Identifying and Boosting "Gazelles": Evidence from Business Accelerators[1]

August 2019

Juanita González-Uribe

LSE

Santiago Reyes

IADB

Why is high-growth entrepreneurship scarce in developing countries? Does this scarcity reflect optimal allocations, or constraints? We explore these questions using as laboratory an accelerator in Colombia that selects participants using scores from randomly assigned judges and offers them training, customized advice, and visibility, but no cash. Exploiting exogenous differences in judges' scoring generosity, we show that alleviating constraints to firm capabilities increases average high-growth by unlocking innovative-entrepreneurs' potential, rather than by transforming low-quality ideas. Results demonstrate that some high-potential entrepreneurs in developing economies face firm capabilities constraints, and that accelerators can help identify these entrepreneurs and boost their growth.

JEL Classification: G24, L26, M13
Keywords: High-Growth Entrepreneurship, Business Accelerators, Young Firms, Firm Capabilities

---

[1]Corresponding authors: Juanita Gonzalez-Uribe (j.gonzalez-uribe@lse.ac.uk) and Santiago Reyes (sreyes@IADB.ORG).

We thank Ulf Axelson, Jean Noel Barrot (discussant), Shai Bernstein, Vicente Cunat, Paul Gertler, Dirk Jenter, Borja Larrain, Igor Makarov, David McKenzie, Daniel Paravisini, Julio Riutort, Antoinette Schoar, Betty Simkins (discussant), Consuelo Silva (discussant), Christopher Woodruff, and seminar participants at ESSEC, FMA, IADB, LSE, Nova, Oxford, Universidad de los Andes, International Conference Finance PUC Chile, and HEC Workshop on Entrepreneurship and Economic Development for comments. We also thank Isabela Echeverry and Esteban Piedrahita at Cali Chamber of Commerce for access to the data and helpful comments. Marcela Torres and Lina Zarama provided excellent research assistance.

A large literature in economics aims to understand growth differences between developed and developing countries. Most of the focus has been on the growth constraints of microenterprises in developing economies.[2] However, recent work suggests that the development problem can be explained by cross-country differences in "high-growth young firms," also informally known as "gazelles" (Schoar (2010); Ayyagari et al. (2014); Hsieh and Klenow (2014); OECD (2017); Grover, Mesdvedev, and Olafsen (2019)). For example, Haltiwanger, Jarmin, and Miranda (2017) show that gazelles drive average growth in the U.S., and Eslava, Haltiwanger, and Pinzon (2018) show that a scarcity of Colombian gazelles drives the average lifecycle growth gap between the U.S. and Colombia. Why is high-growth entrepreneurship scarce in developing countries? Does this scarcity reflect an optimal allocation of the most talented individuals outside of entrepreneurship due to social norms and incentives (e.g., Lerner and Schoar (2010))? Or are there constrained entrepreneurs in emerging economies that are unable to grow high-potential ideas? If so, what type of constraints are faced by these entrepreneurs? And how can these entrepreneurs with high potential be identified in the population of businesses?

In this paper, we take an initial step in exploring these questions using "ValleE", an ecosystem business accelerator in Colombia, as a laboratory.[3] Business accelerators are ideal research laboratories because their aim is to identify and boost high-potential entrepreneurs by alleviating their constraints to growth. It follows that the success of accelerators in achieving these goals could contribute to solving the development problem. Focusing on ValleE is particularly useful because it allows us to address several empirical challenges that are common in exploring these questions, including: identifying high-potential entrepreneurs, distinguishing their constraints to growth, and measuring young firm performance. Our empirical strategy exploits the fact that participants were selected based on scores from three randomly assigned judges that independently evaluated their business plans. While the accelerator provided uniform criteria by which a judge should score applications, we show there is substantial variation in the interpretation of these criteria across judges.[4] As a result, otherwise identical applicants substantially differ in their acceleration probability, because they were randomly assigned to judges with different "scoring generosities." Our approach is similar to that found in the "judge leniency literature" (e.g., Doyle (2007) and the literature thereafter). The main departure is that we exploit the multiple judge assignment per applicant to control for unobserved applicant heterogeneity as perceived

[2] Empirical studies have examined financial and "managerial capital" constraints. The impact of microcredit results shows modest but not transformative results (for a summary of this literature see Banerjee (2004)). Evidence using cash grant experiments shows large average effects driven by a minority of businesses (see de Mel, McKenzie, and Woodruff (2008); Karlan, Knight, and Udry (2015)). The impact of business training programs is more mixed (for a summary of this literature, see McKenzie and Woodruff (2014)).

[3] Ecosystem accelerators are one of the three types of business accelerators. These programs are generally sponsored by governments, universities, or non-profits, and their aim is to stimulate the entrepreneurship ecosystem rather than profit (see Clarysse, Wright, and Van Hove (2015)).

[4] This variation in interpretation can reflect variation in the subjective meaning of scores; for example, in a scale from 0 to 1, a mediocre proposal might score 0.7 from "generous" judge A but only 0.4 from "strict" judge B.

by the judges. We show that controlling for this heterogeneity is crucial in our setting, given the skewness in young firm potential, and the ability of ValleE judges to identify high growers (that we show is orthogonal to scoring generosity). To track performance, we use novel administrative data from the Colombian business registry two years before application and five years after, together with annual surveys.

We focus on identifying constraints to "firm capabilities," as ValleE provided participants with *no cash,* but instead offered standardized business training, customized business advice, and visibility. By firm capabilities we mean key elements of the growth process that firms cannot readily buy in the market and that impede growth even if ventures are injected with cash, such as identifying the correct market need or gaining market recognition (cf., Sutton (2012)). Little is known about the importance of these constraints for high growth, and skeptics argue that these constraints cannot be alleviated, because high-potential entrepreneurs are born with skills that cannot be taught.

We find compelling evidence that participation in the accelerator has large positive impacts for the "marginal applicants" whose selection decision is altered by the judge assignment and that would (not) have been selected but for the strictness (generosity) of their randomly assigned judges. Using an instrumental variables (IV) approach, we estimate that over the first three post-application years, participation in the accelerator increases the marginal applicant's annual revenue by $66.3M COP ($20K USD), a 166 (130) percent increase from the rejected applicant (average applicant) revenue.[5] Exploiting the continuity of our instrument, we show large heterogeneity in impacts, and demonstrate that the IV results come from the remarkable growth of high-potential participants, rather that from any apparent performance improvements of low-potential participants that were mistakenly accepted into the program because they were assigned to more generous judges. Taken together, results show that alleviating constraints to firms' capabilities increases average high growth, particularly by unlocking innovative entrepreneurs' potential.

Examining the mechanisms by which average performance increases, we find evidence that customized advice and visibility are more impactful aspects of the intervention than the standardized business training. Consistent with entrepreneurs' perceptions in business accelerator programs worldwide (see Roberts et al. (2017)), 74% of surveyed ValleE participants say that advice and visibility added the most value, whereas only 8% described the standardized business training as a key impact driver. Furthermore, estimated effects are stronger among applicants that at the time of application reported needing strategic advice rather than standardized training; and among applicants that could take the most advantage of strategic advice and visibility, such as applicants with already existing businesses (rather than business ideas) at application. This evidence substantiates the documented

[5]There is also evidence that acceleration has large positive impacts on profits and employment.

outperformance of customized, over standardized, interventions in the context of microenterprises.[6] As one limitation, we note that acceleration can also trigger additional impact mechanisms such as capabilities-enhancing network, certification and internal validation effects, which, as is common in the literature on capabilities building interventions, we cannot rigorously distinguish with our current data (e.g., Chevalier, Harmon, Walker and Zgu, (2004)).

Overall, our findings provide an initial step in understanding why high-growth entrepreneurship is missing in developing countries. A back-of-the-envelope calculation based on our findings implies that marginal applicants impacted by ValleE surpass Colombian top growers, and reach about 3 times their initial revenue by the 4th year of the business on average, which roughly doubles the 90th percentile of 4-year lifecycle revenue growth of Colombian businesses (cf., Eslava and Haltiwanger (2018)). Taken at face value, this calculation shows that the accelerator unlocked the "gazelle potential" of marginal applicants by accelerating these applicants' growth towards the rates of top growers in the U.S. Our findings suggest that in developing countries constraints to firms' capabilities can hinder the ability of high-potential entrepreneurs to reach their full gazelle potential. These findings question many of the models of entrepreneurial activity in developing countries, which treat firm capabilities as fixed and focus instead on financial constraints (Banerjee and Newman (1993); Paulson and Townsend (2004)).

Our findings also provide compelling evidence that accelerators can help solve the development problem by identifying and boosting top growers, but not necessarily by transforming low-quality projects. As such, they also highlight the limits of these programs as policy instruments. As one caveat, we note that, as all firm capabilities' interventions that have been conducted so far, our empirical exercise is a joint test of two closely related hypotheses: first, that firm capabilities are first-order constraints to high growth, and second, that firm capabilities can be conveyed via the intervention in the first place. Therefore, failure to find effects for low-potential ideas does not necessarily prove that firm capabilities do not matter for the growth of these projects or that firm capabilities cannot be taught to low-potential entrepreneurs. Instead, another simple explanation could be that the firm capabilities provided by the business accelerator were not enough for these types of projects and that more intensive interventions are needed (cf., Bruhn, Karlan and Schoar (2018)).

This paper is related to two main literatures. The first addresses the sources of and constraints to high-growth entrepreneurship. An increasing body of work shows that a small number of young

[6] Several papers in the business training literature show the importance of accompanying standardized business training with more personalized components such as mentoring and technical assistance; see Bloom, Eifert, Mahjan, McKenzie, and Roberts (2013); Valdivia (2015); Brooks, Donovan, and Johnson (2018); Bruhn, Karlan, and Schoar (2018); Campos, Goldstein, Frese, Iacovone, Johnson, McKenzie, and Mensmann (2018); Ubfal, Arraiz, Beuermann, Frese, Maffioli, and Verch (2019). For reviews see also: McKenzie and Woodruff (2014) and Quinn and Woodruff (2019).

firms—variously termed "high-growth entrepreneurs," "gazelles," "transformational entrepreneurs," or "super-start firms," among other monikers—make disproportionate contributions to economic growth.[7] At the same time, a long-established literature suggests that a perceived lack of sufficiently impactful high-growth entrepreneurs in developing economies may be in part due to market failures. For example, limited access to firm capabilities—also called "entrepreneurial capital" or "managerial capital," among other terms—can prevent individuals with high-potential ideas from successfully growing their businesses (e.g., Bruhn, Karlan and Schoar (2010)). Our findings bring new insight to this literature by showing that a deficiency of high-potential participants and/or of more customized intervention services can help explain the historically lackluster performance of firm capabilities building programs (cf., de Mel, McKenzie, and Woodruff (2014)). The closest paper to ours is McKenzie (2018), which shows no impact of standardized business training on venture growth, but does find large effects from the provision of bundled training and cash in the context of a high-growth business plan competition in Nigeria. Buttressed by the heterogeneous impact patterns and the survey evidence, a compelling explanation for the comparative success of ValleE is its provision of intensive customized advice, which is not commonly included in business plan competitions given their short-term and more "at arm's length" nature.

Our work is also related to the growing literature on business accelerators where selection issues are a main concern to identify their impacts. Business accelerators are an increasingly popular method used by governments to help high-growth firms.[8] While these "business schools for entrepreneurs" have been touted by the popular press as critical to the development of start-up ecosystems, the evidentiary support for their impact remains thin, especially outside the innovation hubs of Silicon Valley and Boston. Measuring the impacts of business accelerators is hard, because cohorts are typically very small, and because accelerators put a lot of emphasis on trying to select the best people for the cohort. This raises concerns that approaches that use matching/selection-on-observables (as is common in previous studies) will be severely biased.[9] But on the other hand, it is going to be hard to convince many accelerators to randomize which applicants they take.

[7] The disproportionate contribution of high-growth young firms to economic growth both in developed and developing economies is a well-established fact in economics (see Birch and Medoff (1994); Henrekson and Johansson (2008); Schoar (2010); Haltiwanger, Jarmin, and Miranda (2013); Ayyagari et al. (2014); Hsieh and Klenow (2014); Grover, Mesdvedev, and Olafsen (2019); Eslava, Haltiwanger, and Pinzon (2018)).

[8] The proliferation of business accelerators is well documented; see, for example: Cohen and Hochberg (2014); Grover, Mesdvedev, and Olafsen (2019); Clarysse, Wright, and Van Hove (2015); Roberts, Davidson, Edens, and Lall (2017). The increasing prevalence of public funds for these programs is also well documented; an estimated 40% of businesses accelerators receive some form of government support (e.g., Bone, Allen, and Haley (2017)).

[9] A non-exhaustive list of papers using matching/selection-on-observables to assess accelerator/incubator impacts includes: Colombo and Delmastro (2002); Schwartz (2009); Yu (2016); Smith and Hannigan (2015); Hallen et al. (2014); Lasrado et al. (2016). See also Bone, González-Uribe, Haley, and Lahr (2019) for a summary of the literature.

We make two main contributions to the literature on accelerators. First, we propose a novel identification strategy that can work even when programs are trying to choose the best firms, precisely by showing the amount of randomness that comes from human decision-making in judging success. This approach is thus potentially applicable to many efforts to evaluate these types of programs worldwide. Second, we characterize for the first time the treatment heterogeneity of the non-financial services provided by these programs. Instead, most prior impact estimates cannot be extrapolated beyond a small subsample of entrepreneurs with similar growth potential; for example, in papers exploiting qualifying thresholds to assess impacts using regression discontinuity designs, as is the case in the closest paper to ours in this literature, González-Uribe and Leatherbee (2018a). The patterns of treatment heterogeneity show that selection lies at the heart of these programs' success, and question sceptics' contention regarding the inability of these programs to impact high-potential entrepreneurs.

Our research laboratory is, however, not without limitations, and the main drawback is the small cross-section of applicants in business accelerator programs. We have 675 firm-year observations, corresponding to a 5-year panel for the 135 applicants that were evaluated by the expert judges. To address potential small sample biases—namely, the possibility that our effects pick up the 5% chance we would see an effect when no such effect exists—we show the robustness of results to using different measures of scoring generosity, sets of controls, methodologies, and outcome variables. We address potential issues from high serial correlation in outcomes using the approach by Fee, Hadlock, and Pierce (2013). We also show that the results continue to hold when using local randomization methods (Rosenbaum 2002) that conduct exact finite sample inference and remain valid even when the number of observations is small. We propose an intuitive adaptation of these methods to our setting, which can also be generalized to other contexts. In terms of external validity, we show that the sample is not special by tracing similarities between the entrepreneurs in our sample and the average applicant to ecosystem accelerators worldwide. We emphasize, however, that the external validity of our findings is likely confined to other ecosystem accelerators in developing countries that attract young businesses with traction and have access to high-quality resources, including staff, mentors, and judges.

The rest of this paper proceeds as follows. In Section 1, we describe the context and data. In Section 2, we detail the empirical strategy and present results. We discuss the interpretation of results and their external validity in Section 3. We present robustness checks in Section 4, and offer concluding remarks in Section 5.

1. Institutional Setting

ValleE is a local ecosystem business accelerator that was launched during 2015 after an intense local advertising campaign using social media and radio in the city of Cali, the third most important city in Colombia in terms of population.[10] The accelerator is the brainchild of the Regional Network of Entrepreneurship in ValleE del Cauca (a private organization that aims to encourage entrepreneurship in the ValleE del Cauca region), and is operated by the city Chamber of Commerce, a private entity that has been delegated public duties such as the management of the Colombia business registry.[11] As is common among ecosystem accelerators, ValleE's main objective is to encourage local growth by identifying and boosting high-growth entrepreneurs (cf. Clarysse, Wright, and Van Hove (2015)).[12] Examples of ValleE participants include "Luces projects" a company offering residential wind energy solutions and "Contratan.do," an information and communication technologies' business-to-business hiring platform in Latin America.

Like other business accelerators worldwide, ValleE is a fixed-term, cohort-based program that selects participants based on the relative quality of applications submitted online, as evaluated by a panel of judges (cf. Cohen and Hochberg (2014); Leatherbee and González-Uribe (2018b)). As explained in more detail in Section 1.2, participants are selected based on *average scores* from partially overlapping 3-judge panels in order to satisfy pre-determined budget and space restrictions, as well as judges' time constraints. Any person proposing the creation of a new business or the scale of an existing young (0-3 years) business located in the region is in principle eligible for the program. However, the program focuses on high-growth entrepreneurs, and many applicants are *de facto* incompatible and thus rejected (as explained in more detail in Section 3).

Also similar to traditional business accelerators, ValleE provides participants with firm capabilities (which we describe in more detail below). The distinguishing feature of our setting, however, is that the program offers no cash (as is nevertheless common among the subset of ecosystem accelerators worldwide, cf. Clarysse, Wright, and Van Hove (2015)).[13] The perception is that for many young businesses the foremost constraint to growth is their lack of firm capabilities, which contrasts with the usual academic narrative. The starting point of most entrepreneurial finance models are entrepreneurs with access to, and ability to execute, positive NPV opportunities that cannot convince investors to give them funds for a myriad of potential reasons (unrelated to the notion of firm capabilities), such as investors' fear of moral hazard or adverse selection. The narrative in ecosystem business accelerators questions this starting point, and instead argues that entrepreneurs with access to positive NPV opportunities may have no actual ability or market recognition to successfully execute

[10] Ecosystem business accelerators are popular in low and lower-middle income countries: 37.9% of the ecosystem accelerators in the Entrepreneurship Database at Emory University are located in Africa (17.9%), Latin America (10.3%), and India (10.3%).
[11] Chambers of Commerce oversee the private sector development policies in their region. They are key connecting actors that execute programs aimed at improving regional competitiveness.
[12] The top two impact objectives among ecosystem accelerators are employment generation (35%) and community development (30%). Source: The Entrepreneurship Database at Emory University.
[13] Circa 55% of the ecosystem accelerators in the Entrepreneurship Database at Emory University provide no seed capital. Source: https://www.galidata.org/accelerators/.

those opportunities. Consequently, the businesses of these entrepreneurs would not grow even if they received cash injections.

Like traditional business accelerators, ValleE provides participants with firm capabilities through a variety of services, including: standardized grouped business training, one-to-one customized advice, and increased visibility.

The business training sessions are highly structured and simultaneously attended by all participants in the offices of the Chamber of Commerce. They consist of roughly 8 weekly hours of standardized content (100 hours overall in a space of three months) delivered by hired local and national experts. Bootcamps combine lecture-based conceptual sessions together with case-based sessions discussing real-life practical examples, and cover the topics of business modelling, early-stage financing, market validation, prototyping, accounting, and pitching. Two types of one-to-one customized advice sessions are provided. The first type consists of bi-monthly meetings to discuss business strategy with high-level advisors assigned based on industry, which include renowned CEOs in the region, as well as managers at the Chamber of Commerce. Assigned advisors may provide introductions to potential clients or industry contacts, which are likely to be high impact, as the selected CEOs and Chamber of Commerce managers are well connected within the local ecosystem. The second type of mentoring sessions are handled by program coordinators who take a more hands-on approach: sessions are conducted weekly and are of varying duration. Coordinators are junior to advisors and focus on helping entrepreneurs throughout the day-to-day operations rather than designing avenues for growth. Finally, ValleE provides several opportunities to increase visibility: participants are showcased on the Chamber of Commerce website, in their monthly publications, and exhibited at different events. At the end of their term, participating businesses "graduate" through a "demo day" competition (i.e., a formal presentation of the companies to potential investors).

It is possible that participants benefit through mechanisms other than the services *per se*. For example, services such as customized advice and visibility can trigger potential network, certification, or internal validation effects that have been shown to have large impacts in other contexts such as among prestigious business school students (e.g., Lerner and Malmendier, 2013). We return to this point in Section 2.7, where we discuss the potential channels of impact.

1.1. Sample

ValleE provided us with all the application data, including application scores by each judge and final selection decisions, for the program's first cohort.[14] All selected applicants in this cohort

[14] The judges' identities were not provided by ValleE for confidentiality reasons. For the purpose of our investigation, we were provided with anonymized information that includes judge identifiers in order to track different projects evaluated by the same judge.

participated in the accelerator for three months, during May, June, and July of 2015. Our sample consists of 135 projects (35 participants and 100 nonparticipants) that applied to the accelerator in March of 2015 and were deemed to have high-growth potential by the staff, as explained in Section 2.

Our sample size is standard for business accelerator programs, and exceeds that of similar papers exploring the impact of business training (e.g., 14 participants and 14 control plants: Bloom et al. (2013); 47 participants and 66 control business owners: Mano et al. (2012)). However, it is small enough for concerns to be raised, for instance regarding our ability to detect the impact of the accelerator if such an effect exists. We return to this issue in Section 2, where we describe several robustness checks we run to address this issue, and in Section 4 where we conduct exact finite sample inference using a randomization inference approach that is valid even when the number of observations is small.

Based on the program's records, we constructed several variables to use as controls in our empirical strategy: the age of the firm (*Firm Age*); the founder's sector experience in years (*Experience*); indicator variables for male applicants, serial entrepreneurs, and projects with founding teams (*Male, Serial, Team*); projects located in Cali (*Cali*); founder's education (*High school, Technical degree, College, Graduate*); founder's motivation to start a business (*Stable income, Own boss, Opportunity)*; and industry and location indicators. Baseline information on *Revenues, Profits,* and *Employees* are also included.

Table 1 reports summary statistics for the main variables in the application forms. On average, applicants have 5.6 years in sector experience, are male (79%), educated (67% have a bachelor's or master's degree), are likely to be serial entrepreneurs (61%), and have a founding team (88%). The average number of founders is 3, and the average number of employees is 4. The likelihood of positive revenues is 45%, and median (positive) annual revenue is $52M COP—approx. $15,000 USD. Most applicants are in the services industry (56%), have participated in other entrepreneurship contests (59%), and applied with business ideas (53%) rather than already established firms (47%). Applicants classified as having "business ideas" include informal businesses that at the time of application were trading but had not been incorporated in the business registry of Colombia. Accordingly, the average revenues and employment at baseline for these businesses were positive but modest ($4.61M COP [$1.3K USD] in revenue and 2.7 employees in 2014; see Table 1).

Our sample is comparable to average applicants of ecosystem accelerators worldwide, based on information from the Entrepreneurship Database (ED) program at Emory University. The average applicant to ValleE is similarly sized (ED applicants have an average of 3.5 employees, a 43.2% likelihood of positive profits, and median (positive) revenue of $12,000 USD) but is more educated (47% of ED applicants have a bachelor's or master's degree), less likely to be female (29% of ED applicants are female), and has a more mature business (19% of ED applicants report positive revenues prior to application).[15] Our sample is also comparable to that used in prior work on ecosystem business accelerators: González-Uribe and Leatherbee (2018a) show that applicants to Start-up Chile, a renowned ecosystem accelerator sponsored by the Chilean government, are likely to be male (86%), have between two and three employees, and are predominantly from services industries such as E-commerce (18%). Finally, our sample is also similar to that in prior work on early stage ventures. Haltiwanger, Jarmin, and Miranda (2013) document that 33% of young firms (less than a year old) in the U.S. have between one and four employees, and Puri and Zarutskie (2012) show that the distribution of VC-backed firms is concentrated in the services industry.

1.2. Accelerator Selection Process

Selection into ValleE is a four-part process. First, aspiring participants submit an online application that requests information about the entrepreneurs and their detailed business plans. Next, the accelerator filters applicants to exclude projects that are deemed to have no high-growth potential (e.g., taxi drivers, shopkeepers). Filtered applications are then randomly assigned to three judges that individually score the application.[16] The total number of judges is 50, and thus judges only partially overlap across applicants. The judges evaluate the applications according to five criteria: (i) clarity of the business model proposal, (ii) innovation, (iii) scalability, (iv) potential profitability, and (v) entrepreneurial team. Finally, the staff at the accelerator makes the final decision by picking the top 35 applicants based on average scores. It is impossible for judges and applicants to manipulate the ranking process. Judges are unaware of the weight of each criteria in the final score, they independently score projects, are not aware of the identity of the other judges in the panel, and no judge sees all applications. Applicants do not know who their judges are, nor do they know their position in the ranking. [17] The capacity threshold of 35 participants was determined prior to the launch of the program and is due to budget and space limits.

In the first cohort of ValleE (our sample source), there were 255 applicants who submitted a complete application online. Of these, only 135 businesses were deemed to have "high potential for growth" and therefore correspond to our analysis sample.[18] The maximum length allowed for business plans submitted with the applications and read by the judges was 2 pages. The average number of projects scored by any given judge was 8, and the minimum (maximum) was 5 (14). The program

[15] See https://www.galidata.org/accelerators/.

[16] The main reasons behind using judge panels (rather than individual judges) are to minimize the burden on individual judges (given their time constraints), and mitigate the chance that one judge determines the treatment status of any given project, as this could lead to unwanted biases such as judges favoring projects from their own industries, regions, or communities.

[17] Entrepreneurs were never given their ranking or scores in order to avoid any negative psychological effects or create rivalry among participants.

[18] The characteristics of the final 135 projects differ slightly from the 120 businesses removed by the initial filter, which were more likely to have a female founder, have less educated founders, and refer to nonpecuniary benefits (e.g., being their own boss) as the main motivation behind their business.

picked the judges based on the relevance of their backgrounds to help sort applicants. Judges were not compensated for evaluating applicants, and their identities are private to us. The pool of 50 judges included individuals with substantial experience in business and entrepreneurship, such as c-level executives in local businesses, independent business consultants, industry experts, as well as managers in entrepreneurship departments in development agencies and two staff members. This average business and industry expertise of judges is not necessarily common among other business accelerator programs, where applications are managed by platforms that rely on a wider variety of less 'hands-on' experienced judges such as academics (cf., Gonzalez-Uribe and Leatherbee, 2018a).

Compliance with the selection rule was perfect: the top 35 applicants (based on judges' average scores) were selected, and all selected applicants participated (see Figure 1, Panel A). Table 2 shows statistically significant differences at the time of application between accelerated and nonaccelerated applicants: participants have bigger founding teams, are slightly more educated, have more sectorial experience, and are more likely to be serial entrepreneurs. The economic significance of most of these differences is, however, small, in part due to the filter applied by the program to remove the non-transformational entrepreneurs from the sample.

While the accelerator provided uniform criteria by which a judge should score proposals (see Appendix 1), we show in Section 2.1 that there was substantial variation in the interpretation of these criteria across judges in the first cohort of ValleE. This heterogeneity in "scoring generosity" is reminiscent of the systematic differences in "judge leniency" reported in other settings, such as in bankruptcy courts in the U.S. (e.g., Dobbie and Song, 2015). In Sections 2.2 and 2.3, we discuss how we use this heterogeneity in scoring generosity across the randomly assigned judges to estimate the causal impact of participation in the accelerator.

1.3. Outcome Data

Table 3 presents summary statistics of the main variables used in the regression analysis, including data on firm outcomes.

We use two complementary strategies to collect outcome data. First, we collected novel administrative data from the business registry in Colombia on registration, survival, and annual revenues. The Colombian registry includes information on annual revenues and closures, because Colombian firms submit annual mandatory business license renewals with the Chambers of Commerce that manage the registry. The renewal of an operating license for firms in Colombia is mandatory *de jure* and *de facto*, as companies are required to submit this license to validate their operations with their banks and corporate clients, among others. Using these data, we track annual revenues 2 years before and 3 years after application to the program. We also use these data to discern which applicant *ideas* eventually turned into actual businesses (i.e., start-up rates), and to distinguish which applicant *established firms* continued operating (i.e., survival rates) from those that instead did not renew their operating license (i.e., closure rates). Access to administrative data represents a major advantage relative to most other work in the literature of accelerators and in the literature of business training interventions. With a few exceptions, most prior related work relies on surveys to measure performance (see Woodruff (2018)). We also improve upon prior work where the median number of follow-up observations per firm is one (cf., McKenzie and Woodruff (2014)). Longer follow-ups are important, as short- and long-term impacts of many policies can differ substantially (cf. King and Berhman (2009)).

Our second strategy to collect performance data was to partner with ValleE to design a performance survey that was sent by the program to all applicants every year for three years after application to the program. The yearly surveys included questions regarding revenue, employment, and profits. Our main objective for the survey was to collect additional performance metrics based on employment, profits and fund-raisings, which are not included in the business registry, and which we used to explore other outcomes, and run robustness checks (see Appendix 7). Survey response rates were 77%, 67%, and 60%, respectively, in the years 2016, 2017, and 2018; with participants having slightly higher rates (82%, 77%, and 65%) than nonparticipants (75%, 64%, and 58%). These annual survey response rates were much higher than that found in prior work on business accelerators (e.g., 10% in González-Uribe and Leatherbee (2018a)), and imply similar survey attrition rates to those in related papers exploring the effect of business training in microenterprises and small- and medium-sized firms (e.g., 25% in Karlan and Valdivia (2011); 26% in Calderon, Cunha, and de Giorgi (2013); 28% in Klinger and Schundeln (2011)).

One novel feature of our data collection strategies is that we have information on revenues from both administrative data and surveys. Thus, we can cross-check self-reported revenues in the surveys with those in the registry to gauge the degree of potential selective survey attrition and selective survey responses. We find little evidence of either, which mitigates concerns of data quality from the survey variables, and which lends credence to the additional analysis using the information on employment, profits, and fundraising (see Appendix 7). On average across the survey years, most (74%) survey attrition can be explained by real business closures rather than the refusal of ongoing businesses to answer the survey questions. The refusal rate is no different between participants and nonparticipants, which suggests that refusal is not endogenous to participation (e.g., nonrespondents are busy firms, rather than mistrustful nonparticipants) and helps mitigate concerns regarding the impacts of the program on quality of outcome data besides the potential effect of acceleration (cf. McKenzie and Woodruff (2014)).[19] There is little discrepancy between self-reported and registry-based revenues with

[19] P-values for differences in response rates between treatment and control groups are 0.352, 0.167, and 0.427 for follow-up surveys 1 to 3 years after treatment.

a 95% coincidence rate (and with 60% of the discrepancies being due to mistakes—e.g., missing or extra zeros), and discrepancies do not vary across participants and nonparticipants.

2. Empirical Strategy and Results

In this section, we explain how we exploit the random allocation of projects across judges with different scoring generosities to show causal and heterogeneous impacts of the accelerator on growth. We begin by showing that there is large variation in scoring generosity across judges—i.e., some judges tend to assign high scores, whereas some judges tend to assign low scores. We then show that scoring generosity strongly predicts selection into the accelerator. Next, we use an instrumental variables (IV) approach based on scoring generosity as an exogenous predictor of acceleration to assess the program's impact. Finally, we characterize impact heterogeneity by exploiting the continuity of our instrument.

2.1 Scoring Generosity

We provide evidence of systematic differences across judges in scoring generosity by exploiting the multiple judge assignment per applicant to run fixed effects models of application scores against judge and applicant fixed effects. Our approach is similar to the methodologies in papers assessing the importance of managers in corporations (cf. Bertrand and Schoar (2003)) and general partners in limited partnerships (Ewens and Rhodes-Kropf (2015)). The idea is that judge fixed effects would be jointly significant if judges systematically vary in their tendency to assign high or low scores to projects.

We begin by decomposing individual scores into company and judge fixed effects using the following regression:

$$(1)\ Score_{ij} = \alpha_i + \gamma_j + \varphi_{ij}\ ,$$

where α_i are project fixed effects and γ_j are judge fixed effects. We normalize scores so that they vary between 0 and 1, corresponding, respectively, to the "bottom score" project (scored at 1 by all judges) and the "top score" project (scored at 5 by all judges). Each judge fixed effect is estimated with 8 observations on average (see Section 1.2). These judge fixed effects are meant to capture heterogeneity across judges in their scoring generosity. By contrast, the project fixed effects can be understood as the quality of the project that all judges agree on; they represent "adjusted scores" after controlling for potential systematic differences in scoring generosity across judges. For conciseness in exposition, we refer to the estimated firm fixed effects in equation (1) as adjusted scores hereafter. Also, to ease exposition, we plot the results in Figures 2 and 3, rather than report regression estimates. Figure 2 plots the distribution of fixed effects across judges. Figure 3 plots the average score against adjusted scores; the correlation between these two scores is high at 0.83 (significant at the 1% level).

There are four main findings from estimating equation (1):

First, there is statistically significant heterogeneity across judges: the *F*-test on the joint significance of the judge fixed effects is 5.49 (p-value of 0.00). By contrast, if judge heterogeneity was irrelevant (or nonsystematic), then judge fixed effects would not be jointly significant (as judges are randomly assigned by design). To address concerns regarding the validity of *F*-tests in the presence of high serial correlation (Wooldridge (2002)), we scramble the data 500 times, each time randomly assigning judges' scores to different applicants while holding constant the number of projects evaluated by each judge and making sure that each project receives three scores, in the same spirit as in Fee, Hadlock, and Pierce ((2013).[20] Then we proceed to estimate the "scrambled" projects' and judges' fixed effects and test the joint significance of the latter in each scrambled sample. The distribution of the scrambled *F*-tests is plotted in Figure 4 (Panel A). Lending credence to the statistically significant judge heterogeneity in our setting, we reject the null of "no joint significance of the judge fixed effects" in only 3.99% of the placebo assignments (the largest estimated placebo *F*-test is 1.84).

The second finding is that the judge heterogeneity reflects systematic differences across judges, as the fixed effects do not appear to be driven by outliers or capturing noise. We see very small differences across the different "leave-one-out" estimates of judge fixed effects; as Figure 5 shows, the average standard deviation per judge is 0.003, and the maximum is 0.006. Figure 5 plots the distribution of the standard deviation of the estimated leave-one-out judge fixed effects by judge (with each point representing one judge). For a given project *i*, the leave-one-out approach uses all observations except project *i* to estimate the judge fixed effects (see Figure 5 for more estimation details).

The third finding is the sizable *economic* significance of the scoring generosity heterogeneity.[21] Figure 2 shows that the most generous (strict) judge adds (subtracts) an average of 0.26 (0.28) to any given project she scores, roughly a third of the mean average score of 0.7. Relying on a panel of judges rather than on individual judges helps mitigate the effect of judge heterogeneity by averaging out the

---

[20] In the parallel literature, when seeking to identify the "style" of managers using an endogenous assignment of (movers) managers to multiple companies (e.g., Bertrand and Schoar (2003)), concerns have been raised regarding the validity of *F*-tests in the latter settings on the grounds of (a) the particularly acute endogeneity in samples of job movers and (b) the high level of serial correlation in most of the variables of interest (see Fee, Hadlock, and Pierce (2013)). The first reason for concern is not at play in our setting, as judges are randomly assigned by design, but the second concern may still apply. Regarding the second concern, Heckman (1981) and Greene (2001) discuss the ability of small sample sizes per group to allow for meaningful estimates of fixed effects with a rule of thumb of eight observations per group.

[21] In Figure 2, 34% of judges tend to systematically award individual scores that are one standard deviation above or below the average score of the other judges, whereas only 6.8% of judges do so in the 500 placebo assignments (Panel B in Figure 4; see Fee, Hadlock, and Pierce 2013).

scores of a strict judge with a lenient one in some cases. However, it does not fully correct it, because judge panels are small, with only three individual judges assigned per applicant (see Appendix 2).[22]

The fourth finding is that these systematic differences across judges are unrelated to the judges' skill in distinguishing high growers, and instead reflect judges' propensities to assign high or low application scores. Figure 6 shows a nil correlation between judges' generosity and their ability to correctly rank applicants. We measure judges' ranking ability using the correlation between a "judge's ranks" and "actual ranks." To produce this correlation, for every judge we rank the companies she evaluated based on (i) 2017 revenue ("actual rank") and (ii) the judge's score ("judge's rank). Figure 6 is a scatterplot of each judge's generosity and ranking ability for the 50 judges in our sample.

We note that the results in Figure 6 do not imply that judges' have no ranking ability. In fact, Table 4 shows that projects' actual ranks are predicted by the judges' ranks, even after controlling for judge fixed effects. That is, judges are on average good at ordering projects according to their potential, even though they vary in how high or low are the scores they assign. Table 4 shows results from regressions of actual ranks on judges' ranks and includes judge fixed effects (Column 2).

However, the results in Figure 6 do suggest that, in contrast to judges' ranks, average scores are unlikely to be good predictors of performance because scoring generosity artificially inflates (deflates) the judges' perceived potential of lucky (unlucky) projects as shown in Figure 3. Instead, adjusted scores are more likely to reflect the predicting ability of judges evidenced in Table 4, precisely because adjusted scores "clean" the average scores from these systematic differences in scoring generosity across judges.

Consistent with this intuition, Tables 5 and 6 show that adjusted scores predict revenue in a way that adds to the predictive power of the hard information in applications responses, whereas average scores are not correlated with future performance (or high growth). Tables 5 and 6 summarize estimates from variants of the following regression using observations during 2013–2017 for all applicants to the accelerator:

$$(2)\quad k_{it} = \alpha + \beta\, Score_i + \rho\, Score_i \times After_t + \theta Z_{it} + \mu_t + \varepsilon_{it}\ ,$$

where *i* indicates applicants and *t* time, $After_t$ is a dummy that equals 1 during 2015-2017, Z_{it} is a vector of controls, and $Score_i$ corresponds to either the *Average Score or Adjusted Score*. We include time (μ_t) fixed effects and several controls at the founder and project levels (i.e., the hard information described in see Section 1.1), as well as interactions of these characteristics with the variable $After_t$.

[22] The small size also explains why random assignment does not deal with this issue: for a given project the scores of overly generous judges may not tend to cancel out those of overly strict judges, as the probability that both types of judges will be randomly assigned to the same project does not tend to 1.

We bootstrap standard errors clustered at the applicant level to account for any serial correlation across applicants, and for the fact that the adjusted score is a generated regressor (Wooldridge (2002); Young (2018)). Table 5 shows OLS estimates using revenue as dependent variable; Table 6 estimates a probit model using the variable for "gazelles" as a dependent variable and indicator. We define gazelles as applicants that are in the 90th percentile of revenues by 2017.[23]

Table 5 shows that the top adjusted score project has an additional 86 million COP ($26K USD; Column 2, Panel A, Table 5) in annual revenue relative to the bottom adjusted score project, controlling for secular growth and applicants' characteristics (relative to a baseline level of revenues in 2014 of 25 million COP; Table 3).[24] The correlation between adjusted scores and future revenue is strongest for business ideas (see Appendix 3), and is not explained by a potential treatment effect, but rather reflects the predictive power of judges, as it holds even after controlling for participation (see Column 3, Panel A, Table 5).[25] Table 6 shows that adjusted scores also specifically predict high growth. Applicants among the top quartile of adjusted scores are 20% more likely to become gazelles by the end of the sample period (Column 1, Panel A, Table 6), controlling for covariates and acceleration (Columns 3 and 4, Panel A, Table 6). In contrast to adjusted scores, average scores are uncorrelated with revenue (see Columns 5 and 6, Panel A, Table 5), and explain a much lower fraction of the variation in revenue than entrepreneurs' characteristics (see Column 3, Panel B, Table 5).

Why are ValleE judges able to predict high-growth in a way that adds to the predictive power of the hard information in applications responses? The reason is that judges evaluate projects not only based on applicants' characteristics (i.e., the regression controls), but also based on the business plans, which are unobservable to the econometrician and whose information is not easily codified. Results in Tables 5 and 6 thus suggest that business plans have information that helps predict future growth when

---

[23] So defined, these gazelles surpass top Colombian growers: they expand about 5 times their initial revenue, roughly doubling the 90th percentile of the 4-year lifecycle revenue growth of Colombian businesses (cf., Eslava and Haltiwanger (2018); see Panel B, Table 5). We note, however, that there is no general definition of gazelles; Henrekson and Johansson (2008) show a large variation in definitions in their meta-analysis of the empirical literature. Our classification follows other papers in the gazelle literature that define gazelles using performance thresholds (Kirchhoff (1994), Picot and Dupuy (1998), Schreyer (2000), Fritsch and Weyh (2006), McKenzie and Sanson (2017)). Our threshold is based on size (revenues in 2017) to avoid the difficulties of measuring growth rates in our sample: a large fraction of our applicants have zero revenue at baseline (57%; see Table 1). To implement this classification, we split the sample into two groups according to age at application: (i) more than 1 year relative to incorporation (77 applicants) and (ii) less than 1 year since incorporation or not incorporated (58 applicants). We then define as gazelles the top 10% of firms in each group according to revenue in 2017 (9 and 7 applicants in each group, respectively; see Appendix 4). Applicants classified as gazelles would also classify as high growers under definitions based on growth rates rather than levels: Panel B of Table 5 shows that their average annual revenue growth is 68% in the three years following their application, which exceeds the 20% growth rate requirement for gazelles in most other definitions used in the literature (20% for Birch and Medoff (1994) and Birch et al. (1995); 50% for Ahmad and Petersen (2007); Deschryvere (2008), and Autio et al. (2000)).

[24] Adjusted scores explain 4.4% of variation in revenues, almost a third of the contribution of the entire set of entrepreneurs' characteristics (Column 3, Panel B, Table 5).

[25] This is as expected: if the correlation was fully explained by treatment, then average scores should be more predictive of future performance than adjusted scores—after all, participation in the accelerator is defined by the average, rather than the adjusted score.

evaluated by judges. We note, however, that judges' predicting ability may not necessarily replicate well in other settings, as ValleE judges have uncommonly high business and industry experience, as compared to the judges in other business accelerators (see Section 1.2). We return to this point in Section 3 were we discuss the external validity of the findings.

In conclusion, the results from this section show that by selecting participants using average scores rather than judges' ranks, the program made selection mistakes and rejected (accepted) some unlucky (lucky) applicants with strict (generous) judges. We now turn to this point and explain how we exploit these selection mistakes to assess the causal impacts of the program.

2.2 Exploiting Scoring Generosity as an Exogenous Predictor of Acceleration

We estimate the casual impact of acceleration through a two-stage least squares (2SLS) regression using scoring generosity as an instrumental variable for acceleration. The second stage estimating equation is:

$$(3)\ k_{it} = \alpha + \vartheta_i + \rho Acceleration_i \times After_t + \theta X_{it} \times After_t + \mu_t + \varepsilon_{it},$$

where ϑ_i are project fixed effects, μ_t are time fixed effects, and X_{it} includes several controls (the hard information from applications; see Section 1.1), which are interacted with $After_t$, as the main effect of the time-invariant controls is absorbed by the ϑ_i. We also include the interaction between the adjusted score and $After_t$ to control for differential trends across different quality projects. The first-stage estimating equation associated with equation (3) is:

$$(4)\quad Acceleration_i \times After_t = \alpha + \vartheta_i + \beta f(SG) \times After_t + \theta X_{it} \times After_t + \mu_t + \epsilon_{it},$$

where SG is the "project's scoring generosity" defined as the sum of the fixed effects of the three judges that evaluated each project (see Appendix 2). For the sake of space, we refer to project's scoring generosity simply as "scoring generosity" hereafter. We present the results using bootstrap standard errors clustered at the applicant level to account for any serial correlation across applicants, and for the fact that the adjusted score is a generated regressor (Wooldridge (2002); Young (2018)).

Using scoring generosity interacted with $After_t$ to instrument for acceleration yields a consistent two-stage least squares estimate of ρ as the number of applicants grows to infinity, but is potentially biased in finite samples. This bias is the result of the mechanical correlations between an applicant's own outcomes and the estimation of that applicant's judge fixed effects. Following the parallel literature exploiting judge leniency (Kling (2006), and related papers thereafter), we address the own observation problem by using the sum of the (average) leave one out measures of judge's scoring generosity introduced in Section 2 to build our instruments for acceleration. We also estimate "leave-one-out adjusted scores" by subtracting the sum of the (average) leave one out judge fixed effects from the average score of each project. The correlation between the adjusted score and the leave-one-out adjusted score is high at 0.98 (significant at the 1% level). In unreported results, we verify that results are similar using the raw measures of scoring generosity to construct the instrument, as well as the raw measure of adjusted scores.[26]

The ρ estimate measures the local average treatment effect of the accelerator for applicants whose participation is altered by scoring generosity. Three conditions must hold to interpret these estimates as the average (local) causal impact of acceleration: (1) scoring generosity is associated with participation in the accelerator, (2) scoring generosity only impacts venture outcomes through the probability of participating in the accelerator (i.e., the "exclusion restriction"), and (3) the impact of scoring generosity on the probability of acceleration is monotonic across applicants.

The first assumption is empirically testable. Panel A in Figure 7 shows a positive and non-linear association between acceleration and scoring generosity. The figure plots average acceleration versus our leave-one-out measure of scoring generosity, and shows that the relation between these two variables is monotonic and exhibits a larger slope in the upper tail of scoring generosity. The positive association cannot be explained by applicant heterogeneity because judges were randomly assigned; and indeed Panel B in Figure 7 shows evidence of a positive association holding constant applicant quality (as measured by adjusted score). To produce Panel B in Figure 7, we classify applicants into quartiles of scoring generosity, and estimate for each quartile the distribution of acceleration over adjusted scores.[27] The figure shows that for a given adjusted score, the acceleration probability is always highest (lowest) for projects assigned to the top (bottom) quartile of scoring generosity. With this classification the non-linearity in the relation between scoring generosity and acceleration becomes apparent, as further shown in Table 7. The table shows that for any quartile of scoring generosity the probability of acceleration is always below 6% among low-quality applicants (Column 2, Table 7). Instead, for applicants in the 75th percentile of quality, the probability of acceleration decreases from 99.78% to 3.32% when we move from judges in the top to the bottom quartile of scoring generosity (Column 4, Table 7). Given the non-linear relation of scoring generosity and acceleration, we use quartiles of scoring generosity as our main explanatory variables in equation (4).[28] We discuss non-linearity further in Section 2.5, where we present estimates of marginal effects using non-parametric approaches.

---

[26] Results are available upon request; they are not reported to conserve space.

[27] Relative to a mean average score of 0.7, the breakpoints for the scoring generosity quartiles are -0.03, 0.001, and 0.05, and the max (min) scoring generosity is 0.21(-0.13). These numbers imply that projects classified in the top (bottom) quartile of judge generosity received between 0.05 and 0.21 (0.13 and 0.003) additional (fewer) points than their project fixed effects.

[28] We estimate equation (4) with f(SG) as $\sum_{i=2,3,4} Quartile_i \times After_t$, where $Quartile_i$ is a dummy indicating the *ith* quartile of scoring generosity (the left out category is the bottom quartile).

The first stage results in Table 8 show a large and precisely estimated relationship between quartiles of scoring leniency and the probability of acceleration. The results in the table show that for a given adjusted score, applicants in the top quartile of scoring generosity are 49% more likely to be accelerated than applicants in the bottom quartile of scoring generosity (Column 1). The results are similar across applicant business ideas and applicant established firms (Columns 2 and 3, respectively). We formally test the relevance of the instrument and report the *F*-test of joint significance of the quartiles of scoring generosity, showing that the instruments are not weak (Stock and Yogo (2005)).

Regarding the exclusion restriction, we argue that it is likely to hold for a number of reasons. The most natural concern of favoritism (that firms with higher growth potential were assigned the most generous teams of judges) can be ruled out by design, as judges were randomly allocated. Any remaining concerns regarding the unintentional assignment of generous judges to high-quality firms are not consistent with the patterns shown in Figure 3—i.e. projects with high adjusted scores do not systematically have higher average scores than expected. These concerns are also not consistent with the fact that observable characteristics are similar across applicants assigned to judge panels with low and high scoring generosity (see Appendix 5). Differences in the interaction between applicants and judges across applicants in different quartiles of scoring generosity are unlikely because only two of the 50 judges are ValleE staff members, the rest of the judges do not interact with participants as part of the program, and the judges' identities are not revealed to applicants throughout the process. Scoring generosity also does not measure differences in predicting ability across judges, as shown in Figure 6 (see also Section 2.1). Finally, because applicants are not made aware of their scores, nor of the generosity of their judge panel, psychological reactions are also unlikely (e.g., feelings of grandeur or depression). Ultimately, however, the assumption that scoring generosity only systematically affects applicants' performance through acceleration is fundamentally untestable, and our estimates should be interpreted with this identification assumption in mind.

For the monotonicity assumption, we summarize supporting evidence of several tests presented in Appendix 6. The monotonicity assumption implies that being assigned to a more (less) generous panel of judges does not decrease (increase) the likelihood of selection into the accelerator depending on the projects' characteristics. The monotonicity assumption would be violated if judges differ in the types of applications they grade more generously. For example, the monotonicity assumption could be invalidated if some judges score business ideas more generously than established firms. In Appendix 6 we plot scoring generosity measures that are calculated separately for four restricted subsamples: i) using only business ideas, ii) using only established firms, iii) excluding the bottom quartile projects according to adjusted score, and iv) excluding the top quartile according to the same metric. Consistent with the monotonicity assumption, we find that judges exhibit remarkably similar scoring generosity tendencies across observably different applicants. The plots show a strong correlation between the actual fixed effects and the fixed effects from the restricted samples.

2.3 Local Average Impact Results

In this section we summarize results from the IV regressions. We defer the more detailed interpretation of the results to Section 3.

Table 9 shows compelling evidence of causal impacts of acceleration. Over the first three post-application years, acceleration increases the marginal applicant's annual revenue by $66.3M COP ($20K USD) (Column 3). This effect corresponds to a 166 (130) percent increase from the rejected applicant's (average applicant) 2017 revenue of $51M COP ($40M COP).

The IV estimates the local average treatment effect (cf., Imbens and Angrist (1994)) for the applicants at the margin of selection—i.e., applicants whose selection decision was altered by the judge assignment (and would [not] have been selected but for the strictness [generosity] of the judges). Complementary analysis suggests that the local average effect is not driven by a few outliers: Appendix 8 shows a shift in the revenue distribution 3 years after application to the accelerator (2017 versus 2014) for projects at the top quartile of scoring generosity that is not evident for projects in the bottom quartile of scoring generosity.

We contrast the IV estimate with the naïve OLS estimate of equation (3) that compares average performance across participants and nonparticipants. A comparison between columns 2 and 3 in Table 9 reveals that a positive difference exists between the IV and the OLS estimates (66.31 versus 42.91). This positive difference suggests that the projects at the margin of acceptance are most sensitive to acceleration (cf. Card (2001)). We come back to this point in Section 2.5, where we explore in detail the evidence on treatment heterogeneity.

A comparison between Columns 6 and 9 in Table 9 reveals that the increase in revenue for marginal applicants is driven by established firms (Column 9) and is not significant for business ideas (Column 6). Marginal established firms had an additional annual revenue of 116 M COP ($35K USD) during 2015–2017 (or 2.4 times their initial revenue), whereas the estimate for marginal business ideas is negative, albeit not statistically significant. The results in Table 10 provide a possible explanation for this difference in estimated average impact across established firms and business ideas. The table shows that accelerated entrepreneurs that applied with ideas (and not established firms) were less likely to start a firm during the first year after the program, though they often closed that gap during the second year. The first year after treatment, 39% of the rejected applicants created a firm, while only 9% of accelerator participants did. By 2016, those numbers increased to 56% for the controls and 50% for the participants. In the third and last year, 66% and 58% of the firms created were established firms, respectively. The reason for this delay could be explained by the bootcamps, which included discussions on the value of delaying firm creation until product markets are identified. To produce these reduced form results, we regress an indicator variable of firm registration at the Chamber of Commerce against the interaction between the indicator variable for acceleration and the different year fixed effects.

In robustness checks reported in Appendix 7, we show similar evidence of causal impacts if we use survey-based information on employment and profits. In contrast, we find no evidence that acceleration leads to additional survey-based fundraising. Over the first three post-application years, only 21 applicants secured an average external financing of $25K USD, including 9 participants and 12 rejected applicants. Of those that secured external financing, 14 sourced it from bank loans, 4 from combined bank loans and equity, and 3 relied on equity only. These fundraising results are not surprising given the underdevelopment of private equity markets in Colombia, and particularly in Cali, where the first formal network of business angels and the first local private equity were only launched in 2017.

2.4 Impact Heterogeneity: Who Benefits from Acceleration?

The results so far show compelling evidence of average treatment effects among the marginal companies whose acceleration is altered by the judge assignment. One question that remains regards the types of applicants who benefit most from accelerators. Skeptics contend that these programs can only have non-transformational effects on low-quality businesses, because high growers' skills are innate and cannot be taught. No rigorous evidence exists to substantiate these contentions, as prior accelerator impact estimates cannot be extrapolated beyond small subsamples of entrepreneurs with similar growth potential, such as in studies exploiting qualifying thresholds using regression discontinuity designs (e.g., González-Uribe and Leatherbee (2018a)).

To investigate heterogeneous treatment effects across applicants' growth potential, we take advantage of the continuity of our instrument. Intuitively, our methodology exploits the selection mistakes of ValleE. By selection mistakes we mean applicants that were accepted (rejected) because they were assigned to generous (strict) judges who gave them high (low) average scores, but would not have been accepted (rejected) otherwise—that is, had the program selected participants based on adjusted scores or judges' ranks. Our goal is to estimate changes in accelerator impacts as we move from more low-growth-potential applicants to more high-growth-potential applicants. We can approximate this goal because the wide variation in the scoring generosity of the randomly assigned judges implies that the selection mistakes span different parts of the growth-potential distribution of applicants. Panel B in Figure 1 illustrates this wide span, where the "mistakenly" accepted (rejected) applicants correspond to the solid (open) dots at the left (right) of the 35th rank threshold.

To show evidence of heterogeneous accelerator impacts, we begin by running matching regressions, where we match accelerated participants and rejected applicants based on their adjusted score and observed covariates at application. By construction, accelerated applicants and rejected applicants (so-matched) differ on judge scoring generosity only (relative to the information sets of the econometrician and judges). The matching algorithm we use is kernel matching (with a radius of 0.05), a nonparametric matching estimator that uses weighted averages of all individuals in the control group to construct the matched outcome (cf. Heckman, Ichimura, and Todd (1997)). One advantage of this algorithm (over others based on one-to-one to matching such as nearest neighbor matching) is the lower variance which is achieved because more information is used. A drawback is that observations can be used that are bad matches, which we mitigate by restricting applicants to those in the common support. The results are presented in Appendix 9. The majority of participants fall within the common support, and the average absolute difference in propensity of acceleration is 0.014.

Next, we estimate individual accelerator impacts by each level of propensity score for acceleration (within the common support) as the average difference in post-application annual revenues between participants and matched applicants, where the kernel weights are used to weight the outcomes of the matched applicants (cf. Smith and Todd (2005)).[29] Appendix 9 shows that the average of the individual treatment effects is 59.10, which is (by design) very close to the local average treatment effect that we estimated in Section 2.3 using the IV approach.

Finally, we transform our data so that the individual impacts constitute the observed data subject to further modelling. We then apply nonparametric regressions to the transformed data to predict the relationship between the individual impact estimates and the propensity score of acceleration within the common support. The first derivative of this relation is then evaluated at different values of acceleration propensity using the coefficients from the nonparametric regression. We calculate standard errors using the standard deviation of the marginal effect estimates from a bootstrap procedure with 500 iterations. The identification assumption behind the marginal effect estimates is that for any given propensity for acceleration, accelerated participants and matched applicants only differ in their "judge assignment luck," and thus that, absent differences in the scoring generosity of judges, both types of companies would have had the same treatment status (i.e., both accepted or both rejected). We plot the results in Figure 8.

Figure 8 shows a large heterogeneity in impacts, with the evidence pointing to an increasing function based on an applicant's growth potential. The shape of this function sheds light on the types of applicants who benefit most from acceleration, and helps us understand the IV results from Section 2.3. The figure plots the marginal effects at different values of acceleration propensity in the common

[29] We use a symmetric, nonnegative, unimodal kernel; hence, higher weight is placed on applicants who are close in terms of propensity score of a participant and lower weight is placed on more distant observations.

support, and the bootstrapped confidence intervals. For values above/below those thresholds of the acceleration propensity score we cannot estimate marginal effects, as there are no selection mistakes to use in the estimation—i.e., no applicants with an acceleration propensity below (above) 0.35 (0.75) were mistakenly selected (rejected) by the program.

Regarding the types of projects that benefit most from acceleration, Figure 8 reveals that high-potential applicants reap the most benefits, whereas no significant impacts are visible for low-potential entrepreneurs. This pattern of impact heterogeneity runs counter to sceptics' contentions, and demonstrates that these programs can add value by boosting top growers, but not necessarily by transforming low-potential entrepreneurs. In Section 3, we discuss in detail the interpretation of these results in the wider context of constraints on high-growth entrepreneurship in developing countries and of firm capabilities' interventions.

In terms of the IV estimates, the main implication from the heterogeneity patterns in Figure 8 is that the positive LATE results in Section 2.3 come from the remarkable growth of high-potential participants, and not from any apparent performance improvements of low-potential participants that were accelerated because they were assigned to more generous judges.

Our approach is made in a similar spirit as the estimation of marginal treatment effects (MTEs) in the microeconomics literature (e.g., Heckman and Vytaclil (2005)), and particularly in the judge leniency literature (e.g., Doyle 2007). MTE estimates in our context would illustrate how the outcomes of applicants on the margin of acceleration change as we move from more strict to more generous judges. Given the increasing function of marginal effects in the acceleration propensity score of Figure 8, we expect the MTE function for revenue to be decreasing in scoring generosity. This is because the margin for relatively generous judges should entail relatively less deserving applicants, as measured by our acceleration propensity score. Consistent with this intuition, we show in unreported analysis that the MTE function for revenue conditional on adjusted scores has a decreasing slope in the probability of acceleration, as predicted by scoring generosity. One important departure between our analysis of treatment heterogeneity and MTEs' estimation in the judge leniency literature is that we exploit the multiple judge assignment per applicant to control for applicant heterogeneity using the adjusted scores from the fixed effects models estimated in Section 2.1. These controls are not possible in settings with single judge assignment, such as in most prior applications of judge leniency techniques (e.g., children's welfare, Doyle (2007); bankruptcy courts, Dobbie and Song (2015)), but are crucial in our setting as an unconditional revenue MTE function does not show the decreasing slope in the probability of acceleration. The importance of these controls is as expected given the skewness in young firm growth, as well as the ability of ValleE judges to distinguish high growers, which is captured in the adjusted scores (see Section 2.1).

2.5 Challenges in measuring program impacts on young firm growth

We end the presentation of our results by illustrating how the unobservable heterogeneity in young firms' potential affects the interpretation of estimates from different identification strategies typically used in the analysis of entrepreneurship interventions, including business accelerators. In this section, we discuss the advantages and limitations of the different approaches and compare their estimates with our preferred specifications.

Table 11 summarizes results from different estimations of equation (3)—to conserve space we do not present further details of these additional exercises. Columns 1 and 2 replicate the OLS and IV estimates from Table 9 for ease of comparison.

Column 3 in Table 11 reports the estimate from the most popular methodology used to quantify the effect of business accelerators: propensity score matching (PSM; see Bone, González-Uribe, Haley, and Lahr (2019)). To produce this estimate, we match participants with similar rejected applicants as measured by characteristics at application *only*. The number of observations decreases by design relative to those in Column 2; dependent variables include 2015–2017 revenues only, as the match parameters include revenue at application. The popularity of this method relies mostly on the fact that many of these programs are not designed to be evaluated, and therefore evaluations designed ex post must rely on constructing control groups using matching procedures based on observable predictors of growth potential (e.g., serial founder). The main drawback of these methods is their inability to control for heterogeneity in unobservable growth potential.

Column 4 in the table reports results from a second estimate based on PSM where we match applicants based on characteristics at application *and* adjusted scores too. The number of observations decreases relative to Column 3 as matches are additionally required to be in the common support of adjusted scores. This strategy uses judges' scores as a proxy of projects' growth potential that is unobservable to the econometrician, but was observable to the judges. Table 5 validates this proxy by showing that adjusted scores predict growth (even after controlling for participation), which demonstrates the predicting ability of ValleE judges. In settings where no such ability is demonstrated by judges, adjusted scores will be poor proxies of unobservable growth potential (see, for example, McKenzie and Sansone (2017)).

Finally, Column 5 presents estimates from a regression discontinuity (RD) approach, where we exploit the program's ex-ante determined capacity threshold, which implied that only the top 35 projects based on average scores were chosen. We subtract revenues at application from the dependent variable, and estimate the model over the 2015–2017 period to allow for the comparison of coefficients with the rest of the estimates in the table. The advantage of this strategy is the exogeneity of the cutoff. One of the main drawbacks is its reliance on the continuity at the threshold, which in a setting like ours does not hold due to the variation in unobservable quality near the threshold. Another drawback is the inability to directly extrapolate results beyond observations close the threshold.

There are four main insights from the results in Table 11:

First, the similarity between the estimates in Columns 1 and 3 highlights the concern that approaches using matching/selection-on-observables will be severely biased. The simple PSM does not correct the OLS bias, as participants and matched rejected applicants differ on dimensions that are unobservable to the econometrician (i.e. projects' quality).

Second, the similarity between Columns 2 and 4 highlights the intuition behind our IV approach, which is to infer impacts from outcome differences between applicants with similar potential (i.e., adjusted score and covariates), but exogenously different acceleration status due to differences in judge assignment luck.

Third, the positive difference between Columns 5 and 2 highlights the challenges in using RD to assess impacts. In our setting, the RD inflates the impact because several high-potential companies rank close to the qualifying threshold, which, as we showed in Section 2.5, are the ones that benefit the most from acceleration. Instead, the IV estimate averages out the large impacts on high-potential projects and the small effects on low-potential entrepreneurs. The main drawback, however, is external validity, as the estimates are only representative for marginal applicants, which may differ from the potential impact on other applicants in the population.

Finally, the comparative analysis between the IV and RD estimates (Columns 2 and 5, respectively) highlight the utility of exploring treatment heterogeneity in Section 2.4. The results in Figure 8 help explain the difference between the IV and RD estimates, as stemming at least partially from heterogeneity in impact for different levels of acceleration potential.

2.6 Channels of Impact

Why are there such large benefits caused by acceleration? In this section, we explore this question in several ways. First, we look at which accelerator services have the largest apparent effects according to surveys of participants based on different data cuts. We then consider other channels in addition to the accelerator services through which the program could also affect firm capabilities, such as certification effects that improve market recognition. While we cannot provide rigorous evidence of these additional mechanisms with our current data, we discuss related suggestive evidence.

Appendices 10 and 11 provide compelling evidence that customized advice and visibility, rather than the standardized business training, are the most impactful aspects of the intervention. Appendix 10 shows that, consistent with entrepreneurs' perceptions in business accelerator programs worldwide (see Roberts et al. 2017), 74% of ValleE participants selected advice and visibility as the program's most valuable aspects in follow-up surveys. In contrast, only 8% of surveyed ValleE participants described the grouped business training as a key impact driver. Regarding the business practices that where most impacted by the program, 54% of entrepreneurs reported having found a business contact thanks to the program, and 34% reported using the program to show their products/services to other businessmen that shared their interests. Appendix 11 shows evidence that impact effects are larger for entrepreneurs who would presumably benefit the least from grouped business training, such as the more educated entrepreneurs (those with a college degree or graduate studies). Impacts are also larger for projects that indicated in the baseline that strategic advice (as given by mentors and advisers) was their main constraint on growth. These effects are not visible for applicants that did not indicate that they needed strategic advice at application.

This additional evidence confirms prior findings on the outperformance of customized interventions over standardized business training programs (Karlan and Valdivia (2011); Campos et al. (2018); Ubfal et al. (2019); Bruhn, Karlan, and Schoar (2018); Lafortune, Riutor, and Tessada (2018)). It also suggests that impact mechanisms other than the accelerator services *per se* may be at play too, as customized advice and visibility can trigger network, certification, and validation effects which can affect firms' capabilities, and which have been shown to have large impacts in other contexts, such as among students at prestigious business schools (e.g., Malmendier and Lerner (2013)).[30] As is common in the literature on returns from education, with our current data we cannot rigorously distinguish these additional mechanisms, and thus we only discuss suggestive evidence (e.g., Chevalier, Harmon, Walker, and Zgu (2004)). For example, the heterogeneity of impacts discussed in Section 2.4 goes against certification or validation effects, as by definition these effects should be higher for the low-potential applicants (or homogenous across applicants of different quality). Evidence from follow-up surveys points to potential network effects, as 53% of participants reported an improved ability in finding business contacts as a consequence of participating in the program (see Appendix 10).

3. Interpretation of Results and External Validity

Overall, the results in Section 2 provide compelling evidence that alleviating constraints to firm capabilities has a first-order effect on young-firm growth, specifically by unlocking innovative entrepreneurs' potential, and not necessarily by transforming low-potential businesses. The implications are twofold: On the one hand, the results imply that high-potential entrepreneurs face barriers to growth besides financial constraints, which can be mitigated by business accelerators. On the other hand, the results also highlight the limitations of policies aimed at using business accelerators

[30] For evidence on the value of network and certification effects, see Megginson and Weiss (1991); Hsu (2004); Fafchamps and Quinn (2016); Brook et al. (2018); Cai and Szeidl (2018).

to promote firm growth. The impact of such policies would seem to rely on the deal flow of existing high-potential entrepreneurs into the ecosystem, as well as on the ability of accelerator programs to identify high-potential businesses.

In terms of interpretation, the standard "joint test" caveat in firm capabilities' interventions applies here too. That no impacts are visible in low-quality projects does not necessarily mean that capability building interventions cannot add value to projects in the left tail of the growth potential distribution, or that firm capabilities cannot be taught to these types of entrepreneurs. An alternative explanation for why impacts are not visible in low-quality applicants is that the firm capabilities provided by the business accelerator were not enough for these types of projects and more intensive interventions are needed (see Bruhn et al. (2018)).

In terms of magnitude, our impact estimates are similar to those found in evaluations of business accelerators in developing countries (e.g., Goni and Reyes (2019)). However, they are generally larger than those in similar work on business training interventions for traditional microenterprises. For example, Calderon, Cunha, and De Giorgi (2013) and de Mel, McKenzie, and Woodruff (2014) find a 20% and a 41% increase in revenues (within 1 and 1.5 years), whereas we estimate an increase of 166% in annual revenues over the first three post-application years for rejected applicants. Buttressed by the heterogeneous impact patterns and the survey evidence, the differences in magnitude relative to interventions on microenterprises are likely explained by two distinct factors: the high-growth focus of our sample relative to subsistence enterprises and the additional provision of more intensive customized advice and visibility, which is not common in business training interventions. This last factor can also explain why prior work on business plan competitions find that grouped training has little effect for high-potential firms (Fafchamps and Woodruff (2016); McKenzie (2018)), as the short-term nature of business plan competitions also prevents the inclusion of more intensive advice and visibility in these programs.

In terms of external validity, several aspects of our setting suggest that these results are not only confined to ValleE's experience. For starters, ValleE is very similar to the average ecosystem accelerator on many dimensions. For example, the location of the program outside the capital city of Colombia is a common trait among ecosystem accelerators. Roughly 38% of these programs are located in underdeveloped regions. 40% are in the United States, but outside Silicon Valley, Massachusetts, New York, or Washington D.C.; the rest are in Europe, but are typically not in the capital cities. In terms of services, those offered at ValleE are similar to the offerings of these programs worldwide (cf. Clarysse, Wright, and Van Hove (2015)). This is not to say that some differences between ValleE and other ecosystem accelerators do not exist. Perhaps the most distinguishing features of the program include its access to highly qualified staff, mentors, and judges. We are also careful to emphasize the differences between average applicants to ValleE and other ecosystem accelerators, as mentioned in

Section 2. We argue that the external validity of our findings is likely confined to other ecosystem accelerators that attract young businesses with traction and have access to high-quality resources, including staff, mentors, and judges.

What do we learn from our findings about high-growth young firms, and why are they missing in developing countries? Our findings provide evidence that firm capability constraints obstruct the growth of some high-potential entrepreneurs in developing countries, and that business accelerators can help identify these constrained high-growth entrepreneurs and boost their growth. Using a back-of-the-envelope-calculation based on our findings, we show that marginal applicants on average reach about 3 times their initial revenue by age 4. The 4-year lifecycle growth rate of marginal applicants is roughly double that of Colombian top growers, which instead approximately increase 1.5 times their initial revenue (see Figure 3 in Eslava and Haltiwanger (2018)). Our findings thus show that the accelerator unlocked the gazelle potential of marginal applicants, and brought their lifecycle growth closer to the rates exhibited by US top growers.[31] The back-of-the-envelope calculation is as follows: by 2017, rejected applicants increase their initial revenue 1.8 times, from \$22M COP to \$40M COP. Our IV estimates imply that marginal applicants grow 1.66 times more than rejected applicants, so roughly 3 (1.66×1.8) times their revenue at baseline.

4. Robustness Checks

The main concern with the results in Section 2 is the potential biases from the small cross-section; namely, the possibility that our effects pick up the 5% chance we would see an effect when no such effect exists. This concern is minimized by the robustness of results to using different measures of scoring generosity, sets of controls, and methodologies, as shown in Section 2; as well as to different outcomes variables (see Appendix 7).

Nevertheless, to further address this concern we use a randomization inference (RI) approach that conducts exact finite sample inference and which remains valid even when the number of observations is small (cf. Rosenbaum (2002)). This approach is somewhat similar to the bootstrap approach, but is different in spirit. In particular, when estimating bootstrapped p-values the econometrician is looking to address her uncertainty over the specific sample of the population she drew, while randomization inference instead helps the econometrician address her uncertainty over which units within her sample are assigned to the treatment.

There are two steps to our RI approach. In the first step we identify a subsample of applicants where we argue treatment can be assumed to be "as good as randomly assigned." We select this

[31]Eslava, Haltiwanger, and Pinzon (2019) show that U.S. gazelles roughly double the growth rates of Colombian top growers (see Figure 5).

subsample by taking (i) all accelerated applicants with lower adjusted scores than the highest adjusted score of a nonaccelerated applicant and (ii) all nonaccelerated applicants with adjusted scores higher than the lowest adjusted score of the accelerated projects. Overall, we find 62 projects that match our definition, 28 being ideas and 34 being established businesses. For this subsample, we estimate the treatment effect as the relative increase in revenue for participants versus nonparticipants, and then test the sharp null hypothesis of no treatment effect by applying standard exact randomization inference tools (see, among others, Rosenbaum (2002), (2010); Imbens and Rosenbaum (2005)). In particular, we scramble the data 5,000 times, each time randomly assigning different companies to be placebo participants. For each permutation we estimate a placebo effect equal to the average conditional difference between placebo participants and placebo rejected applicants, as estimated using equation (3). We then compare the placebo effect with our estimated treatment effect and keep track of the number of times that our estimate is bigger (in absolute value) than the placebo difference. We then say that we reject the null of no treatment effect if in more than 95% of the permutations our treatment estimate (absolute value) is smaller than the placebo effects. The results are summarized in Table 12 and illustrated in Figure 9.

The results in Table 12 suggest that our main results (Table 9) are unlikely to be driven by small sample bias: our placebo effects for established firms are larger than our real estimates in only 3.5% of the permutations (Column 2, Panel B). The identification assumptions behind this randomization inference test are that the distribution of the score is the same for all observations in the subsample and that initial outcome variables are statistically similar between groups. We present supportive evidence in Panel A of Table 12, where we show that adjusted scores and revenue among treated and nontreated entrepreneurs are similar in the subsample. The differences in sectorial experience and in initial number of employees (for established firms) are controlled for in the regressions presented in Panel B.

5 Conclusion

We show compelling evidence that alleviating constraints to firm capability has a first-order effect on average high growth, particularly by unlocking innovative entrepreneurs' potential rather than by transforming average-quality projects. Our research laboratory is an accelerator in Colombia that provides participants with grouped training, customized advice and visibility, but, *no cash*. We measure firm growth using administrative data. Our empirical strategy exploits the random allocation of applicants to expert judges that differ in their scoring generosity. Our approach is novel in that it provides an identification strategy that can work even when programs are trying to choose the best firms, precisely by showing the amount of randomness that comes from human decision-making in judging success. This approach is thus potentially applicable to many more efforts to evaluate these types of programs worldwide. We estimate that over the first three post-application years, acceleration increases the marginal applicant's annual revenue by 166 percent relative to rejected applicants.

Marginal applicants are those whose selection decision is altered by the judge assignment and would (not) have been selected but for the strictness (generosity) of their randomly assigned judges. We provide the first exploration of accelerator treatment effect heterogeneity by exploiting the continuity of our instrument, and the wide span of selection mistakes along the distribution of applicant growth potential. We show that impacts are increasing in firm growth potential, which debunks skeptics' contentions that capability building interventions cannot influence high-growth entrepreneurship because entrepreneurial skills are innate. Using a back-of-the-envelope calculation, we quantify revenue improvements of 31%–40% for the accelerator had it accounted for judge heterogeneity in scoring generosity and thus selected higher-potential firms (see Appendix 12). The results demonstrate that accelerators add value in developing countries by identifying and boosting top growers. It follows that these programs could contribute to solving the development problem by boosting high-potential entrepreneurs, and thus reducing the shortage of high-growth entrepreneurship between developed and emerging economies.

References

Aizer, Anna, and Joseph Doyle, Jr., 2013, Juvenile incarceration, human capital, and future crime: Evidence from randomly-assigned judges, NBER Working Paper No. 19102.

Ahmad, Nadim and Ditte Rude Petersen, 2007, High-growth enterprises and gazelles – Preliminary and summary sensitivity analysis, OECD-FORA, Paris.

Autio, Ekko, Pia Arenius, and Hannele Wallenius, 2000, Economic impact of gazelle firms in Finland. Working Papers Series 2000:3, Helsinki University of Technology, Institute of Strategy and International Business, Helsinki.

Autor, David, and Susan Houseman, 2010, Do temporary-help jobs improve labor market outcomes for low-skilled workers? Evidence from "Work First," *American Economic Journal: Applied Economics* 2, 96-128.

Bertrand, Marianne, and Antoinette Schoar , 2003, Managing with style: The effect of managers on firm policies. *Quarterly Journal of Economics* 118, 1169-1208.

Birch, David L., Anne Haggerty, and William Parsons, 1995, *Who's Creating Jobs?* (Cognetics Inc: Boston, MA).

Birch, David L. and James Medoff, 1994, Gazelles. In L.C. Solmon & A.R. Levenson, (Eds), *Labor Markets, Employment Policy and Job Creation*, pp. 159-167. (Westview Press: Boulder, CO and London).

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, 2013, Does management matter? Evidence from India, *Quarterly Journal of Economics* 128, 1-51.

Bloom, Nicholas and John van Reenen, 2010, Why do management practices differ across firms and countries?, *Journal of Economic Perspectives* 24, 203-24.

Bone, Jonathan, Olivia Allen, and Christopher Haley, 2017, Business incubator and accelerators: The national picture, Department for Business, Energy and Industrial Strategy, Nesta, BEIS Research Paper Number 7.

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar, 2010, What capital is missing in developing countries?, *The American Economic Review* 100, 629-33.

Calderon, Gabriela, Jesse M. Cunha, and Giacomo De Giorgi, 2013, Business literacy and development: Evidence from a randomized controlled trial in rural Mexico, NBER Working Paper No. 19740.

Card, David, 2001, Estimating the return to schooling: Progress on some persistent econometric problems, *Econometrica* 69, 1127-60.

Clarysse, Bart, Mike Wright, and Jonas Van Hove, 2015, A look inside accelerators: Building businesses. Research Paper, Nesta, London, UK.

Chang, Tom, and Antoinette Schoar, 2013, Judge specific differences in chapter 11 and firm outcomes, MIT Working Paper.

Cohen, Susan G., and Yael V. Hochberg, 2014, Accelerating startups: The seed accelerator phenomenon, Working Paper.

Deschryvere, Matthias, 2008, High-growth firms and job creation in Finland, Discussion Paper No. 1144, Research Institute of the Finnish Economy (ETLA), Helsinki.

Di Tella, Rafael, and Ernesto Schargrodsky., 2013, Criminal recidivism after prison and electronic monitoring, *Journal of Political Economy* 121, 28-73

Dobbie, Will, Jacob Goldin, and Crystal S. Yang, 2018, The effects of pre-trial detention on conviction, future crime, and employment: Evidence from randomly assigned judge, *American Economic Review* 108, 201-240.

Doyle, Joseph, 2007, Child protection and child outcomes: Measuring the effects of foster care, *American Economic Review* 97, 1583-1610.

Doyle, Joseph, 2008, Child protection and adult crime: Using investigator assignment to estimate causal effects of foster care, *Journal of Political Economy* 116, 746-770.

Mel, Suresh de, David McKenzie, and Christopher Woodruff, 2008, Returns to capital in microenterprises: Evidence from a field experiment, *The Quarterly Journal of Economics* 123, 1329–72.

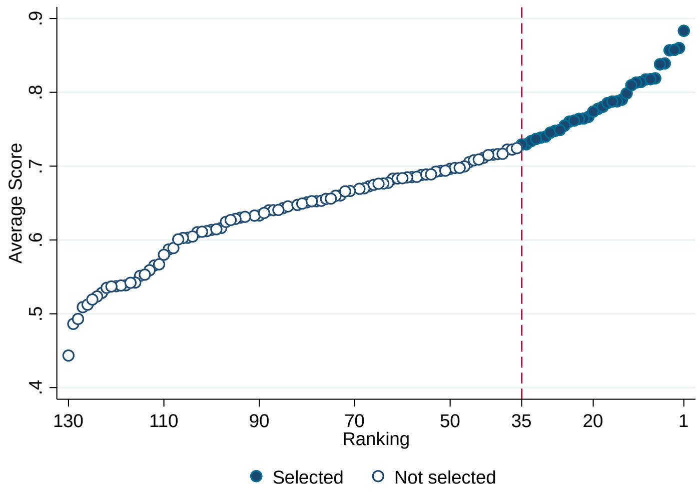
———. 2014. Business training and female enterprise start-up, growth, and dynamics: Experimental evidence from Sri Lanka. *Journal of Development Economics* 106, 199–210.

Dobbie, Will, and Jae Song, 2015, Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection, *American Economic Review* 105, 1272-1311.
Drexler, Alejandro, Greg Fischer, and Antoinette Schoar, 2014, Keeping it simple: Financial literacy and rules of thumb, *American Economic Journal: Applied Economics* 6, 1–31.
Ewens, Michael, and Matthew Rhodes-Kropf, 2015. Is a VC partnership greater than the sum of its partners?, *The Journal of Finance* 70, 1081-1113
Fafchamps, Marcel, and Christopher M. Woodruff, 2016, Identifying gazelles: expert panels vs. surveys as a means to identify firms with rapid growth potential. Policy Research Working Paper, World Bank, WPS 7647.
Fan, Jianqing, and Irene Gijbels, 1996, *Local Polynomial Modelling and Its Applications*. (Chapman and Hall: London).
Fee, C. Edward, Charles J. Hadlock, and Joshua R. Pierce, 2013, Managers with and without style: Evidence using exogenous variation, *The Review of Financial Studies* 26, 567–601
Fehder, Daniel C., and Yael V. Hochberg, 2014, Accelerators and the regional supply of venture capital investment, Working Paper.
French, Eric, and Jae Song, 2011, The effect of disability insurance receipt on labor supply, Federal Reserve Bank of Chicago Working Paper WP-2009-05.
Goldfarb, Brent, David Kirsch, and David A. Miller , 2007, Was there too little entry during the Dot Com Era? *Journal of Financial Economics* 86, 100–144.
González-Uribe, Juanita, and Michael Leatherbee, 2018a, The effects of business accelerators on venture performance: Evidence from Start-Up Chile, *Review of Financial Studies* 31, 1566-1603.
González-Uribe, Juanita, and Michael Leatherbee, 2018b, Selection issues. In 'Accelerators.'Editors: Mike Wright, Imperial College Business School and Israel Drori, VU, Amsterdam, 2018.
González-Uribe, Juanita, Hu, Zhongchen and Koudjis, Peter, 2019, Corporate Accelerators, Working paper.
Hall, Robert E., and Susan E. Woodward, 2010, The burden of the nondiversifiable risk of entrepreneurship, *American Economic Review* 100, 1163-94.
Haltiwanger, John C., Ron S. Jarmin, and Javier Miranda, 2013, Who creates jobs? Small versus large versus young, *The Review of Economics and Statistics* XCV, 347–61.
Heckman, James, Hidehiko Ichimura, and Petra E. Todd, 1997, Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme, *Review of Economic* Studies 64, 605-654
Henrekson, Magnus, and Dan Johansson, 2008, Gazelles as job creators—A survey and interpretation of the evidence. IFN Working Paper No. 733
Imbens, Guido W., and Joshua D. Angrist, 1994, Identification and estimation of local average treatment effects, *Econometrica* 62, 467–75.
Imbens, Guido W., and Paul Rosenbaum , 2005, Randomization inference with an instrumental variable, *Journal of the Royal Statistical Society* Series A 168, 109-126.
Kahneman, Daniel, and Gary Klein, 2009, Conditions for intuitive expertise: A failure to disagree, *American Psychologist* 64, 515–26.
Karlan, Dean and Maria Valdivia, 2011, Teaching entrepreneurship: Impact of business training on microfinance clients and institutions. *The Review of Economics and Statistics* 93, 510-527
King, Elizabeth M., Jere R. Behrman, 2009, Timing and duration of exposure in evaluations of social programs, *The World Bank Research Observer* 24, 55-82, https://doi.org/10.1093/wbro/lkn009
Kirchhoff, Bruce A., 1994, Entrepreneurship and Dynamic Capitalism. (Praeger: Westport, CT).
Kling, Jeffrey R., 2006, Incarceration length, employment, and earnings, *American Economic Review* 96, 863-876.
Klinger, Bailey and Matthias Schundeln, 2011, Can entrepreneurial activity be taught? Quasi-experimental evidence from central America. *World Development* 39, 1592–1610.
Lafortune, Jeanne, Julio Riutort, and José Tessada, 2018, Role models or individual consulting: The impact of personalizing micro-entrepreneurship training, *American Economic Journal: Applied Economics* 10, 222-45.
Landström, Hans, 2005, *Pioneers in Entrepreneurship And Small Business Research* (Springer: Berlin).

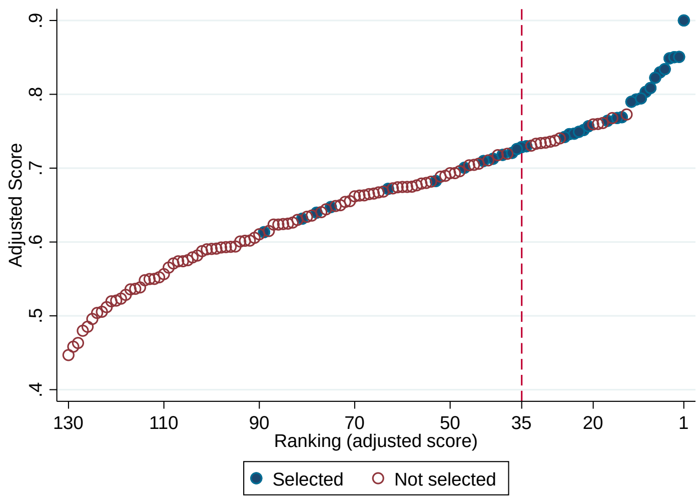
Lerner, Josh and Ulrike Malmendier, 2013, With a little help from my (random) friends: Success and failure in post-business school entrepreneurship, *Review of Financial Studies, Society for Financial Studies* 26, 2411-2452.
Mano, Yukichi, John Akoten, YutakaYoshino, and Tetsushi Sonobe, 2011, Teaching KAIZEN to small business owners: an experiment in a metalworking cluster in Nairobi, Working paper.
Maestas, Nicole, Kathleen Mullen, and Alexander Strand, 2013, Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt, *American Economic Review* 103, 1797-1829.
McKenzie, David, 2017, Identifying and spurring high-growth entrepreneurship: Experimental evidence from a business plan competition. *American Economic Review* 107, 2278-2307.
McKenzie, David J., and Dario Sansone, 2017, Man vs. machine in predicting successful entrepreneurs: evidence from a business plan competition in Nigeria. Policy Research Working Paper Series 8271, The World Bank.
McKenzie, David and Christopher Woodruff, 2008, Experimental evidence on returns to capital and access to finance in Mexico, *The World Bank Economic Review* 22, 457-82.
——— 2014, What are we learning from business training and entrepreneurship evaluations around the developing world? *World Bank Research Observer* 29, 48-82.
Nanda, Ramana, 2006, Financing high-potential entrepreneurship, *IZA World of Labor 2016*, 252.
Puri, Manju and Rebecca Zarutskie, 2012, On the life cycle dynamics of venture capital and non venture-capital financed firms, *The Journal of Finance* 67, 2247-93.
Rosenbaum, Paul R., 2002, *Observational Studies 2nd Edition*, (Springer: New York), doi:10.1007/978-1-4757-3692-2
Rosenbaum, Paul R, 2010, *Design of Observational Studies*, (Springer-Verlag: New York).
Shanteau, James, 1992, Competence in experts: The role of task characteristics, *Organizational Behavior and Human Decision Processes* 53, 252–62.
Smith, Jeffrey, and Petra Todd, 2005, Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125, 305-353.
Stock, James, Motohiro Yogo, 2005, Testing for weak instruments in linear IV regression. In: *Identification and Inference for Econometric Models*. (Cambridge University Press: New York), pp. 80-108.
Wooldridge, Jeffrey M., 2002, *Econometric Analysis of Cross Section and Panel Data* (MIT Press, Cambridge, MA).
Young, Alwyn, 2018, Consistency without inference: Instrumental variables in practical application, Working paper, LSE.
Xie, Yu, Jennie Brand, and Ben Jann, 2012, Estimating heterogeneous treatment effects with observational data, *Sociological Methodology* 42, 314-347.

Figure 1. Distribution of Applicant Scores and Selection into the Accelerator

Panel A—Average Scores

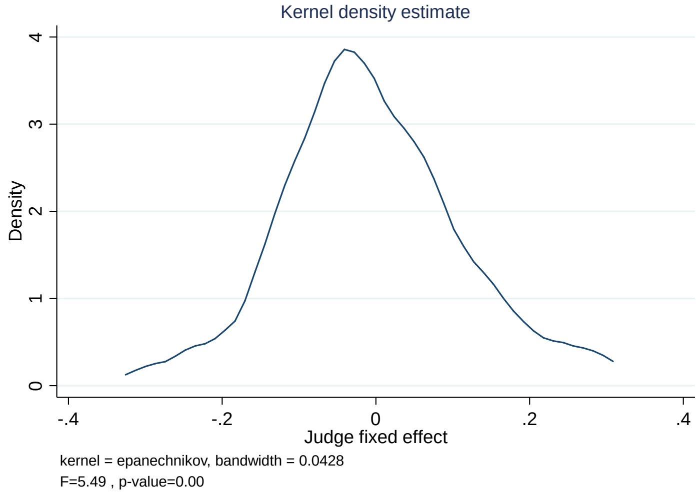


Panel B—Adjusted Scores



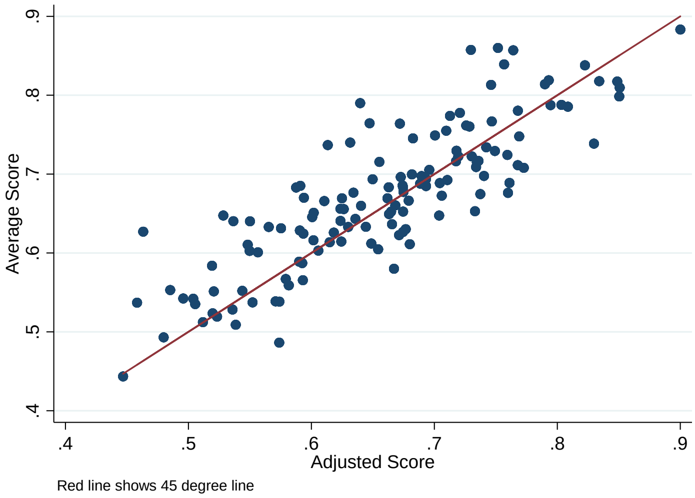
Panel A plots average scores against rankings based on the average score. Panel B plots adjusted scores, estimated as projects' fixed effects from equation (1), against rankings based on the adjusted score. In each panel, each dot represents an applicant; the solid (open) dots indicate the applicants that were (were not) selected into the accelerator.

Figure 2. Distribution of Judges' Fixed Effects



This figure plots the distribution of the estimated judge fixed effects from equation (1), which regresses project scores (by individual judges) against applicant fixed effects and judge fixed effects. Each project was evaluated by 3 randomly selected judges. Judges evaluated an average of 8 projects. The table reports the statistics of an *F*-test showing that the judge fixed effects are jointly significant (*p*-value of 0.00).

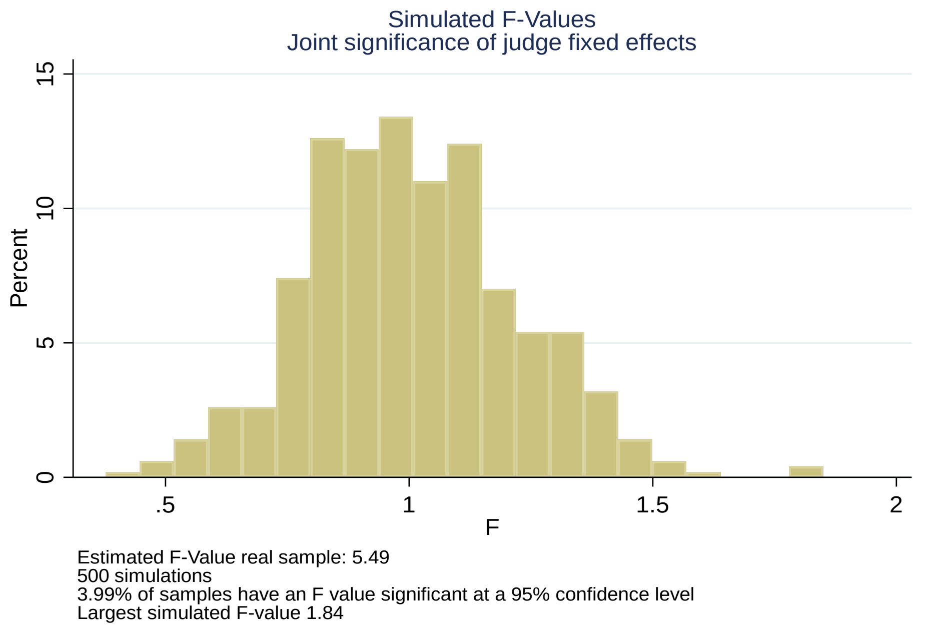
Figure 3. Average Scores and Adjusted Scores



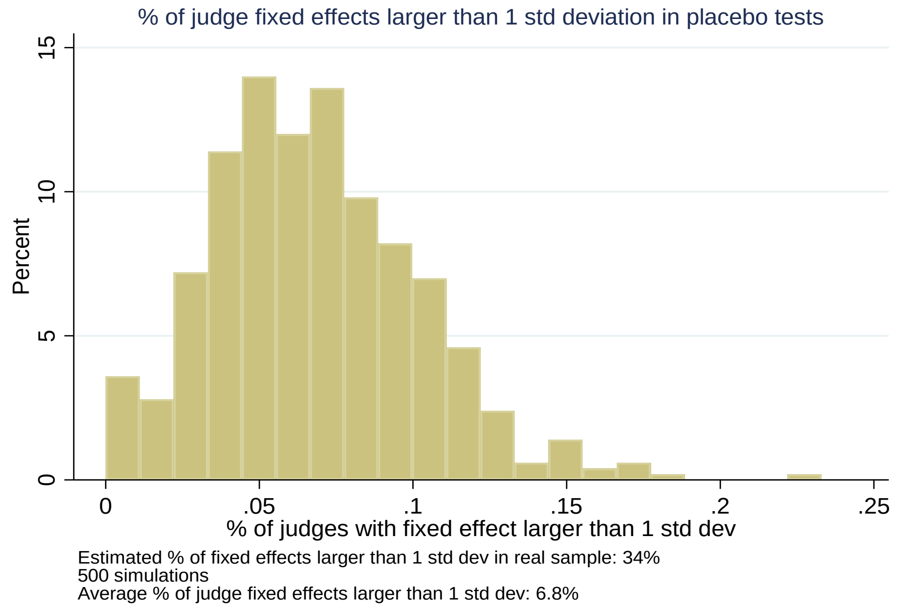
This figure plots average scores against adjusted scores. Each dot represents an applicant. The red line shows the 45-degree line. Applicants with adjusted scores above the 45-degree line were "lucky" in that they drew a generous judge panel, while applicants with average scores below the 45-degree line were "unlucky" and drew a strict judge panel. The correlation between average scores and adjusted scores is 0.825.

Figure 4. Placebo Assignment of Judges' Scores

Panel A—Distribution of *F*-values

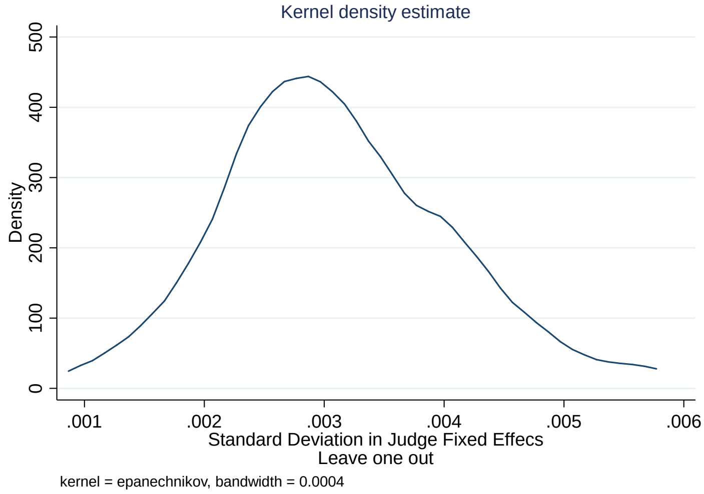


Panel B—Fixed Effects One Standard Deviation Above/Below Project Effect



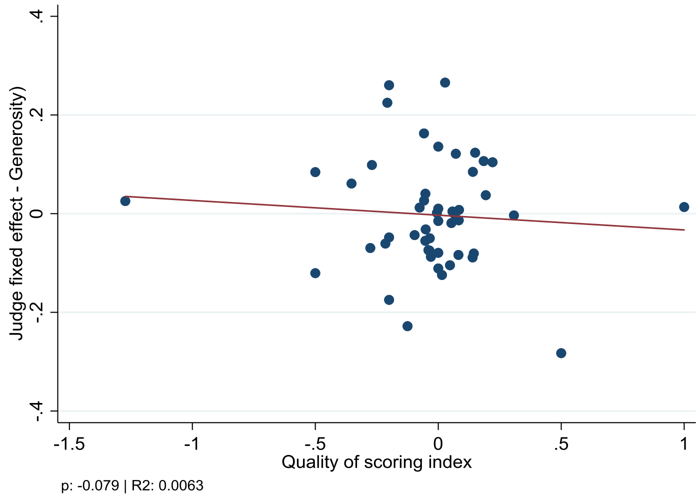
This figure plots the distribution of *F*-tests on the joint significance of the judge fixed effects in 500 placebo assignments.

Figure 5. Standard Deviation of Judge Fixed Effects (Per Judge)



This figure plots the standard deviation of all the "leave-one-out" estimates of the judge fixed effects per judge. For each judge, we estimate 135 judge fixed effects. We produce each estimate by sequentially leaving out of the sample one of the projects. Each judge has more leave-one-out fixed effect estimates than scored projects. This is because the estimated fixed effect of a given judge A from equation (1) varies as we leave out the projects she evaluated, but also as we leave out the projects of other co-judges that judge A did not also evaluate. By co-judges, we mean judges with whom judge A independently co-evaluated at least one project. The average standard deviation per judge of the leave-one-out fixed effects is 0.003 and the maximum is 0.006.

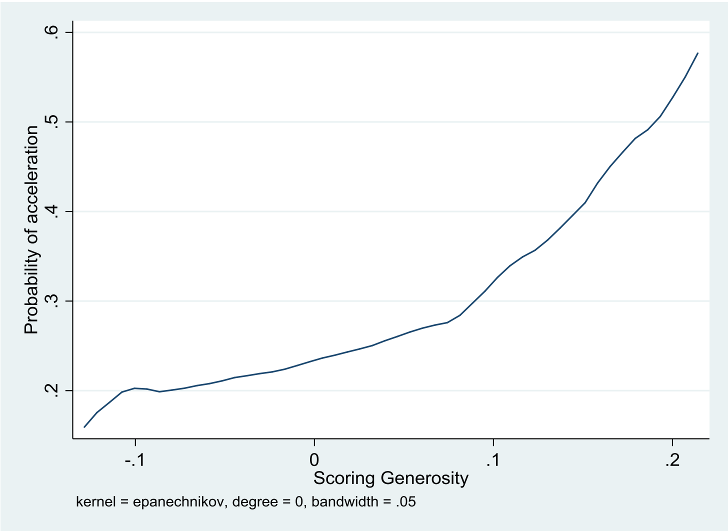
Figure 6. Scoring Generosity and Ranking Ability of Judges



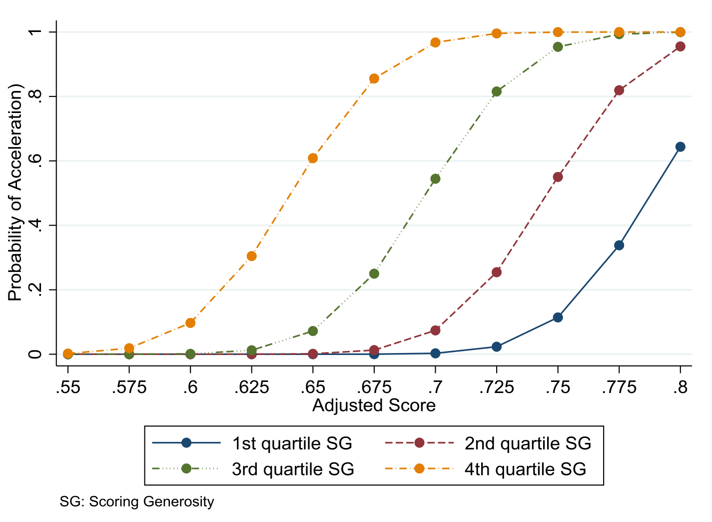
This plot is a scatter plot of judges' generosity and ranking ability. We measure judges' ranking ability using the correlation between a "judge's rank" and "actual rank." To produce this correlation, for every judge we rank the companies she evaluated based on (i) 2017 revenue ("actual ranking") and (ii) the judge's score ("judge's rank")

Figure 7. Acceleration Probability and Scoring Generosity

Panel A—Acceleration Probability and Generosity

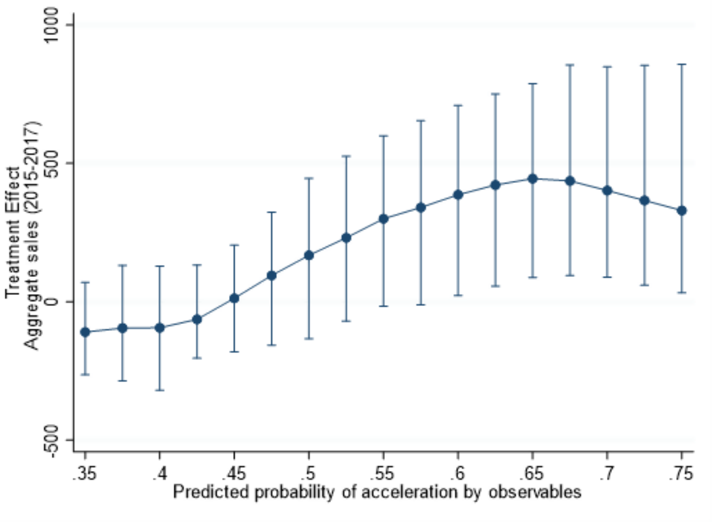


Panel B— Acceleration Probability and Generosity, by Quartiles of Adjusted Score



Panel A plots the probability of acceleration against adjusted score. Panel B plots the probability of acceleration against adjusted score by each quartile of scoring generosity. The top (bottom) quartile of scoring generosity corresponds to the most (least) generous judge panels.

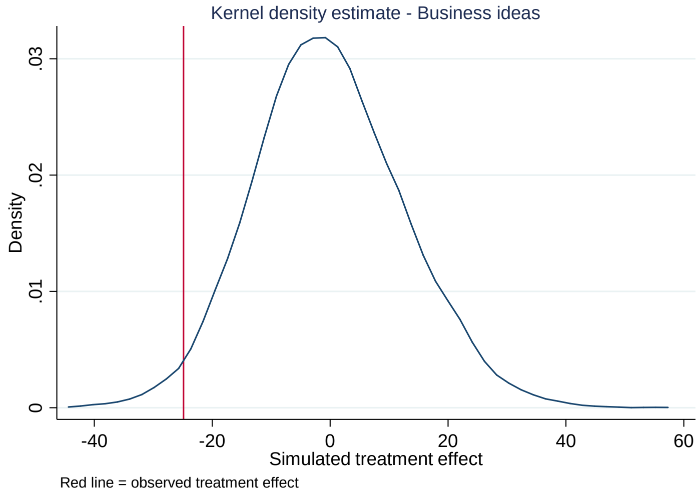
Figure 8. Heterogeneous Acceleration Effects



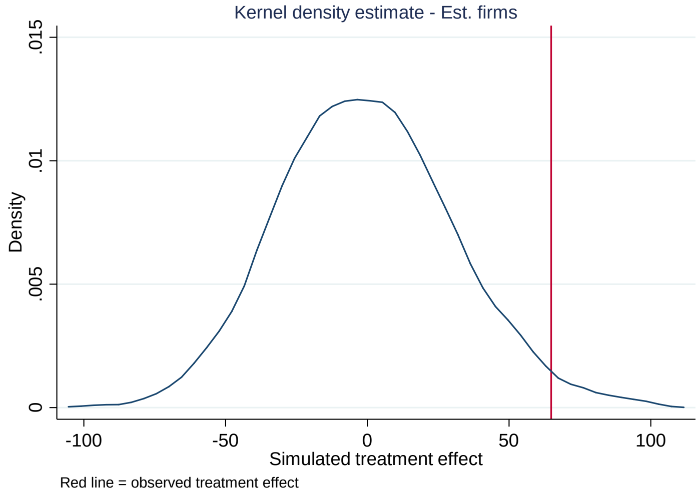
This figure plots the marginal effects of acceleration at different percentiles of applicants' acceleration propensity in the common support. For all other percentiles of the acceleration propensity score we cannot estimate marginal effects, as there are no selection mistakes to use in the estimation—i.e., no applicants with an acceleration propensity below (above) 0.35 (0.75) were mistakenly selected (rejected) by the program. We calculate standard errors using the standard deviation of the marginal effect estimates from a bootstrap procedure with 500 iterations.

Figure 9. Randomization Inference

Panel A. Business Ideas



Panel B. Established Firms



This figure plots results from the randomization inference exercise. Panel A (B) plots the distribution of the estimated acceleration effects from the 5,000 placebo assignments for the applicants that applied as business ideas (established firms).

Table 1. Sample Composition

Variable	All Sample			Business Ideas	Established Firms
	Mean	Min	Max	Mean	Mean
Gender: Male	79%	0	1	75%	84%
Education: High school	12%	0	1	17%	6%
Education: Technical degree	21%	0	1	22%	21%
Education: College	52%	0	1	39%	67%
Education: Masters or PhD	15%	0	1	22%	6%
Location: Cali	85%	0	1	88%	83%
Motivation: Have stable income	12%	0	1	13%	11%
Motivation: Own boss	1%	0	1	0%	2%
Motivation: Business opportunity	87%	0	1	88%	87%
Dedication: Sporadic	6%	0	1	10%	2%
Dedication: Half-time	21%	0	1	25%	17%
Dedication: Full-time	73%	0	1	65%	81%
Sector experience (years)	5.6	0	30	4.7	6.6
Serial entrepreneur	61%	0	1	61%	62%
Has entrepreneurial team	88%	0	1	85%	92%
# people on team	3.0	1	10	2.8	3.3
Sector: Agriculture	16%	0	1	13%	19%
Sector: Manufacturing	21%	0	1	24%	17%
Sector: Water and Electricity	3%	0	1	4%	2%
Sector: Construction	3%	0	1	3%	3%
Sector: Commerce	2%	0	1	1%	3%
Sector: Services	56%	0	1	56%	56%
Participated in other contests	59%	0	1	56%	63%
% Established Firms	47%	0	1	0%	100%
Year founded (established firms)	2013	2010	2015	.	2013
Revenue 2013 (million pesos)	10.62	0	290	1.27	21.48
Revenue 2014 (million pesos)	25.80	0	300	4.61	50.01
Total employees 2014	4.0	0	25	2.7	5.6
Observations		135		72	63

The table presents the composition of the sample and selected summary statistics of the variables in the application forms. The sample includes all 135 applicants that were evaluated by judge panels. The subsample of established firms (business ideas) corresponds to applicants that at the time of the application had (had not) registered as a business with the Chamber of Commerce.

Table 2. Differences Between Accelerated and Non-Accelerated Applicants

Variable	Business Ideas			Established Firms		
	Rejected	Accelerated	*P*-value Diff in means	Rejected	Accelerated	*P*-value Diff in means
Gender: Male	72%	87%	0.25	81%	90%	0.39
Education: High school	19%	7%	0.25	7%	5%	0.77
Education: Technical degree	21%	27%	0.65	28%	5%	0.04**
Education: College	39%	40%	0.92	58%	85%	0.04**
Education: Masters or PhD	21%	27%	0.65	7%	5%	0.77
Location: Cali	84%	100%	0.10	84%	80%	0.72
Motivation: To have stable income	11%	20%	0.33	12%	10%	0.85
Motivation: Own boss	0%	0%	.	2%	0%	0.50
Motivation: Opportunity	89%	80%	0.33	86%	90%	0.67
Dedication: Sporadic	11%	7%	0.66	2%	0%	0.50
Dedication: Half-time	25%	27%	0.87	16%	20%	0.72
Dedication: Full-time	65%	67%	0.90	81%	80%	0.90
Sector experience (years)	5.2	2.9	0.14	5.2	9.7	0.00***
Serial entrepreneur	53%	93%	0.00***	51%	85%	0.01***
Has entrepreneurial team	81%	100%	0.07*	91%	95%	0.56
# people on team	2.7	3.1	0.39	3.3	3.2	0.78
Sector: Agriculture	11%	20%	0.33	16%	25%	0.42
Sector: Manufacturing	28%	7%	0.08*	19%	15%	0.73
Sector: Water and Electricity	2%	13%	0.05**	0%	5%	0.14
Sector: Construction	2%	7%	0.31	0%	10%	0.04**
Sector: Commerce	2%	0%	0.61	5%	0%	0.33
Sector: Services	56%	53%	0.85	60%	45%	0.26
Participated in other contests	51%	73%	0.12	58%	75%	0.20
% Established Firms	0%	0%	.	100%	100%	.
Year founded (est. firms)	.	.	.	2013	2013	0.16
Revenue 2013 (million pesos)	1.25	1.34	0.96	13.37	39.84	0.07
Revenue 2014 (million pesos)	3.84	7.53	0.32	47.54	55.34	0.70
Total employees 2014	2.6	2.9	0.66	5.9	4.9	0.42

The table reports differences between accelerated and non-accelerated applicants, separately for established firms and business ideas. .*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 3—Summary Statistics

Variable	All Sample		
	N	Mean	SD
Sex: Male	135	79.3%	0.407
Location: Cali	135	85.2%	0.357
Sectoral experience (years)	135	5.59	5.643
Serial entrepreneur	135	61%	0.488
Has entrepreneurial team	135	88%	0.324
Motivation: To have stable income	135	12%	0.324
Motivation: Own boss	135	1%	0.086
Motivation: Business opportunity	135	87%	0.333
Education: High school	135	12%	0.324
Education: Technical degree	135	21%	0.412
Education: College	135	52%	0.502
Education: Masters or PhD	135	15%	0.357
Average score	135	0.67	0.090
Adjusted score	135	0.66	0.094
Revenue 2013	135	10.62	37.22
Revenue 2014	135	25.80	56.14
Profits 2014	135	8.20	15.90
Total employees 2014	135	4.03	3.93
Revenue 2015	135	51.66	124.25
Profits 2015	104	14.04	44.42
Total employees 2015	104	5.35	6.90
Fundraising 2015	104	0.086	0.283
Revenue 2016	135	58.32	128.77
Profits 2016	92	16.65	39.86
Total employees 2016	92	4.68	4.65
Fundraising 2016	92	0.077	0.268
Revenue 2017	135	50.64	118.61
Profits 2017	86	8.73	21.62
Total employees 2017	86	4.09	4.88
Fundraising 2017	86	0.135	0.345

The table presents summary statistics of the main variables used in the analysis. The upper panel includes variables from the applications. The lower panel includes performance variables constructed using the application response (data before 2015), survey responses (employees, revenues, profits and fundraising 2015–2017) and the Colombian business registry (revenues 2015–2017).

Table 4 Judges' Ranks and Project Growth Cuts

	(1) Actual Rank	(2) Actual Rank
Judge's Rank	0.295***	0.061***
	(0.0239)	(0.0231)
Judge Fixed Effect	No	Yes
Observations	405	405
R-squared	0.266	0.575

This table shows results from simple regressions of judges' ranks against projects' actual ranks. In column 2 we include judge fixed effects. To implement this regression, for every judge we rank the companies she evaluated based on (i) 2017 revenue ("actual rank") and (ii) the judge's score ("judge's rank"). The total number of observations is 405, as three different judges evaluated each of the 135 applicants. Standard errors are clustered at the applicant level and bootstrapped. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5. Applicant Scores and Project Growth

Panel A-Regression Results

	(1)	(2)	(3)	(4)	(5)	(6)
Adjusted score × After		86.16**	87.33**			
		(37.46)	(35.42)			
Adjusted score		-5.991	-52.37	45.54**		
		(19.12)	(64.64)	(17.58)		
Average score ×After					36.18	
					(45.11)	
Average score					-28.44	-6.861
					(28.12)	(22.42)
Location: Cali	22.59***	20.98*	20.87*	20.85*	22.81*	22.76*
	(7.498)	(12.18)	(12.10)	(12.09)	(11.66)	(12.02)
Gender (1=Male)	15.34**	15.18	13.48	15.28	15.44	15.50
	(6.883)	(10.62)	(8.744)	(10.43)	(10.58)	(10.50)
Has entrepreneurial team	-0.229	-4.479	-2.905	-4.485	-0.275	-0.271
	(8.387)	(8.595)	(8.274)	(7.534)	(7.243)	(7.008)
Serial entrepreneur	9.480	5.631	4.213	5.698	9.586	9.613
	(7.307)	(11.32)	(11.83)	(11.39)	(11.26)	(11.51)
Sector experience	-0.255	-0.293	-0.309	-0.301	-0.263	-0.266
	(0.763)	(0.970)	(0.981)	(0.888)	(0.891)	(0.809)
Motivation: Own boss	-30.63	-28.96	-23.49	-29.13	-31.62	-31.65
	(22.33)	(20.36)	(23.35)	(20.79)	(21.82)	(22.68)
Motivation: Business opportunity	19.95***	20.56	22.38	20.69	19.94	20.00
	(6.365)	(13.70)	(14.63)	(14.07)	(14.32)	(14.67)
Education: Technical degree	-17.65**	-17.37	-19.37	-17.25	-17.49	-17.43
	(8.251)	(12.79)	(12.95)	(10.90)	(12.15)	(11.14)
Education: College	9.047	7.778	5.143	7.822	9.490	9.532
	(8.177)	(10.66)	(8.757)	(8.335)	(11.11)	(9.881)
Education: Masters or PhD	21.27*	19.95	20.20	19.93	21.49	21.48
	(11.75)	(20.90)	(22.22)	(20.80)	(22.01)	(20.68)
Constant	-49.64**	-37.35	-9.736	-71.08***	-30.56	-45.60**
	(22.65)	(22.06)	(32.73)	(23.79)	(18.30)	(17.76)
Observations	675	675	675	675	675	675
R-squared	0.209	0.212	0.215	0.210	0.210	0.209
Control for Acceleration			Yes			

Panel B—Shapley Owen Participation

	(1) Observables	(2) Observables + *Adjusted score*	(3) Observables + *Average score*
Firm's age	61.5%	59.7%	61.0%
Entrepreneur characteristics	17.4%	15.4%	16.9%
Firm's characteristics	11.5%	11.0%	11.3%
Context (time fixed effects)	9.7%	9.6%	9.7%
Score		4.4%	1.2%
R-squared	0.209	0.210	0.209

The table presents results from estimating equation (2). The outcome variable is revenue. The variables average score and adjusted score correspond to the average score from the panel judges and the adjusted score that removes the judge fixed effects. All columns include time fixed effects and industry fixed effects. Standard errors are clustered at the applicant level and bootstrapped for all columns including the adjusted score as a covariate (columns 3, 5 and 6). *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 6. Predicting Gazelles

Panel A—Probit Models

	(1)	(2)	(3)	(4)
Top quartile (by adjusted score)	0.204**	0.189**	0.220**	0.200**
	(0.0818)	(0.0814)	(0.0974)	(0.0838)
Controls for top quartile of average score		Yes	Yes	Yes
Controls for covariates at application			Yes	Yes
Control for acceleration				Yes
Pseudo R^2	0.0873	0.0881	0.184	
Observations	135	135	135	135

Panel B—Revenue Growth Rates

	Initial revenue (2014)	Final revenue (2017)	Growth from baseline	Implied annual growth
Gazelles	63.5	302.3	376%	68.24%
Non-gazelles	20.7	16.8	-19%	-6.70%

Panel A in the table presents results from probit regressions; reported coefficients correspond to marginal effects. The main explanatory variable is an indicator for applicants in the top quartile of adjusted scores. Regression controls vary as specified in each column. The covariates at application include indicator variables for established firms, gender, serial entrepreneurs, and founding team. They also include fixed effects for sectorial experience and entrepreneurs' education. The dependent variable is an indicator for gazelles: the top 10% applicants according to 2017 revenue (and splitting sample into business ideas and established firms). Panel B summarizes revenue growth rates across gazelles and non-gazelles.

Table 7. Unconditional Probability of Acceleration and Scoring Generosity

	(1)	(2)	(3)	(4)	(5)
Quartile of panel judge generosity	Overall (No controls)	Project in 25th percentile (adjusted score= 0.59)	Median project (adjusted score= 0.66)	Project in 75th percentile (adjusted score= 0.73)	Project in 90th percentile (adjusted score= 0.77)
1 (Unlucky)	17.64%	0.00%	0.01%	3.32%	26.89%
2	23.53%	0.00%	0.54%	30.82%	76.38%
3	26.74%	0.03%	16.58%	85.91%	98.91%
4 (Lucky)	36.36%	5.31%	78.96%	99.78%	99.99%

This table shows the probability of acceleration across a double sort of applicants by adjusted score (columns) and quartile of scoring generosity (rows). Column 1 reports results from a probit regression of acceleration, a dummy that indicates applicants that participated in the accelerator, against dummy variables indicating the quartiles of scoring generosity. Columns 2 to 5 report results from the same probit but control for adjusted score.

Table 8. Probability of Acceleration and Scoring Generosity

	(1) All	(2) Business Ideas	(3) Established Firms
2nd Quartile × After	0.203***	0.277***	0.101
	(0.0385)	(0.0472)	(0.0613)
3rd Quartile × After	0.272***	0.162***	0.352***
	(0.0389)	(0.0512)	(0.0580)
4th Quartile × After	0.490***	0.478***	0.482***
	(0.0406)	(0.0519)	(0.0626)
Adjusted Score * After	3.726***	3.161***	3.683***
	(0.201)	(0.259)	(0.318)
Constant	0.000	0.000	0.000
	(0.0144)	(0.0172)	(0.0206)
Observations	675	360	315
R-squared	0.652	0.678	0.739
Number of ids	135	72	63
F	57.58	35.80	39.09
Prob. > F	0.000	0.000	0.000

The table presents results from estimating equation (4). The outcome variable is Acceleration × after. The variables acceleration and after correspond to dummy variables indicating accelerated applicants and years after application to the accelerator, respectively. The bottom quartile of judge scoring generosity is omitted from the regression. All columns include applicant fixed effects and several controls, including adjusted score, firm's age, entrepreneur's age, entrepreneur's education, location, and sectorial and entrepreneurial experience. All columns include time fixed effects, and interactions between the controls and the variable after. Standard errors are bootstrapped and clustered at the applicant level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

.

Table 9. Acceleration and Project Growth

	All			Business ideas			Established firms		
	(1) OLS	(2) OLS	(3) IV	(4) OLS	(5) OLS	(6) IV	(7) OLS	(8) OLS	(9) IV
Acceleration × After	40.94*	42.91**	66.31**	8.962	-3.715	-40.53	62.42**	99.80**	116.8**
	(24.23)	(20.80)	(32.31)	(16.92)	(23.08)	(38.09)	(30.32)	(40.47)	(56.96)
Constant	10.54***	10.54***	10.54***	1.266*	1.266**	1.266*	21.14***	21.14***	21.14***
	(3.166)	(3.512)	(3.284)	(0.660)	(0.637)	(0.661)	(6.519)	(7.716)	(6.399)
Observations	675	675	675	360	360	360	315	315	315
R-squared	0.074	0.095		0.087	0.100		0.092	0.156	
Number of ids	135	135	135	72	72	72	63	63	63
Controls × After	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

The table presents results from estimating equation (3). The outcome variable is revenue. The variables acceleration and after correspond to dummy variables indicating applicants that were accelerated and years after application to the accelerator, respectively. All columns include applicant fixed effects and several controls, including adjusted score, firm's age, entrepreneur's age, entrepreneur's education, location, and sectorial and entrepreneurial experience. All columns include time effects. Some specifications also include interactions between the controls and the variable after, as specified in each column under the row "Controls × After". Standard errors are bootstrapped and clustered at the applicant level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 10. Delay in Firm Creation

	(1)
Acceleration × 2015	-0.259*
	(0.133)
Acceleration × 2016	0.297*
	(0.160)
Acceleration × 2017	0.069
	(0.211)
Constant	0.773
	(0.450)
Observations	107
R-squared	0.140

The table presents results from regressing a dummy indicating firm registration at the Chamber of Commerce against interactions of the Acceleration indicator variable and year fixed effects. The estimation includes several controls, including adjusted score, firm's age, entrepreneur's age, entrepreneur's education, location, and sectorial and entrepreneurial experience. Standard errors are bootstrapped and clustered at the applicant level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 11—Comparison Impact Estimates Based on Different Methodologies

	(1)	(2)	(3)	(4)	(5)
	OLS	IV	PSM without Adjusted Score	PSM with Adjusted Score	RD
Treatment * After	42.91**	66.31**	40.73**	59.10**	81.38**
	(20.71)	(30.62)	(20.31)	(28.12)	(28.02)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	675	675	399	354	405
Number of ids	135	135	133	118	135

This table presents the results from different methodologies to estimate the accelerator impacts. For ease of exposition, Columns 1 and 2 replicate the OLS and IV estimates from Table 9. Column 3 estimates the effects using propensity score matching based on observables including initial revenue, firm's age, entrepreneur's age entrepreneur's sex, education, location, and sectorial and entrepreneurial experience. Column 4 estimates the effects using propensity score matching based on the same observables as Column 3, but also including the adjusted score in the matching procedure. Column 5 estimates the effects based on a discontinuity regression approach that exploits the 35th rank threshold on average score. The dependent variable is the change of revenue (each after year vs 2014) variable. Standard errors are bootstrapped and clustered at the applicant level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 12. Randomization Inference

Panel A--Mean Differences Pre-Acceleration in Subsample

	Business Ideas			Established Firms		
	Not accelerated	Accelerated	*P*-value Diff in means	Not accelerated	Accelerated	*P*-value Diff in means
Observations	19	9		20	14	
Adjusted score	0.69	0.71	0.25	0.70	0.72	0.35
Gender: Male	68%	89%	0.26	85%	93%	0.50
Location: Cali	95%	100%	0.50	90%	79%	0.37
Has entrepreneurial team	84%	100%	0.22	100%	93%	0.24
Sectorial experience (years)	6.7	2.9	0.05**	5.1	10.9	0.00***
Serial entrepreneur	74%	89%	0.38	70%	79%	0.59
Advanced education (college or grad)	68%	67%	0.93	75%	93%	0.19
Firm's age				2.8	3.1	0.42
Total employees 2014	2.4	2.0	0.64	7.6	4.5	0.09*
Revenue 2014 (million pesos)	3.77	6.00	0.68	64.79	64.26	0.99

Panel B—Acceleration and revenue

	(1) Business ideas	(2) Established firms
Treatment * After	-24.86*	64.84**
p = c/n	0.053	0.035
SE (p)	0.003	0.0026
Controls and time fixed effects	Yes	Yes
Number of permutations	5000	5000
Note: c = #{|T| >= |T(obs)|}		

The table reports results for the randomization inference exercise. The sample is restricted to 62 applicants that belong to one of two subsamples: (i) applicants that were not accelerated but whose adjusted score is higher than the lowest adjusted score of the accelerated projects and (ii) accelerated applicants whose average score is lower than the highest average score among nonaccelerated applicants. Panel A summarizes the differences between accelerated and nonaccelerated businesses in the restricted sample for business ideas and established firms, separately. Panel B summarizes the randomization inference results of regressing post-application revenue against a dummy indicating whether the applicant was accelerated, time fixed effects, and controls.

. **ONLINE APPENDIX**

Appendix 1—Evaluation Guidelines for ValleE Judges

Por favor complete el siguiente formato de evaluación para cada uno de los proyectos a evaluar. Por favor use el ID que le fue asignado al igual que el ID de la empresa que va a evaluar y que encontrará en la ficha técnica de cada proyecto.

1) Por favor ingrese su ID de evaluador*

2) Por favor ingrese el ID del proyecto que está evaluando*

3) Nombre del proyecto que está evaluando*

4) Por favor califique de 1 a 5, qué tan bien definidos están los componentes del modelo de negocio. Por favor para esta pregunta califique ÚNICAMENTE la claridad en los aspectos y no el valor o potencial del modelo de negocio

5) Por favor califique de 1 a 5 el nivel de innovación del modelo de negocio, siendo 1 un modelo de negocio con muy bajo nivel de innovación (varios productos similares) y 5 un modelo de negocio con muy alto nivel de innovación (producto, servicio y/o mercado nuevo)*

6) Por favor califique de 1 a 5 el nivel de escalabilidad del proyecto, siendo 1 un modelo de negocio con muy poco escalable y 5 un modelo de negocio con alta probabilidad de escalabilidad.

Un proyecto es escalable cuando puede acceder a un siguiente nivel en términos de: volumen de ventas, cobertura geográfica, cobertura de otro nicho de clientes, crecimiento sostenible de empleados, entre otros

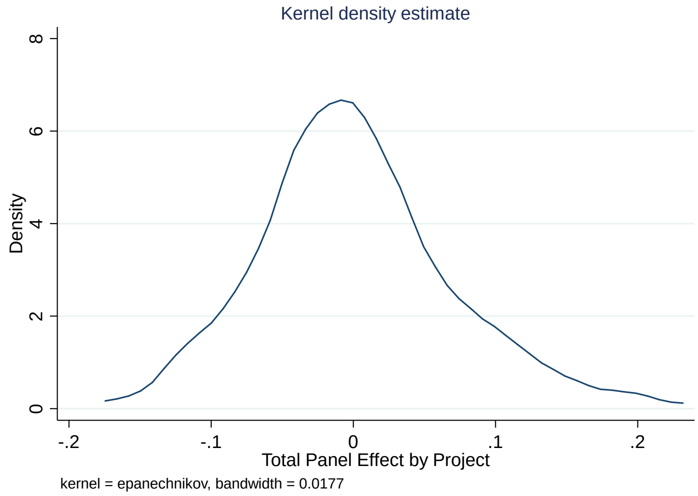
7) Por favor califique de 1 a 5 el potencial de rentabilidad del modelo de negocio, siendo 1 un modelo de negocio con muy baja probabilidad de generar rentabilidades mayores al promedio de su sector y 5 un modelo de negocio con altas probabilidades de generar rentabilidades mayores al promedio de su sector

8) Por favor califique de 1 a 5 las capacidades del equipo emprendedor para desarrollar el proyecto, siendo 1 un equipo con bajas capacidades y 5 un equipo experimentado y con las capacidades para desarrollarlo

Appendix 2—Judge Panels and Scoring Generosity

Relying on a panel of judges rather than on individual judges to rank projects helps mitigate the effect of judge heterogeneity but does not fully correct it, because judge panels are small (only three judges). We provide supporting evidence by comparing the distribution of *individual* scoring generosity (Figure 3) with the distribution of *overall* scoring generosity (Figure A2, below), defined as the sum of the corresponding judge fixed effects for each project. A comparison between the figures reveals that the distribution of the overall scoring generosity is more concentrated around 0 than the distribution of individual judge fixed effects, yet economically significant heterogeneity across judge panels remains, as 28% of judge panels tend to award individual scores that are one standard deviation above or below the average score of the other judge panels. The most generous (strict) judge panel adds (subtracts) an average of 0.12 (0.21) to any given project the group scores (relative to a mean normalized score of 0.66).

Figure A2. Overall Scoring Generosity



This figure plots the distribution of the estimated overall scoring generosity by project, which corresponds to the sum of the judge fixed effects from the estimates of equation (2).

Appendix 3—Sorting Ability of Judges

We characterize in more detail the evidence on the sorting ability of judges. We cut the data and regress residual revenue (post-application: 2015–2017) against adjusted scores. Residual revenue correspond to the residuals from regressions of post-application revenue (i.e., 2015–2017 period) against all the growth determinants from applications (i.e., all variables included in column 1 of Panel A in Table 4). Table A.3.1 summarizes the results. There are three findings regarding the apparent sorting ability of judges.

The first finding shows that judges seem better at evaluating the prospects of business ideas rather than the prospects of already established firms. A comparison between Columns 2 and 3 in Table A.3.1 shows that the adjusted scores predict the revenue of projects that apply as business ideas, whereas the estimated coefficient for established firms is not statistically significant. An alternative interpretation is that we do not have sufficient statistical power to reject the no-prediction null. However, the difference in observations between the two sample cuts is small (216 against 189), whereas the increase in the variance of the estimator is substantial.

The second finding is that judges' sorting ability appears higher at the top, rather than at the bottom, of the distribution of projects. Column 4 shows that when we leave out the bottom quartile of the distribution by adjusted scores, going from the lowest ranked to the top ranked project increases future annual revenue by 197 million COP. Instead, Column 5 shows no predictive power for the adjusted score once we drop the top quartile of participants (as indicated by the adjusted scores). This asymmetry in the predictive power of judges is consistent with prior work evaluating judges' predictions in business plan competitions (e.g., Fafchamps and Woodruf (2016)). However, Fafchamps and Woodruf (2016) show that in their sample the asymmetry runs in the opposite direction: judges in business plan competitions are better at cleaving off the bottom of the distribution, but not as effective at distinguishing within the top. One potential explanation behind the difference in results is the preliminary filter in ValleE: recall that 120 (out of 255 total applications) were not assigned to judges for evaluation, as their potential for growth was deemed to be too low. Had judges also evaluated the filtered projects, it is possible that adjusted scores would have had a high correlation with future performance for the bottom of the distribution (which presumably would include the projects that were filtered out).

Table A.3.1 Applicant Scores and Project Growth: Sample Cuts

	Residual revenue 2015 to 2017					
	All	Business ideas	Established firms	Excl. bottom Q.	Excl. top Q.	Excl. treated and bottom Q.
	(1)	(2)	(3)	(4)	(5)	(6)
Adjusted score	59.30*	57.74*	78.43	197.4***	-65.93	50.78**
	(35.38)	(33.37)	(59.24)	(55.95)	(78.85)	(23.63)
Observations	405	216	189	303	306	198
R-squared	0.003	0.008	0.002	0.013	0.002	0.030

Panel A reports the Shapley-Owen participation of each set of explanatory variables in equation (1) on the *R*-squared across the different regression models in Table 3. Note that some variables that are not significant in the estimation may still have a high Shapley-Owen participation value because of high correlation among explanatory variables in the model. Panel B reports the Shapley-Owen participation of each set of explanatory variables in equation (1) on the *R*-squared of the model in column (5) of Table 3 across different subsamples as detailed on the title of each column.

Appendix 4—Classification of Gazelles

Type of applicant	Age at application	Number	Initial revenue (2014)	Final revenue (2017)	Growth from baseline	Implied annual growth
Gazelles	Business idea, <1 year incorporation	9	14.0	205.0	1364%	144.6%
Nongazelles	Business idea, <1 year incorporation	68	8.2	3.0	-63%	-28.5%
Gazelles	>1 year incorporation	7	127.0	427.0	236%	49.8%
Nongazelles	>1 year incorporation	51	28.2	44.3	57%	16.2%

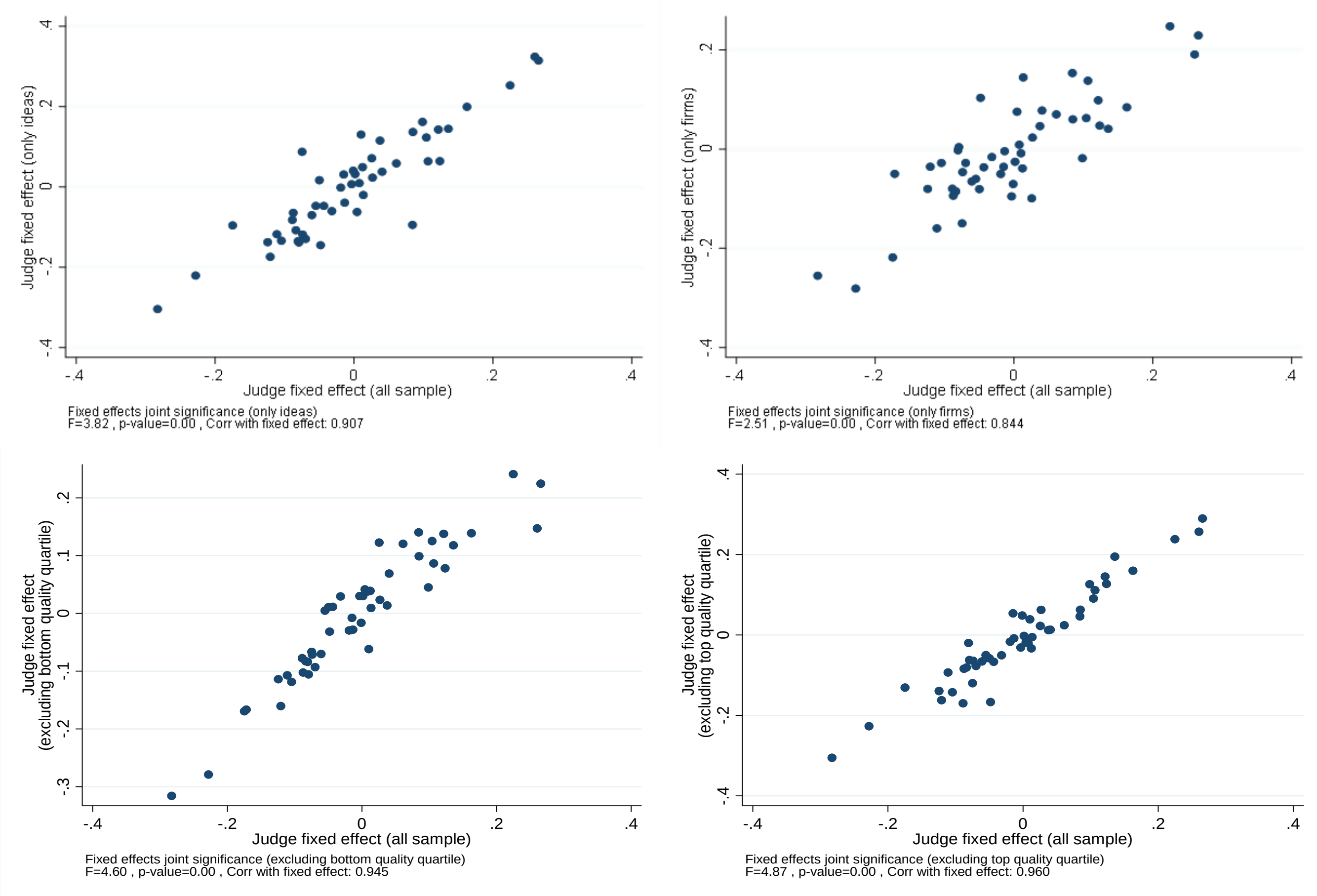
The table shows the revenue growth rates of gazelles and nongazelles across two types of applicants: (i) established businesses that have been registered with the Chamber of Commerce for more than one year at application and (ii) business ideas that were not incorporated at application and established businesses that had been registered with the Chamber of Commerce for less than one year at application.

Appendix 5—Balanced Sample Across Scoring Generosity Quartiles

Variable	Q 1	Other Q	*p*-value diff. in means	Q 2	Other Q	*p*-value diff. in means	Q 3	Other Q	*p*-value diff. in means	Q 4	Other Q	*p*-value diff. in means
Sex: Male	76%	80%	0.65	76%	80%	0.65	76%	80%	0.65	88%	76%	0.16
Education: High school	3%	15%	0.06	24%	8%	0.01	15%	11%	0.56	6%	14%	0.24
Education: Technical degree	24%	21%	0.74	18%	23%	0.53	18%	23%	0.53	27%	20%	0.36
Education: College	53%	51%	0.88	44%	54%	0.30	56%	50%	0.59	55%	51%	0.72
Education: Masters or PhD	21%	13%	0.28	15%	15%	0.98	12%	16%	0.57	12%	16%	0.62
Location: Cali	91%	83%	0.26	85%	85%	0.98	85%	85%	0.98	79%	87%	0.24
Motivation: To have stable income	21%	9%	0.07	12%	12%	0.99	3%	15%	0.06	12%	12%	0.96
Motivation: Own boss	3%	0%	0.08	0%	1%	0.56	0%	1%	0.56	0%	1%	0.57
Motivation: Business opportunity	76%	91%	0.03	88%	87%	0.87	97%	84%	0.05	88%	87%	0.93
Dedication: Sporadic	12%	4%	0.10	9%	5%	0.41	3%	7%	0.40	0%	8%	0.10
Dedication: Half-time	24%	21%	0.74	21%	22%	0.88	21%	22%	0.88	21%	22%	0.97
Dedication: Full-time	65%	75%	0.24	71%	73%	0.76	76%	71%	0.56	79%	71%	0.36
Sectoral experience (years)	6.1	5.4	0.51	4.6	5.9	0.23	5.3	5.7	0.72	6.4	5.3	0.37
Serial entrepreneur	62%	61%	0.97	53%	64%	0.24	62%	61%	0.97	70%	59%	0.27
Has entrepreneurial team	91%	87%	0.53	85%	89%	0.56	91%	87%	0.53	85%	89%	0.50
# people on team	3.2	2.9	0.43	2.9	3.0	0.73	2.9	3.0	0.73	3.0	3.0	0.91
Sector: Agriculture	18%	15%	0.70	9%	18%	0.21	15%	16%	0.88	21%	14%	0.31
Sector: Manufacturing	18%	22%	0.61	21%	21%	0.98	26%	19%	0.34	18%	22%	0.68
Sector: Water and electricity	0%	4%	0.24	9%	1%	0.02	0%	4%	0.24	3%	3%	0.98
Sector: Construction	3%	3%	0.99	0%	4%	0.24	3%	3%	0.99	6%	2%	0.23
Sector: Commerce	0%	3%	0.31	3%	2%	0.74	3%	2%	0.74	3%	2%	0.72
Sector: Services	62%	53%	0.40	59%	54%	0.66	53%	56%	0.73	48%	58%	0.35
Participated in other contests	62%	58%	0.73	62%	58%	0.73	59%	59%	0.95	55%	61%	0.53
% Established firms	47%	47%	0.96	44%	48%	0.73	53%	45%	0.40	42%	48%	0.58
Year founded (established firms)	2013	2013	0.58	2013	2013	0.79	2013	2013	0.38	2014	2013	0.07
Revenue 2013 (million pesos)	17.68	8.22	0.20	7.68	11.62	0.60	14.48	9.36	0.49	2.52	13.27	0.15
Revenue 2014 (million pesos)	31.82	23.77	0.47	21.40	27.28	0.60	34.93	22.72	0.27	14.72	29.38	0.19
Total employees 2014	5.0	3.7	0.10	3.7	4.1	0.55	4.2	4.0	0.80	3.2	4.3	0.19

The table compares applicants' characteristics (at application) across the different quartiles of panel judge scoring generosity.

Appendix 6—Robustness Checks Monotonicity



The figure plots scoring generosity measures that are calculated separately for different restricted samples (as specified in the y-axis of each subplot) against the corresponding judge fixed effects estimated for the full sample.

Appendix 7—Employment, Profits, and Fundraising

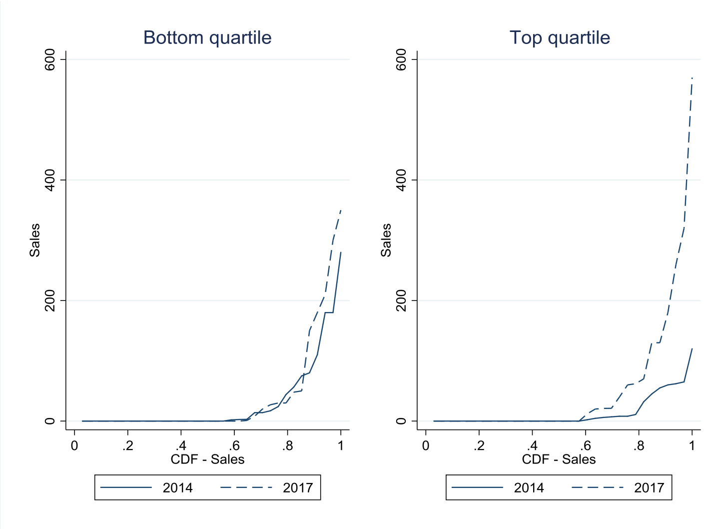
Table A7 summarizes the impact of the accelerator using profits and employment as outcome variables. We construct these additional performance measures using survey responses. The main advantage of this exercise is to provide a robustness check for our findings on acceleration impacts using alternative metrics. The main drawback is that the Colombian Registry has no information on these variables, and thus results based on these performance metrics can be subject to survey reporting bias (which is why we do not use these variables in our main set of results). With this caveat in mind, Table A7 shows evidence of large acceleration effects for employment. Acceleration increases number of workers by 4.4, relative to a baseline average of 4.0 employees per firm (Column 4). In terms of profits, the evidence is more nuanced; the point estimate is positive, but not statistically significant. When we restrict the sample to established firms, however, the evidence on profits is compelling: profits increase by 26M COP ($6.5K USD), 1.4 times relative to the baseline (Column 3). Our results on employment are similar to those in the literature; Glaub et al. (2012) estimate that treated firms have roughly twice as many workers as control firms after five to seven months of a three-day training intervention.

Table A7. Acceleration, Profitability, Employment and Fundraising

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Profits			Total Employment		Fundraising
Sample	All	Business ideas	Established firms	All	Business ideas	Established firms	All
Acceleration × After	18.50	-2.261	26.05**	4.416**	0.138	6.901**	0.227
	(11.99)	(11.55)	(10.59)	(2.014)	(2.821)	(3.315)	(0.216)
Year 2015	43.11**	-8.556	111.7*	10.96**	4.404	16.87*	
	(20.66)	(16.96)	(63.66)	(4.860)	(4.854)	(10.15)	
Year 2016	43.60**	-5.025	108.7*	10.20**	4.516	15.12	
	(20.35)	(16.95)	(62.63)	(4.772)	(4.944)	(10.33)	
Year 2017	40.24**	-4.955	101.5	9.528**	3.891	14.39	
	(20.37)	(15.96)	(62.64)	(4.714)	(4.744)	(10.41)	
Constant	6.116***	0	13.11***	4.030***	2.667***	5.587***	0.472
	(1.361)	(0)	(2.587)	(0.348)	(0.361)	(0.601)	(0.435)
Observations	540	288	252	540	288	252	117
Number of applicants	135	72	63	135	72	63	117

The table presents results from estimating equation (3), instrumenting Acceleration × After with quartiles of panel judge leniency. The outcome variable is indicated at the top of the columns. The variables *Acceleration* and *After* correspond to dummy variables indicating applicants that were accelerated and years after application to the accelerator, respectively. All columns include applicant fixed effects and several controls, including adjusted score, firm's age, entrepreneur's age entrepreneur's education, location, and sectorial and entrepreneurial experience. The specifications also include interactions between the controls and the variable *After*. Standard errors are clustered at the applicant level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Variable of receiving finance is a dummy variable that takes the value of 1 if in any of the three years after treatment the firm received capital from banks or private investors. We construct the measure only for those projects that we have at least one follow-up response.

Appendix 8. Cumulative Distribution of Revenue



This figure plots the cumulative distribution of revenues for 2014 and 2017 for projects in the bottom quartile of scoring generosity (left panel) and in the top quartile of scoring generosity (right panel).

Appendix 9—Propensity Score Matching

Panel A—Probit

	(1) Acceleration
Established firm	0.00681
	(0.0875)
Male	0.153**
	(0.0711)
Sector experience	0.00500
	(0.00651)
Serial entrepreneur	0.109
	(0.0808)
Adjusted score	-9.030
	(11.07)
Squared adjusted score	8.950
	(8.555)
Education: Technical degree	0.0836
	(0.177)
Education: College	0.128
	(0.125)
Education: Masters or PhD	-0.0498
	(0.132)
Has team	0.127
	(0.0839)
Motivation: Business opportunity	0.240
	(0.176)
Revenue 2014	-0.000358
	(0.000607)
Pseudo R^2	0.4709
Observations	135

Panel B—Common Support and Differences in Propensity Scores

	Off support	On support	Total
Untreated	1	99	100
Treated	16	19	35
Total	17	118	135
Average absolute difference in propensity scores		0.014	

Panel C—Kernel Matching Estimator

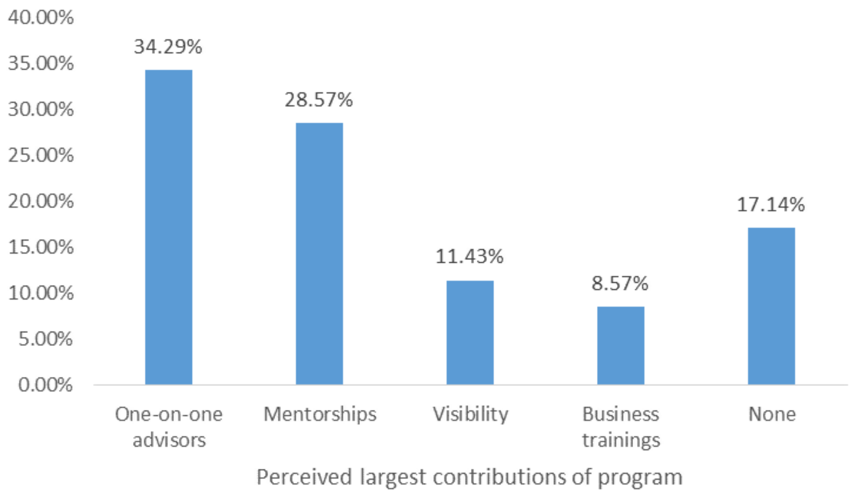
	All	Ideas	Firms
Average (Weighted) difference in revenue	59.10***	4.48	122.01***
	(25.79)	(33.51)	(50.97)
	354		

The matching procedure relies on a kernel matching of propensity scores. The matching begins with a probit regression at the applicant level of a binary variable indicating acceleration against different controls, including adjusted scores. Panel A presents the coefficient estimates and the adjusted R^2. They reveal that the regression captures a significant amount in selection, as indicated by the R^2 of 0.47. We then use the predicted probabilities from this estimation, the propensity scores for acceleration, to perform a kernel match with a radius of 0.05 that forces the matches to be in the common support. This requirement results in 16 accelerated companies for which we are unable to find a corresponding match (i.e., they have propensity scores that are too high relative to the rejected applicants). We report the number of applicants in the common support and the number of matched participants in Panel B. Panel B shows that the majority of differences in the estimated propensity scores for acceleration between the accelerated companies and their matches are inconsequential. The average absolute difference between the matched propensity scores is 1.4%. Panel C reports average annual differences in revenues post-application (2015–2017) between accelerated participants and their matches. Bootstrap standard errors are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

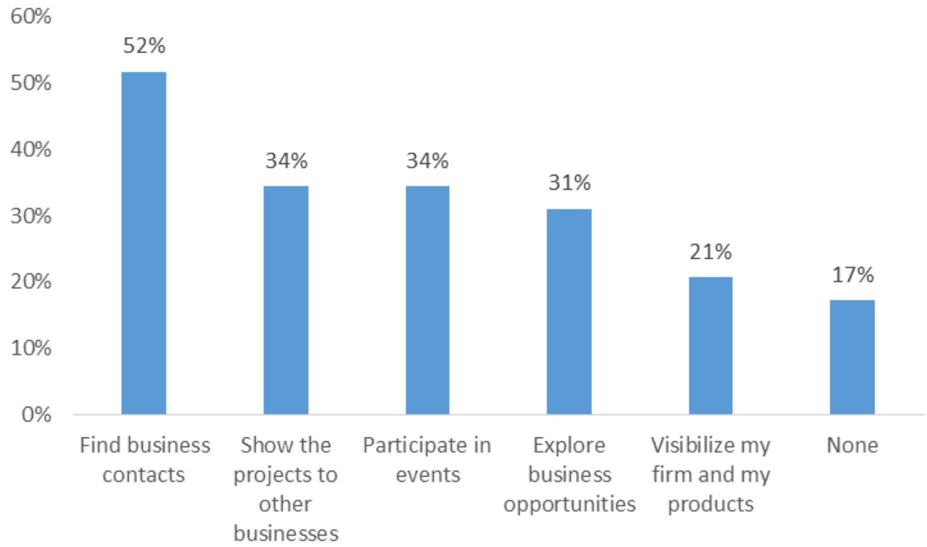
Appendix 10—Responses to Follow-up Surveys on Perceived Impacts

In the follow-up surveys we included questions to explore which services had the largest impacts as perceived by participants, as well as potential changes in firms' practices. Panel A plots the answer to the following survey question: "Please indicate which of these services had the largest impact on your company", where the options are as indicated in the plot's bar titles. Panel B plots the answer to the following question: "Did you achieve any of the following objectives via your participation in ValleE?", where the objectives are as indicated in the plot's bar titles. We use the response to this question using the last follow-up survey, which had a response rate of 60%. For those entrepreneurs who did not answer the survey, we use their last response from previous follow-up surveys (response rate of 67%).

Panel A—Perceived Benefits



Panel B—Perceived Changes in Practices



Appendix 11—Heterogeneity Effect by Applicants' Characteristics at Application

	(1) College education	(2) High school or technical education	(3) Considered needing strategic advice	(4) Did not consider needing strategic advice
Treatment Effect	55.79**	-17.83	54.33**	-12.2
	(23.30)	(13.8)	(22.98)	(21.4)
Observations	81	21	81	21
Number of ids	27	7	27	7

This table presents the t-test results of the estimated impact for different groups of the treated sample. The individual estimated impact comes from using a propensity score matching including the adjusted score as measure of a project's quality. The baseline survey asked entrepreneurs what type of advice they consider their project needed the most, and they could select up to three between the following categories: Management skills, strategic advice, marketing, prototyping, development of new products, legal advice, accounting skills, presenting skills, and financial advice.

Appendix 12—Back-of-the-Envelope-Calculation: Revenue Costs of Selection Mistakes

We use a back-of-the-envelope calculation of our findings to estimate ValleE's costs from failing to correct for heterogeneity in judges' scoring generosity. The results are presented in Table A12. We find that program revenues could have increased by 31%–40% had the accelerator used different selection mechanisms, where program revenues correspond to the sum of the revenues from all participants.

To produce these estimates, we start by calculating counterfactual revenues to the accelerator under the assumption that the program selected applicants based on adjusted scores rather than average scores. To make this calculation, we split the sample of participants into two groups: marginal (12) and nonmarginal, as explained in Section 2.4. We then estimate counterfactual revenues in 2015–2017 for the marginal projects using the implied revenues from Figure 8 (according to the heterogeneous treatment effects by propensity of acceleration). The counterfactual revenues of nonmarginal projects are the same as observed revenues. Finally, we compare observed and counterfactual revenues. The results are presented in Table A12. They show that counterfactual revenues are 31% higher than actual revenues.

We then consider a second alternative scenario, where the program selected participants based both on adjusted scores and covariates at application. We consider this second scenario because the results in Table 5 (Panel B) show that entrepreneurs' characteristics (even after controlling for adjusted scores) are highly correlated with firm growth. Therefore, an algorithm that takes into account both the hard (observables) and soft (adjusted scores) information, rather than just adjusted scores, is likely to generate further selection improvements. To calculate counterfactual revenues in this scenario, we use the propensity score calculated in Section 3.4 to create a ranking of the best 35 entrepreneurs according to the ex-ante observables and the adjusted scores. According to this new ranking, 26 out of the 35 accelerated entrepreneurs are not marginal in that they would also had been selected by this method, whereas 9 are classified as marginal. Performing the same exercise as above (i.e., replacing the revenues of marginal projects for those of the rejected applicants plus the implied treatment effect), we find potential revenue increases of 40%, as shown in Panel B of Table A12.[1]

Our results have implications for business accelerators, as well as other interventions that select participants based partly on scores from partially overlapping judge panels. They imply that sizable revenue improvements can exist if they control for heterogeneity across judges in scoring generosities. Further potential improvements can also be had if selection processes rely on hard information too, rather than only on judges' evaluations. It is possible that the costs from failing to correct for judge heterogeneity in scoring generosity extend beyond the accelerator and imply more general welfare inefficiencies. This will happen if a number of conditions are satisfied, including if (i) the schooling opportunities forgone by the accelerator are not arbitraged away by other accelerators (or other economic agents) and (ii) the provision of schooling through the program is more efficient (including potential spillover effects) than provision through alternative mechanisms (i.e., mentoring by angel investors). Given the high-quality resources at ValleE relative to other support organizations available in the region, efficiency costs are likely, but given data limitations, a judicious welfare quantification is beyond the scope here.

[1] In unreported analysis, we show that the additional revenue improvements in the second counterfactual stem from keeping four applicants that rank below the top 35 adjusted scores but rank high in propensity scores, and from eliminating one company that ranks low in propensity score but is among the top 35 adjusted scores.

Table A12. Correcting for Heterogeneity in Scoring Generosity and Program Revenues

Panel A—Adjusted Scores

Real status	Counterfactual status	Number	Real revenue (2015–2017)	Sum estimated individual impact	Counterfactual revenue
Accelerated	Accelerated	23	$7,773	$10,380	$7,773
Accelerated	Rejected	12	$2,121	$924	
Rejected	Accelerated	12	$2,510	$2,651	$5,161=$2,510+$2,610
Revenue accelerator			$9,894		$12,934
Revenue improvement controlling for judge fixed effects					$3,040=$12,934-$9,894
% Revenue improvement					31%=$3,040/$9,894

Panel B—Propensity for Acceleration

Real status	Counterfactual status	Number	Real revenue (2015–2017)	Sum estimated individual impact	Counterfactual revenue
Accelerated	Accelerated	26	$9,169	$11,419	$9,169
Accelerated	Rejected	9	$724	-$115	
Rejected	Accelerated	9	$1,363	$3,274	$4,637=$1,363+$3,274
Revenue accelerator			$9,894		$13,806
Revenue improvement controlling for judge fixed effects					$3,912=$13,806-$9,894
% Revenue improvement					40%=$3,912/$9,894

The table report our estimates of the revenue costs to the accelerator for not correcting for judge heterogeneity in scoring generosity. Panel A (B) includes the estimates for the counterfactual scenario where the accelerator selects participants based on the adjusted score (propensity score for acceleration, which is a function of the adjusted score and covariates at application).