Promoting High Impact Entrepreneurship in Mexico
Concept Note

First version: April 2016
This Version: May 2017

Abstract
The High Impact Entrepreneurship Program aims to provide support to Mexican SMEs through a matching-grant scheme to support their development and increase their ability to succeed. The targeted firms are start-ups and more-established but still young (“scale-up”) firms that offer an innovative product, service or business model with high potential to compete globally and generate high impact in economic, social and environmental outcomes. Through a randomized control trial, we will investigate the overall impact of the program on firms’ performance as measured by outcomes such as productivity, sales, job creation and innovation. Of equal importance, we will also randomize across two different selection panels to select beneficiaries, one panel that has no reason to be biased but does not have industry experience, and one panel that is potentially biased but has industry experience. This will contribute to a better understanding of how to implement matching-grant programs by improving governments’ selection criteria, procedures and targeting strategy.

1 This Concept Note was prepared by David Atkin (Assistant Professor, MIT. atkin@mit.edu), Leonardo Iacovone (Senior Economist, World Bank Group, liacovone@worldbank.org), Alejandra Mendoza Alcántara (Impact Evaluation Specialist, World Bank Group. amendoza1@worldbank.org), and Eric Verhoogen (Associate Professor of Economics and International Affairs, Columbia University. eric.verhoogen@columbia.edu).
# Table of Contents

Abstract

Table of Contents

1. Background

2. Intervention to be Evaluated

3. Theory of Change

4. Literature Review

5. Hypotheses and Evaluation Questions

6. Evaluation Design and Sampling Strategy

7. Data Collection
   - 7.1 Quantitative Instruments
   - 7.2 Management of Data Quality
   - 7.3 Ethical Issues
   - 7.4 Qualitative Instruments
   - 7.5 IE Implementation Monitoring System

8. Data Processing and Analysis

9. Study Limitations and Risks

10. Evaluation Team

11. Milestones, Deliverables, and Timeline

12. References
1. Background

One of the key objectives of the current Mexican government is to enhance innovation and growth from the small and medium enterprise (SME) sector. According to the OECD, Mexico lags behind other countries at a similar level of technological and scientific development in the creation and growth of knowledge-based SMEs (OECD, 2013). Weak financial markets, regulatory barriers and inefficient support programs are among the main culprits identified. For example, venture capital (VC) investments play an increasingly important role in financing knowledge-based industries. Unfortunately, Mexico has very low levels of VC investments compared to other emerging economies: In China and India venture capital investments represent .15% to .20% of GDP; Latin American countries such as Brazil and Uruguay are closer to the .07% while Mexico is barely .01% (IADB, 2013). While the first-best solution may be to establish a vibrant financial sector that can lend to SMEs and support a VC industry in Mexico, government matching grants can act as second-best solution to restricted access to finance to SMEs. In addition, in the presence of learning spillovers from new technology development or entering new product segments, investments in innovation by private firms are likely to be below what would be optimal for the society as a whole (e.g. Hausman and Rodrik, 2003); this provides a further justification for a matching-grant program.

The High Impact Entrepreneurship Program (HIEP), run by the Instituto Nacional del Emprendedor (INADEM) [National Institute of the Entrepreneur] in the Mexican Ministry of the Economy, aims to provide support to innovative Mexican SMEs through a matching-grant scheme to support their development and increase their capabilities to succeed. The program’s main objective is to promote innovation to boost productivity and employment growth in the SME sector. The HIEP aims to target firms with either high growth impact potential or high social and environmental impact. Firms are required to offer an innovative product or service, or an innovative component in their business model.

However, after three years of running the program, there is concern whether the program is achieving the expected results, whether the design of the program is the most effective to maximize its impact, whether firms with high impact potential are being properly identified and whether this type of entrepreneurship needs government’s assistance to grow. Therefore the World Bank is providing technical assistance to INADEM to evaluate the impact of the program and test firm selection mechanisms that may improve the targeting and therefore the effectiveness of the program.

In January 2016, the Board of Directors of INADEM – made up of diverse government representatives – approved the integration of the impact evaluation into the design of the HIEP program. The use of alternative selection mechanisms and randomization of the matching grants was accepted and has been published in the Official Gazette of the Mexican Government. The program will be launched in June 2017.
High Impact Entrepreneurship

This program focuses on high impact firms. There is no consensus on the definition of high impact firms. For example, the OECD defines them as firms “with average annualized growth greater than 20% per annum, over a three-year period, and with ten or more employees at the beginning of the observation period”. Gazelles are defined as firms up to five year old that comply with the characteristics mentioned above. However studies find that both metrics—employment and sales growth—depend on the sector.2 On the other hand, the U.S. Small Business Administration (SBA) defines high impact firms as “enterprises whose sales have at least doubled over a four-year period and which have an employment growth quantifier of two or more over the same period.”3

The great interest in High Impact Firms derives from their large economic impacts and contribution to employment. For example, a study found (Tracy, 2011) that high impact firms4 in USA, despite only representing around 5-7% of all businesses, were responsible for almost all the net job growth in the US between 1994-2008. Without these firms, the study estimates that the US economy would have lost about 16.3 million jobs. A more recent report (World Economic Forum, 2014), found that a sample of high impact firms5, representing 0.01% of American firms, created over 5% of all jobs created in a period of two years. More interestingly, these firms created wealth not only for their owners but made a significant contribution to society through the identification of market niches or generating innovative solutions to market failures.

INADEM defines High Impact Entrepreneurship as firms that have high economic growth potential and potential to create high-value jobs based on new or improved solutions that translate into new/improved products, services or business models with a social and environmental impact. The targeted firms are SMEs with less than two years old (startups) and SMEs older than two years (scaleups).

Diagnostic fieldwork was conducted in order to align the selection criteria according to INADEM’s definition:

- Four focus groups were conducted with startups and scale up firms that received the grant in previous years.
- More than 10 interviews with VC funds and recognized organizations in the entrepreneur ecosystem.
- Three focus groups with incubators/accelerators, innovation and project evaluation specialists.

Relevant findings:

2 Cella and Morrone, 2008; Moreno and Casillas 2007.
3 Tracy, 2011.
4 Defined as firms with high growth of sales, and absolute and percent change of employment.
5 Defined as firms that grow fast, create jobs, contribute to society and transforms industries.
• There was consensus that a fixed definition of high growth -based on sales and employment- was inaccurate given that the differential rates heavily depend on the sector.

• A common finding across interviews and focus groups is that, for firms at early stages (startups) it is very difficult to assess the potential impact of a project, therefore most of the weight in the decision to invest on a firm or not relies on the entrepreneur profile/personality, and team composition (multidisciplinary). For scale-up firms it is easier to determine the growth potential given that the product/service has been validated in the market and there is evidence of the firm behavior.

• Innovation is a critical component; either innovation is disruptive or incremental. Although, ownership of the technology and flexibility to adapt are key to ensure firm survival and growth (entry barrier for competitors).

As a result of the diagnostic fieldwork and literature review, four main criteria will be assessed for firm selection:

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Examples of components being measured</th>
</tr>
</thead>
</table>
| Entrepreneur's experience, team and strategic alliances | Entrepreneur’s profile, and communication skills  
Team profile  
Objective of the project |
| Technical, Financial and Business Viability | Statements and projections  
Technical resources, applicable regulations,  
Identification of competition, entry barriers,  
validation of product |
| Innovation                           | Potential to scale-up the business model  
Innovation |
| Value added of project's high impact proposal | Type of impacts (economic, social, environmental),  
Justification of impacts and feasibility, |
2. Intervention to be Evaluated

The impact evaluation will serve two purposes. First, it will allow an assessment of the existing approach to selecting firms for matching grants, which we refer to as the “Traditional Panel.” Second, the evaluation will allow a comparison of the Traditional Panel with a new, alternative review process carried out by a panel of experts (highly experienced entrepreneurs, investors and experts in the sector) simulating a venture-capital approach; we refer to this panel as the “VC panel”. The VC panel has a subgroup composed of volunteer industry experts, and another of paid consultants from accounting firms that are experts in business. There are several reasons to believe that these subgroups may be different from each other: first, the latter has different incentives than the volunteer experts; second, the consulting firms for which the paid consultants work have a reputation to uphold and may monitor the work of the consultants; finally, the paid consultants may have fewer network links in particular industries than the volunteer experts and therefore may be less likely to be biased but at the same time may be less informed about the viability of particular proposals. Approximately 150-200 entrepreneurs/experts will be needed to review the applications.

INADEM evaluators (“Traditional Panel”) currently review the projects of all INADEM’s programs (from the basic SME funding initiatives to the High Impact Entrepreneurship Program). These evaluators have a graduate degree in social sciences and took a specialized course on cost-benefit analysis. Some of them work for INADEM full time; others hold other positions outside INADEM. However, these evaluators do not typically have entrepreneurial experience, nor are they experts on the sectors that are evaluated under this program (e.g. IT, biotechnology, financial services, etc.). Evidence from the field and feedback from INADEM’s staff suggested that the main weakness of the program has been the selection process. Essentially, the traditional panel was established to reduce any potential bias in the evaluation process. This bias would come from local experts picking applications of their friends or applications that would benefit them in other ways. By systematically recruiting evaluators without specific industry expertise, the hope was that they would not know the firms in question or be in a position to benefit. However, this comes at the cost of the panel being poorly informed about the quality of the applications, potentially a major issue for a program that aims to promote knowledge-based and innovative industries.

Under our design, both panels, the VC Panel and the Traditional Panel, will review all applications. The review process will be the same for both panels as per INADEM’s scoring rules: projects will be randomly assigned to two reviewers; if grades differ by more than 15 points, a randomly selected third reviewer will evaluate the project. If there are two reviewers, the scores will be averaged. If three reviewers, the mean of the two closest scores will be taken into account.

For the VC panel, the first review will be conducted by volunteer experts of the same industry of the applying firm, the second review will be conducted by the paid business consultants, and in case of a third review, it will be conducted by a randomly assigned paid or volunteer expert.
The High Impact Entrepreneurship program will provide funding to “start-up” and “scale-up” firms (up to 3 million pesos, ~$170,000 USD) with high impact potential in the following investment categories: a) IT and software, b) certifications, c) consulting/professional services, d) machinery and equipment. Depending on the expenditure category, INADEM will grant 70% to 80% of the amount requested, while the entrepreneur commits to finance the remaining 20% to 30%.

The funding allocated for the HIEP this year sums up to a total of 400 million pesos (22 million US dollars) that may fund 150-200 firms depending on the amount requested per firm.

Grants will be randomly assigned among firms with scores above a certain threshold.

3. Theory of Change

The basic theory of change for matching grants is that there are financial frictions preventing private investments with high impacts being made. If this is true, then a matching-grant program can alleviate these financial frictions by providing capital for profitable investments, and loosen a bottleneck to firm growth. There are a variety of firm growth outcomes that we may observe changing: investment, innovation, employment, revenue and sales growth, new product entry, skill upgrading.

A second justification for matching grants is the presence of spillovers. In the absence of government intervention, private investment will be suboptimal: one firm’s innovation benefits other firms and these benefits cannot be appropriated by the original innovator. This theory predicts that matching grants may improve firm outcomes beyond the firm that receives the grant. Therefore, we plan to look at outcomes for other firms in the same industry and/or location.

Finally, we have a theory of change for the evaluation panel itself. Given that traditional evaluators do not have an industry background, they are unlikely to have links with the applying firms and hence unlikely to be biased towards favoring firms based on these links. However, their lack of industry background also means they do not have the expertise to accurately judge the quality of the applications. In contrast, the VC panel should be better able to judge quality but may also be potentially biased toward favoring firms panel members have links with. Therefore, we expect higher treatment effects (a measure of quality) for the firms selected by the VC panel, but also a relationship between firms selected and network links after conditioning on quality.
Figure I - Theory of Change for Panel Selection

Output

SCORED APPLICATIONS

Outcome

SELECTION OF FIRMS

Impact

FIRM’S PERFORMANCE

Smaller treatment effects

TRADITIONAL PANEL

SCORING WITH INFORMATION ASYMMETRY
No links with applicants
No industry knowledge

Pool of eligible firms - random quality

PANEL OF EXPERTS

SCORING WITHOUT INFORMATION ASYMMETRY
Links with applicants
Industry knowledge

Pool of eligible firms – high quality

Pool of eligible firms with strong networks with experts

Larger/Smaller treatment effects

Figure II - Theory of Change for Matching Grants

Output

Expenditure increases in categories stated in application:

- Certification
- Consulting/Professional Services
- Machinery & Equipment
- IT and Software

Mid- term Outcome

Improved skills

Improved business practices

Improved Efficiency

Technology adoption

Development of new product/service

Impact

Labor productivity

Access to market (national/international)

Employment creation

Sales, Revenues, Profits

Next stage funding round

Innovative product/service or business model operational/in the market

Spillover effects:
- Improved firm’s performance
- Same industry
- Same location
- Networks
- Others (TBD)

Association link - -
The table below lists and defines the main outcomes, assumptions and threats. These are based on initial evidence from fieldwork and available literature (see literature review below).

<table>
<thead>
<tr>
<th>Inputs, outputs, outcomes</th>
<th>Objectives</th>
<th>Indicators (preliminary)</th>
<th>Assumptions or threats</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inputs</strong></td>
<td>To provide finance to high impact entrepreneurs</td>
<td>Amount of money provided to firm and concept of funding</td>
<td>-Money is effectively disbursed.</td>
</tr>
<tr>
<td><strong>Outputs</strong></td>
<td>To increase expenditure in key areas for firm’s development</td>
<td>Amount of money spent in each concept to which firm applied for funding (self-reported)</td>
<td>-Firms effectively use the money on those concepts.</td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td>Improved Skills and business practices</td>
<td>-Assessment of financial and HR practices -Number of certifications</td>
<td>-Improved business practices may vary i.e. from financial or HR management to pricing strategy.</td>
</tr>
<tr>
<td></td>
<td>Firm’s efficiency</td>
<td>- TFP calculations - output per hour</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Technology adoption Increase in R&amp;D New product/service</td>
<td>- Adoption/development of new systems -Increase in expenditure in R&amp;D -Share of employees dedicated to R&amp;D -Minimum viable product or service available/tested</td>
<td></td>
</tr>
<tr>
<td><strong>Impact</strong></td>
<td>Improve firm’s performance outcomes and innovation activity</td>
<td>-Sales (variance and levels) -Revenues (variance and levels) -Labor and capital productivity -Value of fixed assets</td>
<td>-Growth rates may depend on age of firm6</td>
</tr>
</tbody>
</table>

6 See Lawless, 2014
4. Literature Review

Several non-experimental evaluations aim to provide evidence of the impact of matching grants on firm’s performance outcomes such as productivity, employment and exports. For example, Gourdon et. al. (2011) found that a matching-grant program for exporting SMEs did not have long-term effects on level of exports. An innovation program in Colombia showed positive effects on product diversification and labor productivity (Crespi, et. al. 2011). A matching grant to receive technical assistance for SMEs had positive effects on employment and wages (Castillo, et. al. 2011). However, these results should be taken with caution given that these studies have several well-known weaknesses, most critically selection bias since better firms typically get the grants.

Very few experimental evaluations have assessed the impact of matching grants. Bruhn, et. al. (2013) found positive impacts of subsidized consulting services for SMEs in Mexico on sales, profits and productivity but not significant impacts on employment. An experiment in Yemen where SMEs received up to $10,000 as a 50 percent subsidy towards the cost of business services (i.e. finance, accounting systems, IT equipment and marketing) had a positive impact on product innovation, marketing and adoption of accounting systems, pointing out that the grants helped firms to undertake additional new activities and investments. However, researchers were not able to assess long-term impacts due to civil conflict in the country (McKenzie, et. al. 2015). Consequently, more evidence is needed on the impact of matching grants on firm growth and new investments, especially for SMEs that have high growth potential. Furthermore, there is a knowledge gap on whether this type of firms need government assistance due to market constraints or if this type of grants “crowds out” private investment or complements it.

Cash transfer programs can shed some light on the impact of external capital shocks on firm performance. An experiment in Nigeria targeting fast growing firms---by providing cash awards to SMEs

---

7 Campos et. al. (2014) explains that political pressures and other implementation difficulties are the reason why many World Bank-funded evaluations have been derailed.
that won a business plan competition—found positive impacts on survival rates, employment, sales and profits (McKenzie 2015). Other evidence, mainly focused on manufacturing microenterprises, is not conclusive regarding whether cash transfers have positive and sustained effects on firm’s growth.\textsuperscript{8} Additionally, although several studies relate outcomes to metrics that are ex-ante observable, e.g. the age of the firm or the entrepreneur’s experience, there is little consensus on how a large scale program can identify and target firms with high growth potential.\textsuperscript{9}

Few studies explore the issue of targeting. Fafchamps et. al. (2014) tests whether a panel of experts or a survey instrument is more effective at identifying fast growing firms. They find that growth is strongly correlated with measures of ability, although experts add predictive power in identifying the high growth entrepreneurs after controlling for measures of ability. That said, survey measures still explain more of the variance in outcomes than the VC panel’s recommendations do.

The trade-off between potentially biased experts and unbiased non-experts has not been explored in the context of industrial grants, but has been explored in academic settings. Zinovyeva et al. (2015) explore the role of connections in academic promotions finding that applicants with prior connections to the evaluator have a much higher likelihood of being promoted. This is problematic as candidates with weaker links have higher-quality research and are more productive than those with stronger links. Li (2015) finds that evaluators are biased in favor of research projects in their own area of expertise, but also have better information about the quality of those projects. However, the benefits of expertise dominate the cost of bias: if applications were not judged by intellectually-proximate evaluators, the quality of successful applications (as measured by subsequent publications) would be lower. Similarly, Bodreau et. al. (2012) finds a negative bias from experts evaluating novel research projects. But the evidence points out that this effect is smaller when there is an intellectual relationship with their own area of expertise, suggesting that the bias is inherent to information asymmetries rather than “conflict of interest or strategic incentives in the evaluation process.” These finding suggest that relationships improve information asymmetries and that more restricted rules regarding conflict of interest can have negative effects on quality of projects funded.

\textsuperscript{8} See Karlan et. al. (2014) and Mckenzie et. al. 2008

\textsuperscript{9} See: Fafchamps, et. al. 2014; Schoar, et. al. 2010; De Mel, et. al. 2010; Wadhwa, et.al. 2009.
5. Hypotheses and Evaluation Questions

Primary questions:

1. Which evaluation/selection model is most effective at identifying high-impact entrepreneurs who will benefit most from the matching-grant program (i.e. firms with large treatment effects from the program)?

2. What is the impact of matching grants aimed at high-impact entrepreneurs on firms’ performance (productivity, sales, job creation) and on innovation?
   a. How heterogeneous are the outcomes depending on initial firm characteristics?

Hypotheses:

a) A panel of experts will be more effective in identifying high impact-entrepreneurs (i.e. firms with large treatment effects), but they will also be more likely to choose firms that they have network links to.

b) Firms are capital constrained in Mexico, and so firms that receive the grant will increase their investment in productive activities, will be more likely to grow (increasing sales, profits and employment) and enhance innovation (launch of the new product, service or business model).

6. Evaluation Design and Sampling Strategy

The evaluation has two key elements. First there will be two different selection panels. By comparing outcomes with the scores of the two panels (and its subgroups for the VC panel), we hope to understand which selection mechanisms best identify high-impact entrepreneurs. Second, not all eligible firms will receive grants, allowing us to tease out the impact of matching grants on firm performance.

The evaluation design is as follows. According to the application rate of previous years, it is expected that the HIEP will receive approximately 1,000 applications. The first screening -- a documents and requirements checklist -- usually reduces the number of applications by 30%. Both the Traditional Panel and VC Panel will review all of the remaining 700 applications. Firms will be scored using the existing score sheets that score firms between 1 and 100.

10 Sections 5, 6, 7, 8, and 9, related to the IE design, have been taken from DIME/I2i Concept Note template to ensure comparability across concept notes.
As discussed above, the review process will be the same for both panels as dictated by INADEM’s scoring rules: initially, projects will be randomly assigned to two reviewers; if grades differ by more than 15 points, a randomly selected third reviewer will evaluate the project. If there are two reviewers, the two scores will be averaged (after the rescaling described below). If three reviewers, the two closest scores will be averaged (after the rescaling described below).

For the VC panel, the first review will be conducted by volunteer experts of the same industry of the applying firm, the second review will be conducted by the paid business consultants, and in case of a third review, it will be conducted by a randomly assigned paid or volunteer expert.

In order to ensure fairness across panels, we will rescale the evaluator scores. In the rescaling process, we will consider the two subgroups of the VC panel, the volunteer and paid experts, separately. Comparing the three groups (Traditional Panel and two VC subgroups), the scores of the two less-generous groups will be shifted upwards (simply by adding points, with a cap at 100), so that all three groups will have the same number of applications above a particular cut-off, where the cut-off will be chosen to identify approximately 400 eligible firms (using the eligibility criteria below). The rescaled scores will then be used to calculate the two averages described above, one for the Traditional and one for the VC Panel.

Finally, the averaged scores will be converted from a 1 to 100 scale to a 1 to 5 scale, with 5 corresponding to a score above the cut-off mentioned above. Applicants that receive a score of 5 from either the Traditional Panel or the VC Panel or both will be considered “eligible.” Given the above procedure, approximately 400 firms will be eligible. Since the program only has funds for approximately 200 firms, the program will be oversubscribed by eligible firms. We will then randomize among the eligible firms.

To increase our ability to distinguish between the different panels and subgroups, we will stratify the randomization according to the following eight groups, which exhaust the set of eligible firms. As for the rescaling, we consider the paid and volunteer subgroups of the VC Panel separately. In determining the strata, if more than one score of a particular type (Traditional, paid, volunteer subgroups) is available, we use the average of the scores for the type.

Group 1: Traditional score = 5; paid, volunteer scores < 5.
Group 2: volunteer score = 5; Traditional, paid scores < 5.
Group 3: paid score = 5; Traditional, volunteer scores < 5.
Group 4: paid, volunteer scores = 5; Traditional score < 5.
Group 5: Traditional, volunteer scores = 5; paid score < 5.
Group 6: Traditional, paid scores = 5; volunteer score < 5.
Group 7: Traditional, paid, volunteer scores = 5.
Group 8: Traditional, paid, volunteer scores < 5. (Note, this is likely to be a small or even empty group, but it is a possibility given that eligibility is based on an average of the two closest paid and volunteer scores and/or two closest traditional scores, while group allocations are based on average scores within groups.)

The figure illustrates these groups.

Eligibility by group & subgroup

Within each stratum, firms will be picked randomly. The number of firms selected will be proportional to the size of the stratum. The selected firms we refer to as the treatment group. The non-selected eligible firms will be our control group.

Comparing firm outcomes between treatment and control tells us about the impact of matching grants on firm outcomes. Interacting treatment with the initial average score (as well as other observable firm characteristics) allows us to explore heterogeneity in this treatment effect. Comparing treatment effects between Group 1, Group 2 and Group 3 shows which panel’s choices have better outcomes. Finally, we will also explore spillovers from receiving a grant on connected firms (both in our sample, and in other Mexican firm samples).
This experimental design will allow us to evaluate whether accuracy and bias come from being in the same industry as the applicant firm. The main outcomes to be measured are related to firm’s performance: firm survival, employment growth, labor productivity, revenue and profits, and investment attraction. In addition, indicators to assess innovation will be explored. We will assess the “accuracy” of the scores by evaluator types based on estimated treatment effects, as well as other growth metrics. We will assess bias by the strength of network links to applicants.

To reduce levels of attrition, INADEM will demand that firms in the treatment group answer all our surveys. Failure to do so will require reimbursement of the grant. For the control group, firms when applying to the HIEP will commit to answer the surveys as a condition to access the HIEP or other INADEM’s programs in the future, regardless of the outcome this year. The team will explore other positive incentives to reduce attrition in control group.

**Power Calculations**

Our ability to do power calculations is severely constrained by data availability. INADEM has only made available sales and employment data for firms mainly based in Mexico City that applied in previous years. Hence we do not have data for many other outcome variables; nor are we able to carry out power calculations related to the comparisons across panels.

Assuming that 200 firms will be treated, we obtained a minimum detectable effect (MDE) of 23% increase in employment with 80% power and 95% level of confidence. The minimum target of the program set by the government is to achieve at least a 50% increase in the employment rate. Even if only 100 firms are treated, our MDE rise to 32%, still below the government target. For sales, a minimum increase of 24% will be needed to detect an impact with a sample of 200. We have no evidence of the magnitude of the effects of similar matching-grant programs, but the large cash transfer program in McKenzie et. al. (2015) showed 140% increases in the employment rate.

<table>
<thead>
<tr>
<th>Sample Size for Each Group=200</th>
<th>Sample Size for Each Group=100</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Employment</strong></td>
<td><strong>Sales</strong></td>
</tr>
<tr>
<td>Power</td>
<td>Power</td>
</tr>
<tr>
<td>80%</td>
<td>90%</td>
</tr>
<tr>
<td>MDE</td>
<td></td>
</tr>
<tr>
<td>4.52</td>
<td>1,346,800</td>
</tr>
<tr>
<td>1.26</td>
<td>377,104</td>
</tr>
<tr>
<td>0.28</td>
<td>1,597,281</td>
</tr>
<tr>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td>23%</td>
<td>24%</td>
</tr>
<tr>
<td>Implied % Change in Outcome</td>
<td></td>
</tr>
<tr>
<td>23%</td>
<td>26%</td>
</tr>
<tr>
<td><strong>Sales</strong></td>
<td><strong>Employment</strong></td>
</tr>
<tr>
<td>Power</td>
<td>Power</td>
</tr>
<tr>
<td>80%</td>
<td>90%</td>
</tr>
<tr>
<td>MDE</td>
<td></td>
</tr>
<tr>
<td>4.52</td>
<td>1,346,800</td>
</tr>
<tr>
<td>1.46</td>
<td>436,363</td>
</tr>
<tr>
<td>0.32</td>
<td>1,597,281</td>
</tr>
<tr>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td>27%</td>
<td>28%</td>
</tr>
<tr>
<td>Implied % Change in Outcome</td>
<td></td>
</tr>
<tr>
<td>27%</td>
<td>32%</td>
</tr>
<tr>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td>32%</td>
<td>37%</td>
</tr>
<tr>
<td>Implied % Change in Outcome</td>
<td></td>
</tr>
<tr>
<td>32%</td>
<td>37%</td>
</tr>
<tr>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td>33%</td>
<td>39%</td>
</tr>
</tbody>
</table>

11 Source: INADEM & USAID: High Impact Entrepreneurs Report
We also intend to compare the performance of the VC and traditional panels (as well as comparisons between the two types of VC panel; paid and volunteer experts). Much of the evaluation where we consider different panels is not carried out using a standard treatment effect approach (e.g. see equations 2 and 3 in Section 8 where we explore how scores relate to quality and proximity). In particular, we will rely heavily on raw scores and variation in the network measures (with our experimental design ensuring there is sufficient variation on these dimensions). That said, we can add power calculations for a simple evaluation of whether different panels have higher treatment effects (i.e. equation 1 of Section 8 but with different panels having different treatment dummies). In particular, we now report power calculations to see how large the difference in treatment effects will have to be for us to detect differences between groups in the following two comparisons (assuming that firms are distributed proportionally across the three groups). To do this, we simulate draws of sales and employment assuming they are distributed log normally (using the mean and variance of unlogged sales and employment above), run a regression of the outcome on treatment allowing for treatment effects to vary by group, and test the null hypothesis that treatment effects are equal across groups. We do this 1000 times for any particular difference in effect sizes.

(1) Comparing firms only picked by the VC panel to firms only picked by the traditional panel: For sales, a difference in the treatment effect between comparison groups of at least 70% is needed in order to reject the null that they are equal at the 5% level in at least 80% of the iterations. For employment the difference is around 75%. (We assume that there is no treatment effect in the traditional panel).

(2) Comparing firms picked by both panels to firms picked by just one of the panels: For sales, a difference in the treatment effect between comparison groups of a little above 60% is needed in order to reject the null that they are equal at the 5% level in at least 80% of the iterations. For employment the difference is a little below 70%. (We assume that there is no treatment effect in the group picked by just one of the two panels).

While these magnitudes are sizeable, they are not improbably large given the 140% effect size in McKenzie et. al. (2015). Unfortunately, we have no similar studies we can refer to in order to check if these differences in treatment effects across panels are reasonable. We also note that the finer grained actual scores will provide more power with which to compare panels.

The summary of the simulations is shown below:

<table>
<thead>
<tr>
<th>Effect Size</th>
<th>Sales (Comparison 1)</th>
<th>Employment (Comparison 1)</th>
<th>Sales (Comparison 2)</th>
<th>Employment (Comparison 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>0.489</td>
<td>0.451</td>
<td>0.586</td>
<td>0.579</td>
</tr>
<tr>
<td>60%</td>
<td>0.648</td>
<td>0.594</td>
<td>0.776</td>
<td>0.718</td>
</tr>
<tr>
<td>70%</td>
<td>0.796</td>
<td>0.748</td>
<td>0.885</td>
<td>0.855</td>
</tr>
<tr>
<td>80%</td>
<td>0.878</td>
<td>0.855</td>
<td>0.946</td>
<td>0.938</td>
</tr>
<tr>
<td>90%</td>
<td>0.944</td>
<td>0.893</td>
<td>0.982</td>
<td>0.964</td>
</tr>
<tr>
<td>100%</td>
<td>0.966</td>
<td>0.963</td>
<td>0.986</td>
<td>0.99</td>
</tr>
</tbody>
</table>
Figure III: Selection Process

Step 1: First screening (paperwork requirements)
- Traditional panel
- Volunteer experts
- Paid experts

First screening:
- ~30% do not comply with these requirements.

0-100 scores re-scaled to 1-5 scores

Step 2: Second screening (VC panel)
- Pool of eligible firms: ~400 top firms that get a score of 5 by at least one panel

Step 3: Assignment to groups
- Only traditional panel
- Firms selected only by VC panel
- Firms selected by both

Assign to these 3 groups

Step 4: Objective: 150-200 Beneficiary firms
- Eligible firms will be within these 8 subgroups for reassessment purposes
- Group 1- score of 5 only by traditional panel
- Group 2- score of 5 only by volunteer experts
- Group 3- score of 5 only by paid experts
- Group 4- score of 5 by both volunteer and paid experts
- Group 5- score of 5 by traditional and volunteer experts
- Group 6- score of 5 by traditional and paid experts
- Group 7- score of 5 by traditional, volunteer and paid experts
- Group 8*- score of <5 by traditional, volunteer and paid experts

Figure IV: Evaluation Design: Overview
7. Data Collection

The evaluation will rely on different sources of data:

**Administrative data:** INADEM will collect baseline and final data as part of the launching and monitoring activities of the program. In addition, INADEM is exploring collaboration opportunities with the National Institute of Statistics (INEGI) in order to closely track benefited firms to monitor firm’s compliance with the goals stated in their projects. This data will be collected from November 2017 to March 2020. These include indicators such as: jobs retained, job creation, revenue, profits, sales, value of capital and firm’s characteristics.

**Firm survey data:** Three data collection exercises will be conducted:
Baseline: this will be conducted after the two panels reviewed all applications and eligible firms have been identified (~November 2017)
First follow up- (~March 2019)
Second follow up-Two years after funding was received (~March 2020)
The surveys will capture entrepreneur’s profile, firm characteristics and firm performance.

In order to track connected firms to look for spillovers we will also collect data on network links between firms, and seek access to other government surveys where we can match connected firms based on industry, location, ownership and even potentially employee flows.

**Evaluator survey data:** In addition, we will survey all evaluators. The survey will take two forms. A one-time survey will collect information regarding the evaluator’s characteristics, experience, background etc. A second form will be attached to each proposal the evaluator evaluates, where the evaluator must report network links to the firm in question (previous business relationships, personal friendships etc.). Together these data will allow us to create network proximity measures between evaluators and applicants based on: geographical location, membership of associations, industry experience, financial dealings with applicants, friendships, educational institutions, etc.. In addition, we plan to elicit summary (non-binding) evaluation scores along three dimensions: (1) which firms the evaluators think will benefit most from receiving the grant, (2) which firms the evaluators think are the best, and (3) which projects the evaluators think are the best. These three different evaluation measures will allow us to explore what dimensions evaluators care most about when ranking projects, and how these different measures map to later outcomes. In addition, the scoring system used by INADEM is calculated by summing scores along many dimensions (see the Criteria table in Section 1). Comparing firm outcomes to scores across criteria will also help us understand how to target high growth businesses more effectively.
7.1 Quantitative Instruments

<table>
<thead>
<tr>
<th>Outputs, outcomes</th>
<th>Indicators (preliminary)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outputs</td>
<td>Amount of money spent in each concept (self-reported)</td>
<td>Biannual</td>
</tr>
<tr>
<td>Outcomes</td>
<td>-Assessment of financial and HR practices</td>
<td>Annual</td>
</tr>
<tr>
<td></td>
<td>-Number of certifications</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Efficiency</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Adoption/development of new systems</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Increase in expenditure in R&amp;D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Share of employees dedicated to R&amp;D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Minimum viable product or service available/tested</td>
<td></td>
</tr>
</tbody>
</table>

| Impact            | -Sales (variance and levels)                         | Annual    |
|                   | -Revenues (variance and levels)                      |           |
|                   | -Labor and capital productivity                      |           |
|                   | -Value of fixed assets                               |           |
|                   | -Return on assets                                    |           |
|                   | -Number of clients, share of largest client          |           |
|                   | -Value of exports, share of sales, and countries     |           |
|                   | -Number of employees                                  |           |
|                   | -Angel investment, Series A, Series B?               |           |
|                   | -Patents and innovation indicators (TBD)             |           |

<table>
<thead>
<tr>
<th>To measure links between reviewers and evaluators</th>
<th>Indicators</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct relationship</td>
<td>-Evaluator knows applicant</td>
<td>One time</td>
</tr>
<tr>
<td></td>
<td>-Evaluator has previously worked with applicant</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Evaluator has studied in same class as applicant</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Evaluator has been a mentor/teacher to applicant</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Others.</td>
<td></td>
</tr>
<tr>
<td>Proxies</td>
<td>-Evaluator is a member of a business association as applicant</td>
<td>One time</td>
</tr>
<tr>
<td></td>
<td>-Evaluator’s professional expertise is in the same industry as applicant</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Evaluator location is the same as applicant</td>
<td></td>
</tr>
</tbody>
</table>
7.2 Management of Data Quality

Surveyors will conduct personal interviews and collect data electronically. Training of surveyors, development of protocols of data collection, monitoring and independent check-ups will be conducted according to best practices.

7.3 Ethical Issues

The project obtained human subjects approval from Columbia and MIT IRBs. We will endure all the data remains confidential.

7.4 Qualitative Instruments

In depth-interviews and focus groups will be conducted before, during and at the end of the program to understand channels of impact and interpretation of findings. The instruments and objectives will be developed as research questions arise during the evaluation period.

7.5 IE Implementation Monitoring System

The team will work closely with INADEM’s staff to ensure that all grants are distributed according to the results from the randomization exercise. Transfers are made one time through the banking system. In addition, the team will conduct check back calls to beneficiaries in order to confirm the reception of the grant.

We will work with INADEM to enhance their monitoring system to track firm’s expenditures. The team will work closely with the staff to assess if grants are being used, how and in which concepts.
8. Data Processing and Analysis

Estimation model

In order to evaluate the matching-grant program on outcomes of interest, we will run a standard intent-to-treat (ITT) specification:

\[ y_{it} = \beta Treatment_i + \gamma_s + \alpha_t + y_{i0} + \epsilon_{it} \]

Where \( y_{it} \) is the outcome indicator, \( Treatment_i \) is a dummy variable indicating treatment, \( \gamma_s \) are strata fixed effects, \( \alpha_t \) are time dummies, and \( y_{i0} \) are baseline controls including the baseline value of the outcome variable. To increase precision, we will report additional specifications that include other baseline controls. As we will have multiple outcome variables, we intend to perform an omnibus test. We will run this regression for each of the two follow-up survey rounds, reporting effects year-by-year and combined. We will cluster the standard errors by firm for the combined regression, and potentially use other corrections once the current debates surrounding the poor performance of clustering techniques made by Young (2016) have been resolved.

We will explore multiple types of heterogeneity in the treatment effects estimated by \( \beta \). First, we will explore how these treatment effects vary by group (recall group 1 was chosen only by traditional panel, group 2 only by volunteer experts, group 3 by paid business consultants and group 4 to 7 by combinations of at least two of them), as well as by the average score. A key comparison will be comparing group 1’s treatment effect to group 2 and 3’s, which we can formally test by interacting \( Treatment_i \) by group. This reveals which of the panels picks applications with higher treatment effects.

We also intend to use the scores on various components of the evaluation (described in more detail in the Evaluator Survey Data section in Section 7) to explore how firm outcomes and, if possible, treatment effects, depend on different criteria firms are evaluated on. This will further our knowledge regarding what criteria to as evaluators to judge firms on.

Furthermore, we plan to explore whether reviewers provide more accurate assessments of proposals they are proximate to (based on industry and based on other network measures that we survey). Following Li (2015), we first run the following specification:

\[ D_{ir} = \beta_1 proxi\ity_{ir} + \gamma_r + \epsilon_{ir} \]

Where \( D_{ir} \) is either a dummy variable for whether evaluator \( r \) gave applicant \( i \) a score above the cutoff (the funding eligibility cutoff), or is the actual score itself. We include evaluator fixed effects \( \gamma_r \) in case some evaluators are universally generous. We measure connections between the evaluator and the applicant firm with the variable \( proxi\ity_{ir} \) (e.g. same industry, same location, same school as CEO etc.). The regression just with the proximity term and the fixed effects is interesting and tells us the
impact of proximity on the score. However, a positive value on the coefficient on proximity is potentially a combination of both bias and the fact that proximate evaluators may be better informed (and proximity is correlated with quality of application). The addition of quality controls, and an interaction between quality and proximity addresses this issue:

\[ D_{ir} = \beta_1 \text{proximity}_{ir} + \beta_2 \text{quality}_i + \beta_3 \text{proximity}_{ir} \times \text{quality}_i + \gamma_r + \epsilon_{it} \]

We measure the quality of the application, \(\text{quality}_i\), using the estimated heterogeneous treatment effects above. This means we only have quality metrics that vary by group and potentially by other observables.\(^\text{12}\) The coefficient \(\beta_1\) reveals the bias of proximate evaluators, with a high \(\beta_1\) meaning that conditional on quality they are more likely to score highly applications they are proximate to. The coefficient \(\beta_3\) reveals the expertise of more proximate evaluators, since it corresponds to being better at scoring high quality applicants highly.

Finally, we will explore spillovers. First, we will check whether outcomes differ amongst control firms based on network proximity to treatment firms (same industry, same location, actual network connections based on network module of firm surveys). This is simply a regression of standard outcome variables (employment, investment, sales etc.) on the sum of treatment dummies of all other firms \(i'\) weighted by the inverse of network distance, \(\text{distance}_{i'i}\), metrics (as well as time dummies, strata dummies, and controls for baseline values of the outcome):

\[ y_{it} = \delta \sum_{i'} \frac{\text{Treatment}_{i'}}{\text{distance}_{i'i}} + \gamma_s + \alpha_t + \gamma_{i0} + \epsilon_{it} \]

We will also run similar regressions using other firm datasets we hope to access (e.g. the Encuesta Industrial Mensual and Encuesta Industrial Anual).

To assess the balance of the control and treatment groups we will check on observable characteristics of firms such as: size, age, sales, size of grant requested, region and category of investments. We will conduct a similar balance test on attriting firms. Previous data indicates that attrition of beneficiary firms is extremely rare given that firms have to pay back the money if they do not spend it in the categories they applied for. We will also require beneficiary firms to answer the surveys as a condition to receive and maintain the grant. We also expect limited attrition in the control group. Firms in the control group will be required to answer the surveys as a condition to apply or receive grants from other INADEM programs. We are also currently discussing with INADEM which positive incentives can be

---

\(^{12}\) These estimated treatment effects will clearly be noisy measures of quality, both because they impose the same quality level for all firms with in a grouping, and because the sample size used to estimate that groups treatment effects is small. We are currently exploring other methods of obtaining unbiased and accurate measures of quality. One possibility that solves the former method is to carry out pairwise randomization so that each treated firm has an untreated firm matched on observables ex ante. In principle this will allow each pair of firms to have their own (noisier) treatment effect. Another imperfect alternative is to just focus on firms that received that grant and use the actual outcome measures (e.g. the change in employment) as a quality measure. Note we still have a mix of low and high scores using this strategy since some applications are funded even when one panel of evaluators scores them badly (for example, if the evaluators in the other panel scored them highly enough and they were subsequently randomized into treatment).
offered in order to further reduce attrition. In case all these methods are unsuccessful, and selective attrition seems to be a major issue, we will use bounding techniques as well as propensity score matching techniques to verify the robustness of our findings.

We have registered the IE with the AEA RCT registry.¹³

9. Study Limitations and Risks

Attrition—A threat to the internal validity of the study is random attrition of the control group. To reduce the rate of attrition, firms will agree when submitting their application to respond to three surveys regardless of the outcome of their application. Firms will be informed that responding to the surveys will be a pre-requisite for applications to subsequent rounds of the program or to any other program launched by INADEM.

Take-up—Low take-up does not represent a risk for this program given the high and increasing number of applications received in previous years and the large amount of money offered to treatment firms.

Policy Relevance and Impact

The HIEP is one of the INADEM’s largest programs and is the one that provides the largest amount of money per entrepreneur. Evaluating the impact of the program and assessing whether existing strategies are effective at identifying high impact entrepreneurs is of huge value to INADEM as they allocate resources in future and decide on the structure of future programs. We are working in close collaboration with INADEM, and the results of the evaluation will directly feed into the design of subsequent rounds of this and related programs that target high impact entrepreneurs in Mexico.

Beyond this program in Mexico, the goal of this study is to more broadly inform policy towards entrepreneurship in both emerging and developed economies. First, the program targets small high impact entrepreneurs in Mexico. As far as we are aware, there is no rigorous evidence of the impact of a matching-grant program for this specific type of SME. We think this evidence has applicability beyond Mexico since the entrepreneurs are not dissimilar to high impact entrepreneurs in the US or many other developed and emerging economies.

Second, the project will help us better understand how to make matching-grant programs more effective. The knowledge generated is valuable in many settings because there are many government-funded programs that offer grants to small businesses. These programs face similar difficulties in choosing who evaluates the applications. For example, US programs at the national and subnational level have a substantial variation in the mix of evaluators across business experts, technical experts, public representatives and government employees. Our project hopes to provide guidance on how to design rules regarding panel selection.

Dissemination Plan

The impact evaluation has been discussed with relevant government representatives within and outside INADEM as well as with relevant stakeholders of the entrepreneurship ecosystem. This has facilitated the acceptance of the methodology raising awareness of the usefulness of impact evaluations to test and improve public policies. The team will continue to communicate the advancements in the evaluation design and preliminary findings as the program moves forward. At the end of the evaluation, the team will draft a policy note with the key findings of the evaluation and will organize a workshop with relevant stakeholders to discuss how these findings can feed in the design of the HIEP and similar programs.

In addition, a peer reviewed technical paper aimed for an academic audience and the impact evaluation community will be discussed in international forums such as the SME-IPA events, Nesta Gov. Lab annual conference, IGC growth week, World Bank conferences, economic and entrepreneurship forums in Mexico, etc.
10. Evaluation Team

<table>
<thead>
<tr>
<th>Role</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-Principal Investigator</td>
<td>Leonardo Iacovone, Senior Economist, World Bank Group. <a href="mailto:liacovone@worldbank.org">liacovone@worldbank.org</a></td>
</tr>
<tr>
<td>Principal Investigator</td>
<td>David Atkin, Assistant Professor, MIT. <a href="mailto:atkin@mit.edu">atkin@mit.edu</a></td>
</tr>
<tr>
<td>Principal Investigator</td>
<td>Eric Verhoogen, Associate Professor of Economics and International Affairs, Columbia University. <a href="mailto:eric.verhoogen@columbia.edu">eric.verhoogen@columbia.edu</a></td>
</tr>
<tr>
<td>Impact Evaluation Coordinator, Co-principal Investigator</td>
<td>Alejandra Mendoza, Impact Evaluation Specialist, World Bank Group. <a href="mailto:amendoza1@worldbank.org">amendoza1@worldbank.org</a></td>
</tr>
<tr>
<td>Field Coordinator</td>
<td>Emilio Lopez, Consultant, World Bank Group <a href="mailto:jlopezlopez@worldbank.org">jlopezlopez@worldbank.org</a></td>
</tr>
<tr>
<td>Government counterpart</td>
<td>Itzel Villa, Director of the HIEP, INADEM. <a href="mailto:itzel.villa@inadem.gob.mx">itzel.villa@inadem.gob.mx</a></td>
</tr>
<tr>
<td>Government counterpart</td>
<td>Alejandro Gonzalez, Director of Strategic Planning and Evaluation. <a href="mailto:alejandro.gonzalez@inadem.org.mx">alejandro.gonzalez@inadem.org.mx</a></td>
</tr>
</tbody>
</table>
## 11. Milestones, Deliverables, and Timeline

<table>
<thead>
<tr>
<th>Milestones</th>
<th>Completion Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methodology Note</td>
<td>April 2016</td>
</tr>
<tr>
<td>Intervention</td>
<td>June 2017-December 2017</td>
</tr>
<tr>
<td>Baseline data collection</td>
<td>November 2017</td>
</tr>
<tr>
<td>Midline/final data collection</td>
<td>March 2019</td>
</tr>
<tr>
<td></td>
<td>March 2020</td>
</tr>
<tr>
<td>Final analysis</td>
<td>June 2020</td>
</tr>
</tbody>
</table>
12. References


De Mel, Suresh, David McKenzie, and Christopher Woodruff (2010), “Who are the Microenterprise Owners?: Evidence from Sri Lanka on Tokman v. de Soto,”, in' International' Differences' in' Entrepreneurship, Lerner and Schoar, eds.,' University'of'Chicago'Press,'2010.'


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.


