

Task Force for the Development of a Toolkit for Impact Evaluation of Public Credit Guarantee Schemes for SMEs

Consultative Document

Prepared by the Secretariat of the Task Force



WORLD BANK GROUP



Strengthening Financial Sectors

Contents

| | |
|--|----|
| 1. INTRODUCTION TO THE TOOLKIT | 5 |
| 1.1. Background | 5 |
| 1.2. Evaluation in the World Bank Group/FIRST Principles | 6 |
| 1.3. The Need for Evaluation and Main Issues | 7 |
| 1.4. Purpose of the Toolkit..... | 8 |
| 1.5. Structure of the Toolkit..... | 8 |
| 2. ASSESSING THE IMPACT OF CREDIT GUARANTEE SCHEMES FOR SMEs | 10 |
| 2.1. What is impact evaluation? | 10 |
| 2.2. Causal Inference and Counterfactuals..... | 10 |
| 2.3. Experimental Approach | 11 |
| 2.4. Non-Experimental Approach | 12 |
| 3. IMPLEMENTING THE IMPACT EVALUATION OF CREDIT GUARANTEE SCHEMES FOR SMEs | 16 |
| 3.1. Identifying the Evaluation Questions | 16 |
| 3.2. Selecting the Impact Evaluation Method | 18 |
| 3.3. Data and Sampling | 21 |
| 3.4. Setting Up the Evaluation Team | 23 |
| 3.5. Time and Budget | 24 |
| 3.6. Production and Dissemination..... | 25 |
| References..... | 27 |

Figures

| | |
|--|----|
| Figure 1: Credit Amount in Relation to Firm Size (Post Intervention) | 13 |
| Figure 2: Difference in Difference | 15 |
| Figure 3: Simplified Results Chain for a CGS | 17 |
| Figure 4: CGS Evaluation Methodology Decision Matrix | 21 |

Boxes

| | |
|---|----|
| Box 1: Outline of an Evaluation Report of CGS | 26 |
|---|----|

Glossary

| | |
|-------|---|
| AADFI | Association of African Development Finance Institutions |
| ACSIC | Asian Credit Supplementation Institution Confederation |
| AECM | European Association of Mutual Guarantee Societies |
| CGS | Credit Guarantee Scheme |
| DID | Difference in Difference |
| ED | Encouragement Design |
| FIRST | Financial Sector Reform and Strengthening |
| PSM | Propensity Score Matching |
| RCT | Randomized Control Trial |
| RDD | Regression Discontinuity Design |
| REGAR | Ibero-American Guarantee Network |
| SME | Small and Medium Enterprise |

FOR DISCUSSION ONLY

Task Force for the Development of a Toolkit for Impact Evaluation of Public Credit Guarantee Schemes for SMEs

Secretariat

| | |
|-------------------|--|
| Pietro Calice | Senior Financial Sector Specialist, World Bank Group |
| Bilal Husnain Zia | Senior Economist, World Bank Group |
| Moni Sengupta | Senior Financial Sector Specialist, FIRST Initiative |
| Roya Vakil | Financial Sector Specialist, World Bank Group |

Members

| | |
|--------------------------|--|
| Mohamed Al-Jafari | Director General, Jordan Loan Guarantee Corporation |
| Zac Bentum | Managing Director, Exim Guaranty Company of Ghana |
| José Fernando Figueiredo | Special Honorary Chairman, AECM; Coordinator of the Global Network of Guarantee Institutions |
| Jong-goo Lee | Head of Int'l Relations, Korea Credit Guarantee Fund |
| Pablo Pombo | Secretary General, REGAR |

Technical Focal Points

| | |
|---------------------------|---|
| Abdelmoughite Abdelmoumen | Head of Studies, CCG Morocco |
| Sarangua Amarsanaa | Marketing Specialist, CGFM |
| Lourdes Rosario M. Baula | Group Head, SBGFC |
| Jean-Louis Leloir | Special Adviser to the Board, AECM |
| Perbaganan Kuppusamy | Chief Risk Officer, CGC Malaysia |
| Amin Masudi | Head of Non-Bank Guarantee, Jamkrindo |
| Horacio Molina Sanchez | Professor, University Loyola Andalucia |
| Hiroshi Tahara | Senior Economist, Japan Finance Corporation |

TOOLKIT FOR IMPACT EVALUATION OF PUBLIC CREDIT GUARANTEE SCHEMES FOR SMEs

1. INTRODUCTION TO THE TOOLKIT

1.1. Background

1.1.1. Limited access to finance, particularly bank credit, is a long-standing hurdle for SMEs, with varying severity of financing constraints across countries. In developing countries, between 55 percent and 68 percent of formal SMEs are either unserved or underserved by financial institutions, with a total credit gap estimated in the range of US\$0.9 trillion to US\$1.1 trillion.¹ Financing is also a major constraint in advanced economies, where financing gaps for SMEs were exacerbated by the 2008-09 financial and economic crisis. SMEs are typically at a disadvantage with respect to large firms when accessing financing. SMEs face higher transaction costs and higher risk premiums since they are typically more opaque and have less or inadequate collateral to offer. These market failures and imperfections provide the rationale for government intervention in SME credit markets.

1.1.2. An increasingly popular form of government intervention is represented by credit guarantee schemes (CGSs). These are specialized institutions or programs set up by the government which pledge to repay some or the entire loan amount to the lender in case of default of the SME borrower. This reduces the lender's expected credit losses, acting as a form of insurance against default. CGSs typically charge a fee for their services. A CGS can lower the amount of collateral that the SME needs to pledge to receive a loan, because it effectively provides a substitute for collateral. Similarly, for a given amount of collateral, the CGS can allow riskier SME borrowers to receive a loan and/or to obtain better lending conditions (e.g., longer maturities, lower rates, higher loan amounts), because the guarantee lowers the risk faced by lenders.

1.1.3. In addition to improving access to finance for SMEs, CGSs can potentially play a more important role, especially in countries with weak institutional environments, by improving the information available on SME borrowers in coordination with credit registries and bureaus, and by building the credit origination and risk management capacity of participating lenders, for example through technical assistance for the setup of SME units. Moreover, CGSs can also play an important countercyclical role, providing support to small businesses during a downward economic cycle.

1.1.4. More than half of all countries, including advanced economies, have a CGS in place and the number is growing. CGSs have become a common feature of financial systems across the world. However, their expansion as a policy tool to ease access to credit for SMEs has triggered greater demand for evidence on their impact. This demand

¹ World Bank (2014).

concerns in particular CGSs' quality, efficiency and effectiveness. CGSs are established to improve access to finance for SMEs and in some cases to deliver other important results such as investment and jobs. Whether or not these results are actually achieved is a crucial public policy question yet one that is not often examined. More commonly, CGS managers and policymakers focus on controlling and measuring inputs and immediate outputs – how much money is spent, how many guarantees have been issued, how many loans have been granted – rather than assessing whether the CGS has achieved its intended goal of improving access to finance for SMEs and contributing to economic development.

1.1.5. Impact evaluations of CGSs like any other public policy are needed to inform the government on a range of decisions, from curtailing inefficient programs to scaling up interventions that work and selecting program alternatives. Once impact evaluation results are available, they can also be combined with information on CGS costs to perform a cost-benefit analysis and speak to the efficiency of the policy intervention. This Toolkit aims to offer an accessible introduction to the topic of impact evaluation and seeks to provide a core set of impact evaluation tools which can be applicable to CGS operations.

1.2. Evaluation in the World Bank Group/FIRST Principles

1.2.1. CGSs can play an important role in unlocking lending to SMEs. However, they may add limited value, prove costly and, more importantly, create distortions in credit markets when their design and implementation are flawed. With the objective of identifying internationally-agreed good practices to assist governments to effectively and efficiently establish, operate, and evaluate CGSs for SMEs, in 2015 the World Bank Group and the FIRST (Financial Sector Reform and Strengthening) Initiative convened a global task force of experts. The result was the development of the “Principles for the Design, Implementation, and Evaluation of Public Partial Credit Guarantee Schemes for SMEs” (the Principles).² These are a set of good practices covering four key dimensions deemed critical for the success of a CGS: (i) legal and regulatory framework, (ii) corporate governance and risk management, (iii) operational framework, and (iv) monitoring and evaluation.

1.2.2. Monitoring and evaluation is a critical component of a CGS to report and communicate its activities and achievements. In particular, Principle #16 calls for a systematic and periodic evaluation of the CGS' performance, which should be publicly disclosed. CGS performance involves three key dimensions: outreach, or the capacity of the CGS to meet demand for guarantees; financial sustainability, or the CGS's capacity to contain losses while continually maintaining an adequate capital base relative to its liabilities on a going concern basis; and impact or additionality, both financial additionality and economic additionality.

1.2.3. Financial additionality refers to incremental credit volumes granted to eligible SMEs because of CGS activities. These extensive margin effects include access to credit for

² World Bank and FIRST Initiative (2015). Available at: <http://www.worldbank.org/en/topic/financialsector/publication/principles-for-public-credit-guarantee-schemes-cgss-for-smes>.

SMEs that otherwise would not be able to obtain financing as well as higher loan amounts. On the intensive margin, financial additionality includes more favourable conditions for eligible SMEs in loan size, pricing and maturities, and reduced amount of collateral required to obtain credit. Economic additionality refers to the economic welfare that the CGS creates as a result of its operations. In particular, economic additionality speaks to the effect of guarantees on output, investment and employment.

1.3. The Need for Evaluation and Main Issues

1.3.1. Evaluating a CGS' impact is necessary to account for the effective use of public resources, measure the achievement of the CGS policy objectives, and improve its performance, as reflected in Principle #16. Therefore, an impact evaluation should be a fundamental component of any public CGS. Yet measuring impact is not an easy task and involves a trade-off among evaluation techniques and budget considerations, among others. The design of an impact evaluation framework requires an answer to a number of important questions, whose balance inevitably determines the form and content of the final output.

1.3.2. First is the choice of the analytical method. There are two basic options in undertaking impact evaluations – the quantitative approach and the qualitative approach. Quantitative evaluation involves assessment of the impact of programs through a comparison of outcomes between the group in receipt of the guarantees and some form of “control group”, for example a similar group of SMEs that have not benefited from the guarantees or the same SMEs before and after receipt of CGS support. Such data may be collected directly from the firms themselves, from official data or from both.

1.3.3. Qualitative assessments are commonly based on opinions of program participants and stakeholders about the policy, its success, and its limitations. Through surveys, interviews, focus groups, and/or case studies, qualitative evaluations collect additional information that sheds light on the satisfaction of participants, on the relevant mechanisms responsible for the impact of the intervention, and on general feedback to adjust and improve the operation of the policy or intervention.

1.3.4. Both approaches present advantages and disadvantages. The most evident advantage of quantitative techniques is that they can provide clear answers on the additionality of a CGS. If well done, a quantitative approach can get as close as possible to the true impact of a CGS. On the other hand, quantitative techniques can be technically challenging, require extensive data collection, and can deliver narrow results, focused primarily on issues of effectiveness and efficiency. Qualitative methods, on the other hand, can be more straightforward and have the benefit of providing additional information beyond that associated with quantitative techniques, and can represent a good entry point for impact assessment. However, qualitative evaluations have the major disadvantage of not providing reasonable estimates of the CGS' impact.

1.3.5. Next, and related to the choice of the evaluation technique, is the cost of implementing an impact evaluation. Evaluations can be costly and therefore to justify mobilizing the technical and financial resources needed to carry out a high-quality impact evaluation the stakes should be high, either because the CGS is strategically and financially important for the government or because it reaches a large number of SMEs. Additionally, little should be known about the effectiveness of the CGS's operations due to the absence of any previous study of the CGS or evidence from countries with similar circumstances.

1.4. Purpose of the Toolkit

1.4.1. The *Toolkit for Impact Evaluation of Public Credit Guarantee Schemes for SMEs* has been created with the objective of identifying a set of uniform methodologies for assessing the financial and economic impact of public CGSs as systematically and objectively as possible. A uniform methodology set can ensure comparability across time and countries, and therefore can provide a global reference for impact evaluations of CGSs.

1.4.2. The Toolkit is intended to provide guidance to CGS managers, policymakers and stakeholders on how to design and implement an effective and efficient CGS impact evaluation. Impact evaluations assess whether or not a program has achieved its intended results. Impact evaluations can help strengthen the evidence base for developing CGSs around the world and help direct resources to be spent more effectively to improve access to finance for SMEs.

1.4.3. The Toolkit reviews a variety of existing impact evaluation techniques – randomized experiments, regression discontinuity design, propensity score matching, and difference-in-difference – and proposes a selection process for an impact evaluation framework that is rigorous, credible, and at the same time practical, straightforward, and relatively inexpensive to implement. Accordingly, the Toolkit suggests a hierarchy of evaluation designs to fit the operational context of CGSs.

1.4.4. The approach to impact evaluation in the Toolkit is largely intuitive and technical notations are reduced to a minimum. The emphasis is rather on concepts and methods that underpin any impact evaluation. The methods are drawn directly from applied research in the social sciences. In this sense, the Toolkit brings the empirical tools widely used in economics and other social sciences together with the operational realities of CGSs to pragmatically present an impact evaluation framework for real-world application.

1.5. Structure of the Toolkit

1.5.1. After this introductory Chapter, the Toolkit is divided in two parts. Chapter 2 provides an overview of impact evaluation and introduces different modalities of impact evaluation such as prospective and retrospective evaluations. It then explains what impact evaluations do, what questions they answer, what methods are available to conduct them,

and advantage and disadvantages of each. The approaches discussed include randomized selection methods, regression discontinuity design, propensity score matching and difference-in-difference.

1.5.2. Chapter 3 provides a roadmap for designing and implementing a CGS impact evaluation. It first discusses how to formulate evaluation questions in the context of CGSs and hypotheses useful for policymaking. These questions and hypotheses form the basis of impact evaluation because they determine what the CGS evaluation is looking for, including the outcomes of interest. The chapter then suggests a hierarchy of appropriate methods that fit the operational rules of CGSs. It finally touches upon some operational steps to implement an impact evaluation such as collecting data, setting the evaluation team, budgeting and timing for the evaluation, and producing and disseminating the results.

FOR DISCUSSION ONLY

2. ASSESSING THE IMPACT OF CREDIT GUARANTEE SCHEMES FOR SMEs

2.1. What is Impact Evaluation?

2.1.1. The impact evaluation of a CGS involves evaluating the changes in the outcomes of interest that can be attributed to the CGS itself. Therefore, the key challenge in carrying out a meaningful impact evaluation is to identify the causal relationship between the CGS and the outcomes of interest. In other words, the impact evaluation looks for changes in the outcome of interest that are directly attributable to the CGS. The focus on attribution and causality is the hallmark of impact evaluation and determines the methodology to be used.

2.1.2. In order to estimate the causal impact of a CGS on outcomes, any methodology chosen must estimate the so-called counterfactual, that is, what the outcome would have been for eligible SMEs had they not participated in the scheme. To be truly additional and achieve its policy objectives, a CGS must ensure that beneficiary SMEs obtain outright financing and/or improved terms and conditions that would not otherwise obtain if it was not for the guarantee they receive (financial additionality). Furthermore, guaranteed SMEs are expected to contribute more to investment, exports, job creation, value added and other relevant economic outcomes than non-guaranteed SMEs (economic additionality).

2.1.3. Impact evaluations can be divided in two general categories: prospective and retrospective. Prospective evaluations are developed at the same time as the CGS is being designed and are built into its implementation. Baseline data are collected prior to implementation for both the treatment and control groups. Retrospective evaluations assess impact at a given time after the CGS has started implementation, generating treatment and comparison groups ex-post. In general, both prospective and retrospective evaluations can produce strong and credible results if implemented correctly.

2.2. Causal Inference and Counterfactuals

2.2.1. The impact evaluation essentially constitutes a causal inference problem. Assessing the impact of a CGS is equivalent to assessing the causal effect of the policy on the outcomes of interest (loan amount, interest rate, maturity, collateral, investment, sales, export, jobs). One can think of the impact of a CGS as the difference in the outcome of interest for the same SME with and without the guarantee. Yet measuring the same firm in two different states at the same time is impossible. This is what is commonly referred as “counterfactual problem”. Solving the counterfactual problem requires the identification of a perfect duplicate of the guaranteed SME.

2.2.2. Though no perfect duplicate exists for a single SME, statistical techniques exist that can be employed to generate two groups of SMEs that, if large enough, are statistically indistinguishable from each other. In practice a key objective of the impact evaluation is to identify a group of guaranteed SMEs (treatment group) and a group of non-guaranteed SMEs (control group) that are statistically identical in the absence of the CGS intervention and estimate the average impact of the CGS rather than the impact on each SME. An ideal

control group should be statistically identical to the treatment group in terms of both observable (such as past profitability, size, owner characteristics, etc.) and unobservable (such as owner motivation, preferences, etc.) characteristics. If the two groups are identical with the only exception that one group is guaranteed by the CGS and the other is not, then any difference in the outcomes of interest can be attributed to the CGS.

2.2.3. Impact evaluation techniques deal with these issues and allow the identification of a proper counterfactual group to compare with the group of SMEs that were granted a credit guarantee to estimate as cleanly as possible the effect of the CGS. In general, impact evaluation techniques can be classified in two broad categories: experimental and non-experimental. Experimental methodologies randomly assign SMEs to both the treatment and the control groups, allowing evaluators to cleanly estimate the counterfactual. Non-experimental techniques use statistical analysis to identify the most appropriate set of firms that can form the control group.

2.3. Experimental Approach

2.3.1. Randomized experiments, also known as randomized control trials (RCT), are the best methodology for ensuring a valid counterfactual. The essence of a RCT is the random assignment of the scheme's participants to a fraction of the eligible participants. This ensures by design that each participant is equally likely to be placed in the treatment or control group, allowing for a credible attribution of the outcomes observed. In other words, program participation is the only reason different average outcomes are observed in the two groups. More importantly, with a large enough pool of eligible participants, random assignment means that the treatment and control groups will be statistically identical.

2.3.2. Evaluators can use the RCT approach under a certain set of conditions. The program's conditions must make it feasible to assign participants to the treatment and control group; randomization must be able to occur prior to the program's beginning; the program must have a large enough number of participants to allow for meaningful statistical analysis; and it must be easy for participants to comply with the assignment. It is important to remember that, while the RCT is often considered the "gold standard" for impact evaluation, these evaluations can be costly to implement, and implementation is rarely perfect. Issues such as contamination, attrition (where participants drop out of the study), or other concerns that jeopardize random assignment can make data analysis quite complex. Probably for these reasons, there is no evidence of RCT ever conducted to evaluate the impact of CGSs.

2.3.3. Encouragement design (ED) is a form of RCT, which is useful when evaluating programs with voluntary enrolment or programs with universal coverage, such as CGSs. In this method, some units, i.e. SMEs, selected randomly receive incentives to participate in the scheme that is available to all eligible firms. Such encouragement can be in the form of information, marketing materials, or financial incentive. However, it is important that the program is not already popular, and the promotion device should be effective enough

that it will significantly affect the likelihood that firms will participate in the CGS. An example, of an encouragement design mechanism can consist of reducing the cost of applications for a random subset of SMEs to the CGS. If firms receiving the incentive are more likely to apply to the CGS, this mechanism will predict scheme participation. Like the RCT, no ED has ever been used to assess the impact of CGSs.

2.4. Non-Experimental Approaches

2.4.1. In cases where RCT methodology is not feasible, quasi-experimental methods can be used. There are many instances and many important policy questions where RCTs cannot be employed, for instance where a policy or program has already been implemented and prospective planning is not possible. Examples of non-experimental methods that are particularly suited to evaluating the effectiveness of CGSs include regression discontinuity design, propensity score matching, and difference-in-difference estimation.

2.4.2. Regression Discontinuity

2.4.2.1. Regression discontinuity design (RDD) is a methodology used to assess interventions that have a continuous eligibility index with a clearly defined cut-off score to determine who is eligible and who is not. The idea is that firms who are just above (below) and just below (above) the selection criteria or threshold score are very similar to each other; thus, if a credit guarantee is granted to those who score above (below) this threshold, those who score just above (below) the threshold will be the treatment group, while those who score just below (above) it will be the control group. RDD takes advantage of existing program rules, and thus allows it to be evaluated without changing program design. It can be a retrospective evaluation tool as it does not rely on random assignment.³

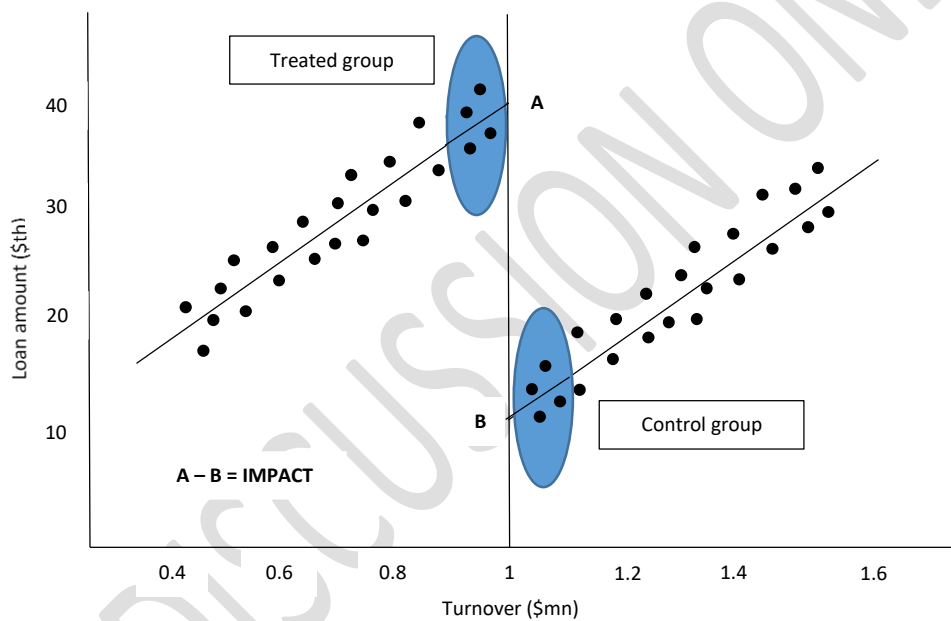
2.4.2.2. CGSs for SMEs generally aim to improve access to credit for eligible firms below (or above) a certain threshold in terms of number of employees, sales, total assets, credit scoring or a combination of these criteria. This exogenous cut-off can provide a design that allows the identification of the intervention's impact, since SMEs at the margin of the threshold would not differ substantially. Assume, for example, that the CGS targets all firms in a country with turnover below \$1 million and that this limit is used as eligibility. This implies that all firms with turnover above \$1 million are ineligible, regardless of their credit quality. In this example, the continuous eligibility index is simply the value of turnover and the cut-off score is \$1 million.

2.4.2.3. The RDD strategy exploits the discontinuity around the cut-off score to estimate the counterfactual. Intuitively, eligible firms with turnover just below \$1 million will be very similar to firms with turnover just above \$1 million. The latter firms can then be used as a comparison group for the former firms. Figure 1 presents a possible post-intervention situation conveying the intuition behind the RDD

³ Examples of RDD applied to impact evaluation of CGSs include Armentos et al. (2015) for Chile; and De Blasio et al. (2017) for Italy.

identification strategy. Average outcomes (for example, amount of credit) for eligible firms with turnover just below the \$1 million threshold (A) are higher than average outcomes for ineligible firms with turnover just above \$1 million cut-off (B). Given the continuous relationship between turnover value and amount of credit before the CGS intervention, the only plausible explanation for the discontinuity we observe post-intervention must be the existence of the credit guarantee issued by the CGS. In other words, since the firms in the vicinity of the \$1 turnover threshold had similar baseline characteristics, observed differences in the loan amounts (A – B) between the two groups is a valid estimate of the CGS impact.

Figure 1: Credit amount in relation to firm size (post-intervention)



2.4.2.4. RDD estimates the average impact of the policy around the eligibility cut-off, where treatment and control units are similar. This implies that the estimates cannot necessarily be generalized to units whose scores are further away from the cut-off score, that is, where eligible and ineligible SMEs may not be similar. The estimation of local treatment effects also raises challenges in terms of the statistical power of the analysis since a relatively large number of observations around the cut-off score is required in order to measure impact estimates. An additional caveat when using the RD design concerns the enforceability of the eligibility cut-off. RDD inferences are valid as long as SMEs are unable to manipulate participation in the scheme. With these limitations in mind, RDD yields unbiased estimates of the impact in the vicinity of the eligibility cut-off, taking advantage of the operational rules of the program.

2.4.3. Propensity Score Matching

2.4.3.1. Propensity Score Matching (PSM) is a non-experimental approach that can be used to identify a control group that is statistically equivalent to the treatment

group. The idea behind matching is to compare each firm in the treatment group to a control-group firm that is very similar to them. Because there are many dimensions (firm size, profitability, leverage, urban-rural location, etc.) along which the evaluator might like to match firms, PSM can be used to incorporate many different characteristics. PSM essentially uses statistical techniques to construct an artificial control group by identifying for every possible SME under treatment a non-treatment SME that has the most similar characteristics possible.⁴

2.4.3.2. PSM takes a number of measures and combines them into a single score, the propensity score, which represents the predicted probability of participating in the CGS. Firms with similar propensity scores have a similar likelihood of receiving the intervention, and thus can be compared across the treatment and control groups. The impact of the intervention will then be measured as the difference in outcomes between the treated group and the control group.

2.4.3.3. PSM analysis can be challenging. First, it requires that only those participants who have a good match in the non-treatment group can be analyzed. This means that PSM requires a large dataset to allow for a large enough set of usable data points, including baseline data. The dataset must also contain enough information to adequately estimate propensity scores for each firm. Finally, as with other retrospective methods, a key limitation of PSM analysis is that the matching across firms is only as good as the available data and firms in the treatment and control groups may still differ on unobservable characteristics.

2.4.4. *Difference-in-Difference*

2.4.4.1. The difference-in-difference (DID) method does what its name suggests. It compares the changes in the outcome of interest over time between the population that is enrolled in a program (treatment group) and the population that is not (control group). The use of the DD estimator requires baseline data, that is, data on the outcomes of interest for both the treatment group and the control group are needed from periods before and after the intervention.⁵

2.4.4.2. Assume that in the example presented in the previous section there are no sufficient observations in the vicinity of the cut-off point of \$1 million turnover, which would not permit the RDD. Simply observing credit amounts for SMEs before and after the participation in the CGS will not give us the causal impact of the CGS because many other factors are likely to influence access to credit over time. At the same time, comparing SMEs that received a guarantee with SMEs that did not receive a guarantee will be problematic if unobserved reasons exist for why some eligible SMEs received the guarantee and others did not. The DID compares the before-and-after-

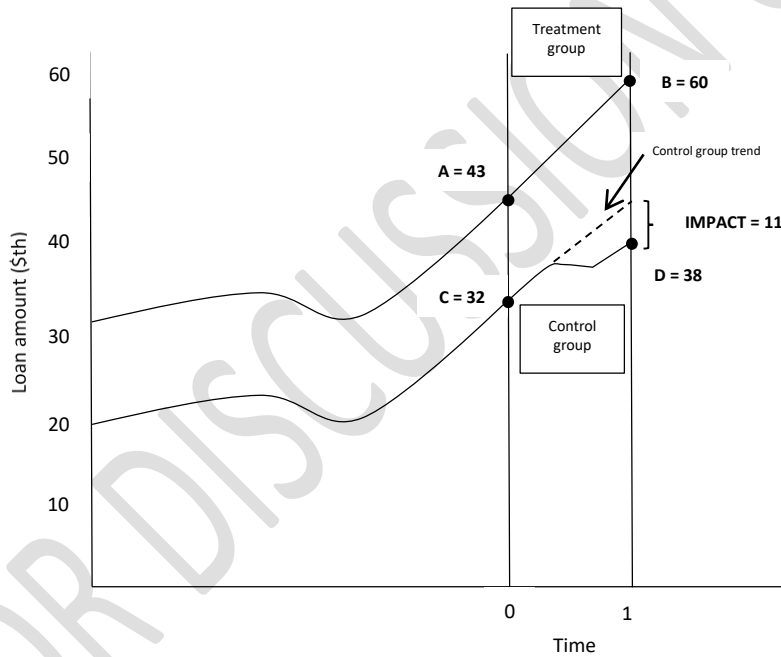
⁴ Examples of this technique applied to CGS impact evaluation include: Oh et al. (2009) for Korea; Uesugi et al. (2010) and Ono et al. (2013) for Japan; Brown and Earle (2017) for USA.

⁵ Examples of DID approach in the context of CGS impact evaluation include: Lelarge et al. (2010) for France; Zucchini and Ventura (2009), D'Ignazio and Menon (2013), and Boschi et al. (2014) for Italy; Kowan et al. (2015) for Chile; Arraiz et al. (2014) for Colombia; Asdrubali and Signore (2015) for Central and Eastern European countries.

changes in the outcome of interest, i.e. credit amount, for the group of firms that benefited from the guarantee to the before-and-after-changes of a group that did not participate in the scheme. In other words, the counterfactual being estimated here is the changes in credit amount for the comparison group.

2.4.4.3. Figure 2 illustrates the DID method. A treatment group of SMEs participates in a CGS program and a control group does not. The before-and-after outcome of interest, i.e. credit amount, for the treatment group are A and B, respectively, while the credit amount for the control group goes from C, before the program is implemented, to D, after the CGS is implemented. In DID, the changes in credit amount for the control group ($D - C$) represents the counterfactual. This amount is then subtracted from the change in credit amount for the treatment group ($B - A$) to obtain the impact. In summary, the impact of the CGS is computed as the difference between two differences: $(B - A) - (D - C) = (60 - 43) - (38 - 32) = 11$.

Figure 2: Difference-in-Difference



2.4.4.4. The key assumption of DID analysis is that in the absence of the treatment the change in outcomes for the treatment group would be identical to the change in outcomes for the control group. While this assumption is not formally testable, its validity should always be carefully examined to ensure that the DD impact is not biased. If data are available for several years before the treatment, then one easy way to assess the validity of the equality of trends assumption is whether pre-treatment trends were equal between the two groups. It is also useful to control for baseline observable firm characteristics between the treatment and control groups.

3. IMPLEMENTING THE IMPACT EVALUATION OF CREDIT GUARANTEE SCHEMES FOR SMES

3.1. Identifying the Evaluation Questions

3.1.1. The starting point of the impact evaluation of a CGS consists of formulating a study question or a set of study questions that focuses the research and that is tailored to the public policy interest. In other words, the research question should draw from the mandate and policy objectives of the CGS described in its mission statement. The assessment then involves generating credible evidence to answer that question.

3.1.2. CGSs are established to address market failures, which prevent and/or constrain SMEs from accessing credit. Typically, the mission statements of CGSs around the world emphasize access to finance for SMEs that lack adequate collateral (financial additionality). However, many CGSs have broader developmental objectives such as supporting job creation, facilitating industrialization programs, fostering entrepreneurship, developing export capacity and promoting investment in innovation (economic additionality).

3.1.3. The fundamental impact evaluation questions for a CGS can, therefore, be formulated as:

What is the effect of the CGS on access to finance for SMEs?

What is the effect of the CGS on economic development?

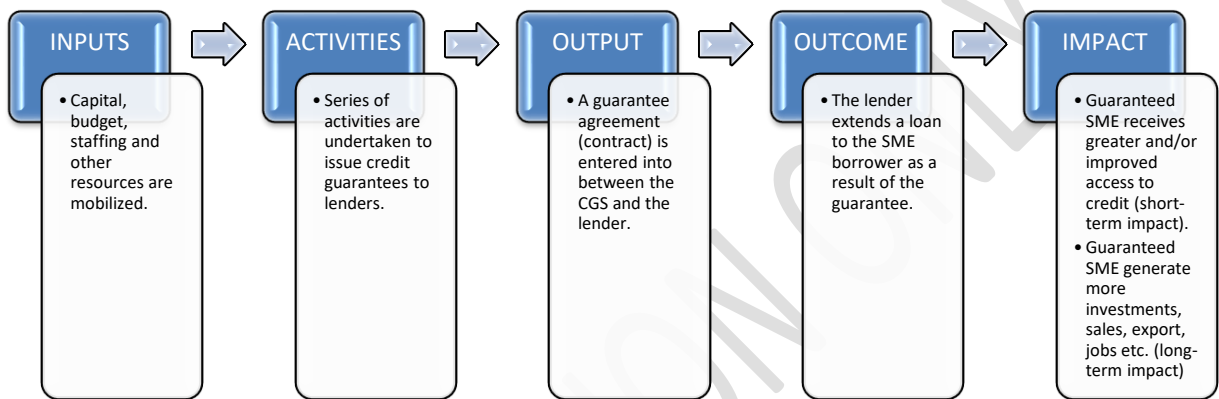
3.1.4. A theory of change can help specify the impact evaluation question. A theory of change describes how an intervention is supposed to deliver the desired result, highlighting the causal logic of how and why a particular policy will achieve its intended outcome. A theory of change is typically modelled using a results chain. This sets out a logical outline of how a sequence of inputs, activities and outputs for which a policy is responsible interact with behaviour to establish pathways through which impact is achieved.

3.1.5. A basic results chain for a CGS, which is described in Figure 3, will map the following elements:

- *Inputs*: resources available to the CGS, including capital, operating budget and staff.
- *Activities*: work performed to issue credit guarantees, including credit analysis, due diligence, etc.
- *Outputs*: the tangible service produced by the CGS, i.e. a guarantee agreement or contract.

- *Outcome*: the result likely to be achieved once the partner lender uses the output, that is the guarantee agreement, and extends a loan to the SME borrower.
- *Impact*: the guaranteed SME obtains better access to credit than it would otherwise (financial additionality). However, the impact concerns also longer-term goals such as SME's increased contribution to economic development (economic additionality).

Figure 3: Simplified Results Chain for a CGS



3.1.6. Based on the results chain outlined above, it is possible to formulate the hypotheses to be tested using the impact evaluation. Specifically, the hypotheses to be tested will be the following:

- The CGS entails first-time SME borrowers to enter the formal financial system.
- Guaranteed SME borrowers obtain higher volumes of credit than non-guaranteed SMEs.
- Guaranteed SMEs pay lower interest rates than non-guarantees SMEs.
- The CGS allows guaranteed SMEs to obtain longer loan maturities than non-guaranteed SME borrowers.
- Guaranteed SME borrowers benefit from reduced collateral requirements.
- Guaranteed SMEs generate more investment, sales, export, jobs etc. than non-guaranteed SMEs.

3.1.7. A clearly articulated results chain provides a useful map for selecting the financial and economic impact indicators to be used for the impact evaluation. Based on the discussion above, this Toolkit suggests the following outcome variables, which can be used alone or jointly:

- *Financial additionality – (short-term impact)*
 - Loan amount (\$).
 - Loan collateral (\$ or %).
 - Loan interest rate (%).
 - Loan tenor (months/years).
- *Economic additionality (long-term impact)*
 - Firm employment (number).
 - Firm investment/value added (\$).
 - Firm sales (\$).
 - Firm exports (\$).

3.2. Selecting an Impact Evaluation Method

3.2.1. The key to estimating the causal impact of a CGS is to find a valid comparison group by applying one of the methods described in chapter 2. The overarching principle guiding the selection of the impact evaluation method is that the operational rules of the CGS determine the evaluation methodology, and not vice versa. The operational rules most relevant for the evaluation design are those that identify who is eligible for receiving a credit guarantee and how they are selected for participation. The most relevant comparison groups can come from those SMEs that are eligible but cannot participate at a given time (for example, when excess demand exists) or those near the threshold for participating in the CGS based on objective, transparent and accountable targeting and/or eligibility rules. In cases where such threshold based comparison is not possible, control firms can be carefully matched on observable characteristics.

3.2.2. CGSs' operational rules typically cover eligibility, allocation rules in light of limited resources, and the phasing in of beneficiary SMEs. More specifically, the key rules generating a roadmap to a method for identifying comparison groups relate to targeting criteria, financial capital (financial resources) and timing.

- *Targeting rules:* CGSs generally use targeting criteria that rely on a continuous indicator or cut-off point which is cheap and easy to collect. This commonly consists of limits on firm size, often defined by the number of employees, turnover amount, value of assets or their combination, and/or firm age. Moreover, to be financially sustainable and preserve their equity base, CGSs often rank eligible SMEs based on their creditworthiness, applying a credit scoring methodology or equivalent credit

analysis technique so that guarantees are only granted to creditworthy eligible SMEs.

- *Capital:* CGSs typically operate with limited financial resources and do not have enough capital to provide program services to all eligible SMEs who apply for a guarantee, even when the leverage of the CGS equity is taken into account. In that case, CGSs have to decide which of the eligible applicants is entitled to receive a guarantee and which are excluded. However, in many instances CGSs limit their operations to specific economic sectors or geographical regions, even though there may be eligible SMEs in other sectors or regions.
- *Timing:* CGSs typically phase implementation of their programs over time. Administrative and resource constraints prevent a CGS from immediately rolling out its program to every SME in its target group. Therefore, the CGS must implement its program over time, and thus it must decide who can participate in the program first and who can join later. This is typically achieved on first-come, first-served basis.

3.2.3. These three dimensions related to targeting rules, capital and timing of implementation are useful to map possible comparison groups and therefore select the most appropriate impact evaluation method. CGSs generally targets SMEs on the basis of criteria that are quantitative and available, and combine a cut-off for eligibility such as size limit with ranking for creditworthiness such as a credit score. Limited financial resources are expected to translate into excess demand for the services of the CGS. However, this is not necessarily the case when the program has not been marketed and communicated effectively to both lenders and SMEs. In any event, operational budget constraints as well as logistical and administrative limitations always result in the CGS to phase in its programs over time.

3.2.4. The above operational rules of CGSs suggest the following hierarchy of impact evaluation methods:

- Prospective evaluations
 - **RCT:** this should be the “method of choice” in cases where the CGS is still in the planning phase and has not started operations yet. When properly implemented, RCT generates comparability between the treatment and the control group in both observed and unobserved characteristics, with low risk for bias. Because this methodology is rather intuitive, requires limited econometrics and generates average treatment effects for the population of interest, it also makes communicating results to policymakers a relatively straightforward task. One option for implementing an RCT approach is to exploit the timing constraint mentioned above, with the scheme randomizing the order of SMEs granted a CGS. In this setup, those SMEs who by chance receive CGS early would be the “treatment” group, and

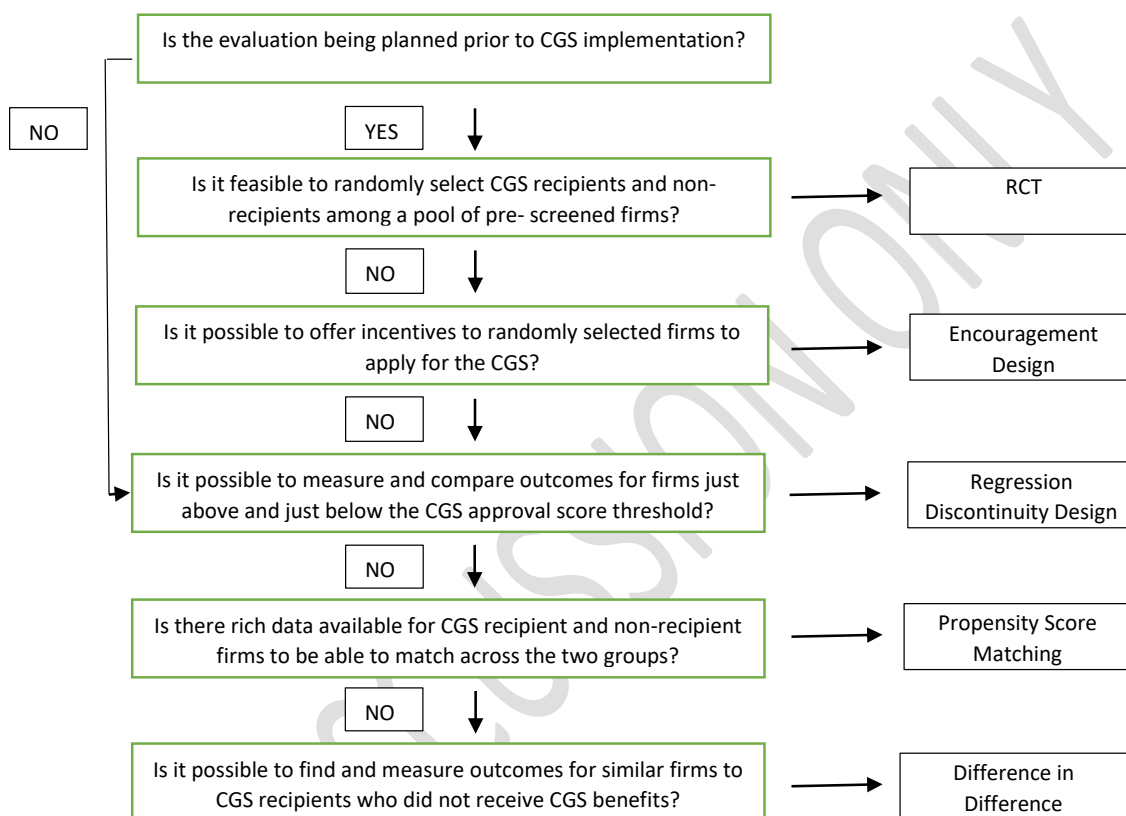
those who receive the CGS at a later date (up to a year later) would be the “control” group.

- **ED:** this has similar statistical power as the RCT and would represent an appropriate methodology where a CGS exists but is underused by firms. In such situations, a targeted advertising campaign could make information about the CGS’ programs available to a randomly selected set of SMEs. The evaluation team could then separately measure the success of the encouragement (i.e. what percentage of those encouraged actually applied to the CGS) as well as the impact of the CGS (i.e. among those SMEs encouraged by the CGS, what was the impact on their performance measures in the future based on selected impact indicators).
- Retrospective evaluations
 - **RDD:** taking advantage of any threshold-based size limit and/or credit scoring model employed by the CGSs as a continuous eligibility rule, RD design is the simplest and most universally adaptable evaluation methodology in the CGS context. In practice, firms that are within a pre-established threshold in terms of size or credit score are included in the CGS program while those that fall below are rejected. As discussed in Chapter 2, firms just above and just below this threshold are likely to be very similar in terms of their past performance measures and other pertinent characteristics. Hence, future firm performance measures could be compared for the two sets of firms, with those receiving the credit guarantee forming the “treatment” group and those just below the threshold forming the “control” group. With this threshold in mind, data would be collected for firms that fall just above the threshold (and hence avail of the CGS), as well as for firms that fall just below the same threshold. The exact sample size to be collected would depend on the percentage bandwidths around the threshold, which the evaluation team will decide in advance. For example, if the firm size limit for eligibility is turnover lower than \$1 million and the evaluation bandwidth is 20 percent, then firms with turnover up to 20 percent above the threshold (i.e. \$1.2 million) will be labelled the “treatment” group, while firms with a turnover up to 20 percent below the threshold (i.e. \$0.8 million) will form part of the “control” group. Subsequent analysis would compare financial and economic impact variables between the two groups of firms.
 - **DID/PSM:** in cases where RDD is not feasible, for example because of limited data availability in the vicinity of the cut-off eligibility point, DID or PSM evaluation methodologies can be considered. However, it is important to note that while DID and PSM techniques are valid alternatives to RDD, they require baseline data and collection of data on a very broad array of

characteristics for both treatment and control groups, which are not always efficiently available in the average country where a CGS operates.

3.2.5. Figure 4 presents the decision-making process for selecting the impact evaluation method for a CGS.

Figure 4: CGS Evaluation Methodology Decision Matrix



3.3. Data and Sampling

3.3.1. The next step in planning an impact evaluation is to determine the data needed and the sample required to estimate difference in the outcomes of interest between the treatment group and the control group. Good quality data are required for a meaningful impact evaluation. The results chain discussed in Section 2 and depicted in Figure 1 provides a basis to define the impact indicators that are directly affected by the CGS and that should be measured. However, indicators are required throughout the results chain, including intermediate impact indicators, measures of the delivery of the intervention, exogenous factors and control characteristics.

3.3.2. One of the most important aspects of data collection for impact evaluation purposes is the requirement to collect data not only for recipients of CGSs, but also non-recipients. While existing data are always needed at the outset to estimate benchmark values or to conduct power calculations, existing data are rarely sufficient. Impact

evaluations require comprehensive datasets covering a sufficiently large sample that is representative of both the treatment and the control group. Still, the possibility of using existing monitoring data for impact evaluation should be seriously considered as it can substantially reduce the cost of the impact assessment. Monitoring data are data collected and maintained by the CGS as part of its regular operations and typically recorded in a monitoring and evaluation system.

3.3.3. A potential obstacle to maintaining relevant monitoring data concerns the case where the CGS issues guarantees on a portfolio basis. In the portfolio approach, lenders are entitled to attach guarantees to loans without previous consultation with the CGS – but within eligible categories that have been clearly specified in contractual agreements between the CGS and the lender. In the portfolio approach there is, therefore, no direct relationship between the CGS and the SME borrower. This may complicate collection of monitoring data by the CGS and is therefore important that the CGS obtains all relevant data from lenders as part of the periodic reporting requirements.

3.3.4. Monitoring data can and must be complemented by administrative data or data collected and maintained by other public agencies and private actors. In many jurisdictions where CGSs operate, central banks, credit registries, credit bureaus, and specialized vendors typically maintain administrative data on firms for the purpose of monitoring credit portfolios and performance. Often, these administrative data are a rich source for detailed firm-level information on loan activity, interest rates, collateral requirements, loan maturity, default and even firm outcomes such as sales, investment, export and number of employees. Such administrative data can be a cost-effective source for evaluating the impact of CGSs.

3.3.5. This Toolkit recommends that CGSs explore options to systematically access and obtain relevant administrative data from the above entities in a standardized format on the basis, for example, of memoranda of understanding or any other relevant subscription-based option available to overcome any potential legal constraints associated with obtaining such data. These data would constitute the input to a reliable and consistent information system housed within the administration of the CGSs. One important task for such a system would be to collect standardized data at regular frequency from relevant agencies and partners to supplement the monitoring data on CGS applicants. This information system would provide a healthy source of data for a potential counter-factual group (i.e. those firms who apply for but marginally do not qualify to receive CGS benefits). Principle #16 requests that CGSs collect and retain relevant data for impact evaluation.

3.3.6. If administrative data are not available or are not sufficient, the impact evaluation will have to rely on survey data. Well-designed and implemented survey instruments are generally able to assess program impact indicators and capture all the information that is important to interpret and understand those indicators, including demographic and entrepreneurship data on preferences, attitudes, behaviours etc.

However, surveys can be very expensive and therefore this Toolkit recommends their use only when administrative data are unavailable.

3.3.7. In addition to quantitative data, qualitative data are an important supplement to impact evaluations as they can be helpful in understanding perceptions and experiences with the CGS. Qualitative data are usually collected through focus groups and interviews with selected CGS participants and other key informants. Although the views and opinions of gathered during focus groups and interviews may not be representative of the CGS's participants, they can provide insights in what is happening in the CGS and are useful to develop and test hypotheses as well as to provide context and explanations for the quantitative results. Therefore, collection of qualitative data is strongly encouraged when carrying out the impact evaluation of a CGS.

3.3.8. Size is not the only relevant factor in ensuring that a sample is appropriate for impact evaluation. The process by which the sample is drawn from the population of interest is also crucial. Selecting a sample requires a balance among cost, time, complexity, and accuracy. For quantitative research, samples can be randomly or systematically selected, and measures can be taken to ensure that certain groups are represented (stratification) or to reduce costs (cluster sampling). This topic is beyond the scope of this Toolkit and the reader can refer to the specialized literature.

3.4. Setting up the Evaluation Team

3.4.1. The credibility of the impact evaluation will depend not only on the quality of data collection, but also on the quality of data analysis. The impact evaluation of a CGS should be seen as a partnership between the CGS's main shareholder (the government), CGS management, and the evaluator. The government is expected to provide strategic guidance, the CGS's management is expected to provide operational coordination, and the evaluator will be responsible for the technical aspects. For an impact evaluation to be successful, the three parties need to work together. In other words, the process should not be divorced from the policy relevance and strategic importance of the assessment and, crucially, from the operational rules of the CGS, which are an essential factor in determining the evaluation design.

3.4.2. An important consideration relates to whether it should be an insider or an outsider who conducts the evaluation. The key argument in favour of external assessors is that they are less likely to be affected by political interference and therefore more likely to be perceived by others as independent. This independence is expected to provide more objectivity to the evaluation. In contrast, the key advantage of using internal evaluators is that they have much better knowledge of the CGS and its operations as well as of the political and policy context. Internal evaluators are more likely to gain support from CGS managers and staff. However, they are more likely to be careful and selective about their policy recommendations.

3.4.3. To manage and coordinate the evaluation process, this Toolkit recommends that CGSs establish independent evaluation units reporting directly to the board of directors. The composition of these independent evaluation units can include a mix of CGS staff, university academics and local and international consultants. The evaluation unit may decide to contract out the entire evaluation at once or only certain sub-components. Either way, the evaluation unit should be responsible for developing an evaluation plan, including engagement guidelines for evaluations, the relevant impact measures for the evaluation, the minimum data requirements and agreements regarding confidential data, the methodology and evaluation framework as well as a results framework, a clear timeline and deliverables for the project, and a budget ceiling. This plan would provide the basic terms of reference to launch a call for technical and financial proposals from external evaluators.

3.4.4. An important determination the evaluation unit has to make is whether the evaluation – or parts of it – can be implemented locally and what kind of supervision and outside assistance will be needed. Evaluation capacity varies greatly from country to country. A related resolution is whether to work with a private firm or a public agency. Private firms can be more dependable in providing timely results but this would come at the expense of building capacity in the public sector. Universities, research institutions or international organizations can also work as evaluators. The reputation and technical expertise of these partners can ensure that evaluations results are widely accepted by stakeholders.

3.5. Time and Budget

3.5.1. Another key issue when designing an evaluation is how much time is needed before results can be meaningfully measured. If the evaluation is undertaken too early there is a risk that only partial or no impact is found. If the evaluation happens too late there is a risk that the CGS might lose public or donor support, or that a badly designed and implemented program might be expanded.

3.5.2. The impact evaluation needs to be fitted to the CGS implementation cycle. The timing of data collection should take into account how much time is needed after a guarantee is granted for results to become apparent. The CGS results chains presented in section 3.1. helps identify impact indicators and the appropriate time to measure them. CGSs typically aim to provide short-term benefits to SMEs such as getting a loan and/or better terms and conditions (financial additionality) as well as longer-term gains such more investment, sales, export and jobs (economic additionality). Therefore, evaluations and data collection will have to be calibrated to the objectives of the evaluation and the impact indicators of interest. As a general guidance, this Toolkit recommends to measure a financial additionality assessment after 1-2 years and economic additionality after 2-3 years.

3.5.3. It is important to note that the production of results should also be timed to inform budgets, program expansions and other policy decisions. In other words, the timing

of an evaluation should take into account when the information is needed to inform policy making and synchronize data collection and analysis to key decision-making points.

3.5.4. Budgeting constitutes one of the last steps to operationalize the impact evaluation of the CGS. While actual impact evaluation costs depend on the country context, international experience shows that impact evaluations constitute only a small fraction of overall CGS budgets. Moreover, the costs of conducting an impact evaluation must be compared to the opportunity cost of not conducting a rigorous assessment and thus potentially running an ineffective program. Clearly, many resources are required to implement a rigorous evaluation, with budget items including staff fees for at least a principal investigator, a research assistant, a sampling expert and project staff, who may provide support throughout the evaluation. Needless to say, the larger costs in an evaluation are those related to data collection. Financing for impact evaluations can come from many sources, including the CGS itself, the government, donors, international organizations, foundations and research institutions.

3.6. Production and Dissemination

3.6.1. The main output of an impact evaluation of the CGS is the *impact evaluation report*. The main objective of the evaluation report is to present the results and provide an answer to all policy questions set out initially. The report also needs to show that the results are grounded in valid estimates of the counterfactual and that the estimated impact is directly attributable to the CGS. The evaluation report should summarize all the work connected with the evaluation and include detailed descriptions of the data and the statistical techniques employed in the analysis in addition to discussions of results and relevant tables, charts and annexes. Box 1 suggest the potential content of an impact evaluation report of a CGS.

3.6.2. In addition to the comprehensive evaluation report, the evaluation team should produce one or more shorter policy briefs to help communicate the results to policymakers and stakeholders. A policy brief focuses on presenting the core findings of the evaluation through graphs, charts and other accessible formats, and on discussing the policy recommendations. It also includes a short summary of the technical aspects of the evaluation.

3.6.3. Beyond producing evaluation results, the ultimate objective of CGS impact evaluation is to make SME finance policies more effective and improve the intended development outcomes. Therefore, to ensure that the evaluation results effectively inform policy decisions, it is essential that the evaluation outputs are disseminated through a well-thought dissemination plan that outlines how key stakeholders will be kept informed and engaged throughout the evaluation cycle.

Box 1: Outline of an Evaluation Report of CGS

1. Introduction
2. Description of the CGS (benefits, eligibility rules and so on)
 - 2.1. Design
 - 2.2. Implementation
3. Objectives of the evaluation
 - 3.1. Hypotheses, theory of change, results chain
 - 3.2. Policy questions
 - 3.3. Key impact indicators (financial additionality, economic additionality or both)
4. Evaluation design
 - 4.1. Theory
 - 4.2. Practice
5. Sampling and data
 - 5.1. Sampling strategy
 - 5.2. Data collected
6. Validation of evaluation design
7. Results
8. Conclusion and policy recommendations

FOR DISCUSSION ONLY

References

Armentos, M., Artellini N., Hoppe A., Urizar, M., Yaros, B. (2015): “Rethinking FOGAPE. Evaluating Chile’s Partial Credit Guarantee Program. Mimeo.

Arráiz, I., Meléndez, M., Stucchi, R. (2014): “Partial credit guarantees and firm performance: evidence from Colombia”, *Small Business Economics*, 43 (3): 711-724.

Asdrubali P., Signore S. (2015): “The Economic Impact of EU Guarantees on Credit to SMEs”. *European Economy Discussion Paper 2*, July 2015.

Boschi, M., Girardi, A., Ventura, M. (2014): “Partial Credit Guarantees and SME Financing”. *Journal of Financial Stability*, 15: 182-194.

Brown J. D., Earle J. S. (2015): “Finance and Growth at the Firm Level: Evidence from SBA Loans”. *Journal of Finance*, 72 (3): 1039–1080.

Cowan, K., Drexler, A., Yañez, Á. (2015): “The Effect of Credit Guarantees on Credit Availability and Delinquency”. *Journal of Banking and Finance*, 59: 98-110.

De Blasio G., De Mitri S., D'Ignazio A., Finaldi Russo P., Stoppani L. (2017): “Public Guarantees to SME Borrowing in Italy. An RDD Evaluation”. *Bank of Italy Temi di Discussione (Working Paper) No. 1111*.

De Blasio G., De Mitri S., D'Ignazio A., Finaldi Russo P., Stoppani L. (2015): “Public Guarantees to SME Borrowing in Italy. An RDD Evaluation”. *Working paper*

D'Ignazio A., Menon C. (2013): “The causal effect of credit guarantees for SMEs: evidence from Italy”. *Bank of Italy Working Papers*, n. 900.

Lelarge C., Sraer D., Thesmar D. (2010): “Entrepreneurship and Credit Constraints: Evidence from a French Loan Guarantee Program”. Chapter in *NBER book: International Differences in Entrepreneurship*, Josh Lerner and Antoinette Schoar editors.

Oh, I., Oh, I., Lee, J.D., Heshmati, A. (2009): “Evaluation of Credit Guarantee Policy Using Propensity Score Matching”. *Small Business Economics*, 33 (3): 335-351.

Ono A., Uesugi I., Yasuda Y. (2013): “Are lending relationships beneficial or harmful for public credit guarantees? Evidence from Japan's Emergency Credit Guarantee Program”. *Journal of Financial stability* Vol 9 pag 151-167

Uesugi I., Sakai K., Yamashiro G. M. (2010): “The effectiveness of Public Credit Guarantees in the Japanese Loan Market”. *Journal of the Japanese and International Market*, Vol. 24, p. 457-480.

World Bank, 2014. Global Financial Development Report: Financial Inclusion. Washington DC.

World Bank and FIRST Initiative, 2015. Principles for the Design, Implementation and Evaluation of Public Credit Guarantee Schemes for SMEs. Washington DC.

FOR DISCUSSION ONLY