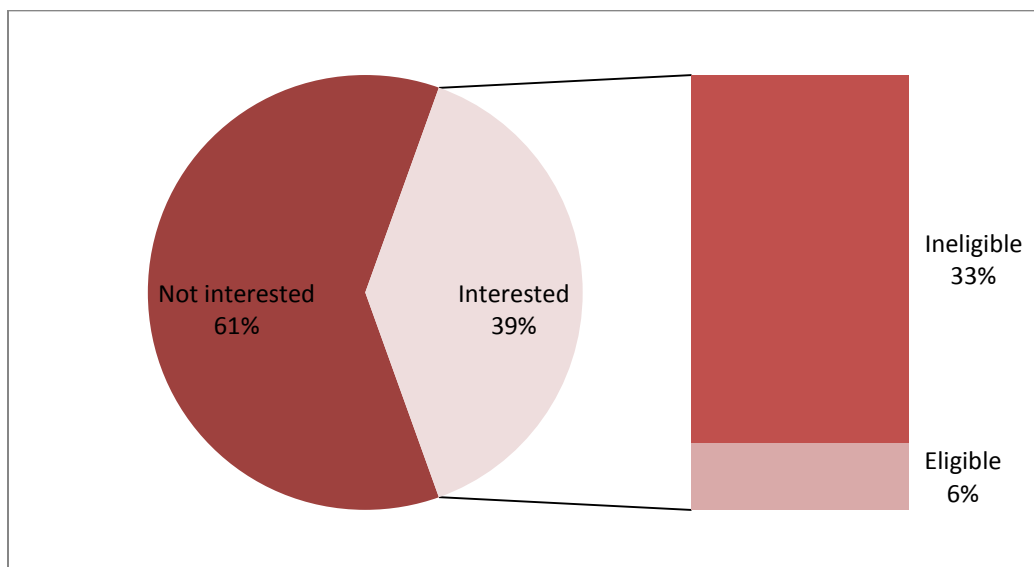
Notes: Projects A through G indicate the seven different proposed randomized trials and are listed in random order.
Govt. decided not to randomize includes cases where there was an initial agreement to randomize, followed by a later change.

Figure 1: A Relaxation of Eligibility Criteria Increased Applications



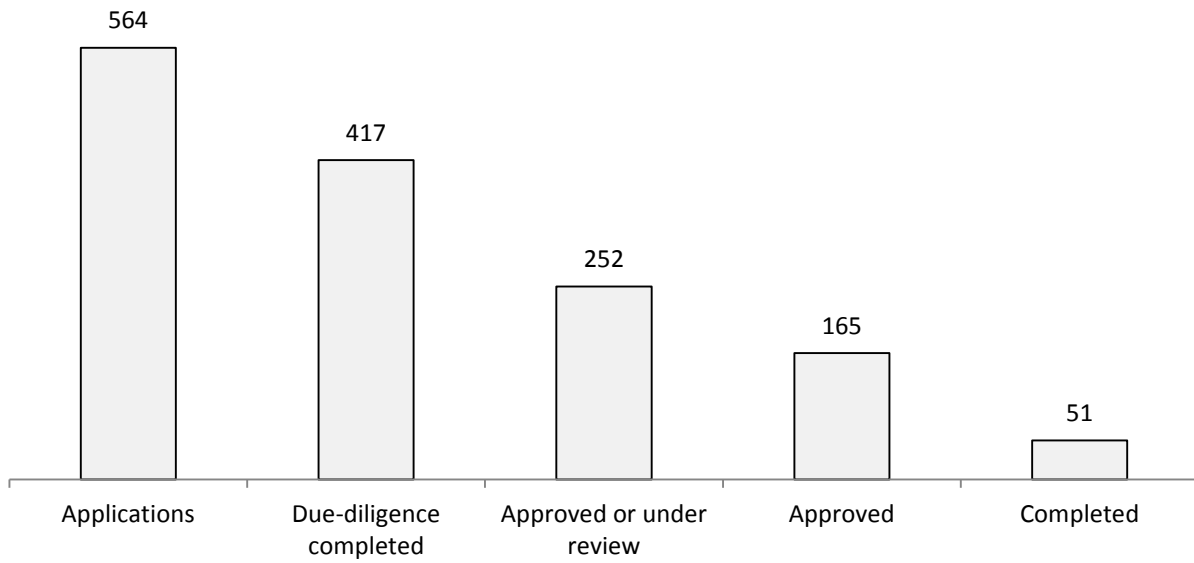
Notes: the figure shows the numbers of applications to the matching grant program. The change in eligibility criteria was from a minimum turnover of 60 thousand USD to a minimum turnover of 30 thousand USD.

Figure 2: Interested and Eligible Businesses Based on Business Awareness Study



Notes: the figure is based on a survey of 209 businesses in one of the seven countries conducted one year after launching the matching grant program.

Figure 3: Number of Firms Applying, Getting Approvals, and Completing Activities



Notes: the figure shows information on the number of firms in each category for a matching grant program in one of the seven countries. On average, each firm applies for circa 2 activities.