

Asking Better Questions

The Effect of Changing Investment Organizations' Evaluation Practices on Gender Disparities in Funding Innovation

Amisha Miller

Saurabh A. Lall

Markus Goldstein

Joao Montalvao



WORLD BANK GROUP

Africa Region

Gender Innovation Lab

December 2023

Abstract

Female innovators raise fewer resources from investors, even when their ventures are similar to those of all-male teams. Efforts to mitigate the disparities have typically focused on changing how founders seek investment. However, the causes of gender disparities are systemic: in uncertain contexts, evaluators value women's competence or leadership potential lower than men's, and investors inquire more about risks when facing female founders than males. What is the effect of investment organizations' evaluation practices on gender disparities in funding innovation? This paper examines a two-stage global field experiment with investors making 1,871 investment decisions on early-stage

startups, which resulted in \$320,000 invested in 16 startups. The experiment changed an organization's evaluation framework to systematize investor inquiry across all ventures by including prompts about (1) risk and reward and (2) progress during the evaluation period. This caused treated investors to (1) assess startups more consistently and (2) assess startup competence more dynamically than control investors. It eliminated, even reversed, the gender gap in investment outcomes. These results have implications for organizations making decisions in uncertain contexts, and those aiming to reduce gender disparities.

This paper is a product of the Gender Innovation Lab, Africa Region. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at jmontalvao@worldbank.org.

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

Asking Better Questions: The Effect of Changing Investment Organizations' Evaluation Practices on Gender Disparities in Funding Innovation*

Amisha Miller Saurabh A. Lall Markus Goldstein Joao Montalvao[†]

JEL Classification: G24, D22

Keywords: Gender gap, Entrepreneurship, Innovation.

* We thank our implementing partner Village Capital. We thank the World Bank Group's Umbrella Facility for Gender Equality (UFGE) and the We-Fi initiative, as well as ANDE and SGB for funding. Thank you also to David McKenzie, Christopher Woodruff, the NESTA IGL Group, and the University of Oregon Policy Group for helpful comments. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. This paper is an output of the World Bank Africa Gender Innovation Lab, within the Office of the Africa Region Chief Economist.

[†] Miller: New York University's Stern School of Business, amisha.miller@nyu.edu; Lall: Adam Smith Business School, University of Glasgow, Saurabh.Lall@glasgow.ac.uk; Montalvao: World Bank, jmontalvao@worldbank.org.

1. Introduction

Women are underrepresented in leadership positions in innovation – among the population of funded ventures, under 12 percent of startups have female founders (e.g., Gompers and Wang 2017, Lerner and Nanda 2020, Luo and Zhang 2022). Early-stage startups with female founders are valued less than those with male founders even when ventures are similar or identical (e.g., Brooks et al. 2015, Guzman and Kacperczyk 2019, Ewens and Townsend 2020, Bapna and Ganco 2021). This disparity exists across the US and developing economies (Roberts and Lall 2018). It inhibits the ability of female-founded ventures to grow (Delecourt and Ng 2021), directs innovation away from novel solutions or female users (e.g., Jeppesen and Lakhani 2010, Koning et al. 2020), and, more broadly, can result in a large misallocation of resources within economies (e.g., Hsieh et al. 2019).

Scholars have theorized systemic reasons that produce and reinforce gender disparities in economic outcomes (e.g., Fernandez-Mateo and Kaplan 2018). Gatekeepers play an important role in resource allocation and are overwhelmingly male (Gompers and Wang 2017). They tend to socialize with, hire, or invest in people who share their gender (Ibarra 1993, Greenberg and Mollick 2017, Howell and Nanda 2019, Bapna and Ganco 2021). Beyond homophily factors, in contexts where quality is uncertain (such as innovation), all evaluators tend to rely on easily accessible indicators of expected quality, including status (Podolny 1993, Simcoe and Waguespack 2011, Kim and King 2014). Because men are typically perceived as higher status and/or more competent than women, this – often unconscious – reliance on gender can lead to lower evaluations for women than men (Ridgeway and Correll 2004, Correll and Benard 2006, Botelho and Abraham 2017, Snellman and Solal 2023). In addition, early evaluators often consider the preferences of others, which pushes evaluators to make more conventional choices, as encoded in status beliefs (Correll et al. 2017).

These gendered differences are also embedded in investors' behaviors during evaluation processes. In the absence of information about organizational performance (Aldrich and Fiol 1994, Cohen et al. 2019a), evaluation typically involves interacting with founders to gather information on the potential of the innovator and their idea within a short time (Kirsch et al. 2009, Petty and Gruber 2011, Huang and Pearce 2015, Huang 2018). During these interactions, evaluators assess innovators, and pattern-match their behaviors to previous successful cases (e.g., Elsbach and Kramer 2003, Huang 2018), who are typically male. Investors also often ask founders questions early in interactions to better understand a venture (Miller et al. 2023), and tend to ask female founders more difficult questions compared to male founders (Kanze et al. 2018). Given these patterns of behaviors, investors' processes of inquiry – how evaluators assess the potential of an innovator and their idea during interactions, in the absence of extant performance data – can disadvantage female innovators.

The extensive research on the sources of women's underrepresentation in innovation is not matched by how they might be mitigated (Jennings and Brush 2013). Scholars have examined female resource-seekers' pitches on investors' decisions (Kanze et al. 2018, Lee and Huang 2018, Balachandra et al. 2019, Huang et al. 2021). By focusing on how individuals might circumvent investors' behavior, scholars have held investor behaviors constant. Yet many investors are embedded in accelerators, angel groups, or venture capital firms, which typically allocate organization-level funds as part of a designed, collective investment process (Drover et al. 2017). Investment organizations and their funders (often limited partners) have invested \$4.8 billion in diversity strategies since 2018 (Cortes 2019, Biegel et al. 2020, DFC 2021), but the strategies they employ to do so, and the effects on investment outcomes are understudied. Without examining organizational evaluation practices – creating agreement on referents, negotiating about criteria, and establishing value by comparing entities (Lamont 2012) – scholars and practitioners cannot fully understand how disparities in

investment outcomes are created or might be reduced. *What is the effect of investment organizations' evaluation practices on gender disparities in funding innovation?*

Examining this question requires access to investment organizations' evaluation templates, investors' evaluation practices, and investment outcomes by gender. This research is difficult to conduct at scale because most investment organizations prefer not to share their evaluation processes and outcomes publicly (e.g., Da Rin et al. 2013). To examine this question, we employ unique data from Village Capital (Vilcap), a global investment organization that selects qualified early-stage startups for consideration by its own investors, and introducing startups to a broader and more diverse set of potential follow-on investors.

Typically, investment organizations evaluate startups using organization-level criteria over three months (Tyebjee and Bruno 1984, Fried and Hisrich 1994, Cohen et al. 2019b, Gompers et al. 2020). Investors receive information – i.e., a pitch deck or overview, meet founders, and then make a decision on whether to continue to diligence (or inquire about) the startup. This process repeats as investors progress through deeper stages of diligence before making a decision to invest in a startup. Working with Vilcap, we designed a two-part intervention to reduce gender disparities in their investment outcomes by changing their evaluation templates that structured inquiry, specifically how Vilcap prompted individual investors to inquire about risk in stage one, and startup progress in stage two. We examined the effects on investors' evaluation, analyzing over 31,000 scores investors allotted to startups during the period of study.

In the first stage of the field experiment, we used a cross-sectional design to assess the first stage of investment outcomes – an investor's decision to conduct further diligence on a startup, or to exclude it from further consideration. We randomized a diverse set of 278 investors into a treatment group, which Vilcap prompted to systematically inquire about risk and reward, and a control group

where investors evaluated startups as normal. After analyzing 87 startups – resulting in 1,341 judge-startup investment decisions on continuing diligence – treated investors sought consistent information across startups while control investors asked more risk-focused questions to startups with a female founder. This reduced gender disparities in evaluation, compared to the control group.

In the second stage, we leveraged a unique facet of Vilcap’s investment thesis – it trains local investors to invest \$320,000 into 16 startups over three months. We tested the effect of an additional treatment to systematically inquire about startups’ progress, added to the first treatment. We assessed this treatment’s effect on investment outcomes of Vilcap’s actual investment capital. Our analysis of a panel dataset of 1,530 decisions (from 510 investor-startup dyads over three time periods) suggests that treated investors assessed startups in a dynamic fashion rather than static – this changed how they evaluated startup competence and its future potential. These small changes in evaluation templates produced changes in investors’ processes of inquiry, which spurred differences in evaluation that eliminated, even reversed, the gender gap in investment outcomes.

Rather than focus on how to prepare startups for evaluation, our research explores a more systemic question: what is the effect of organizations’ evaluation practices on investments? We contribute to a growing literature in entrepreneurship that considers how to level the playing field for female entrepreneurs. This research highlights a novel mechanism to reduce gender disparities in investment outcomes: organizations can change evaluation templates to structure investor inquiry. This can affect which startups continue in due diligence, or receive investment. We extend prior research on designing organizational evaluation practices to reduce gender disparities in outcomes, to contexts where performance data is limited, and inquiry is a core part of evaluation. More broadly, we theorize how processes of inquiry can sustain or reduce status-based inequities in society.

2. The Role of Organizations in Evaluation

Organizations play an important role in designing evaluation practices that shape collective evaluation (e.g., Lamont 2012, Zuckerman 2012). They can design templates and processes that shape how decisions are made (e.g., March and Simon 1958). For example, organizations can specify how data that informs decision-making is shared or analyzed (e.g., Fayard and Metiu 2014, Anthony 2021), and how data is presented, can affect the decisions that are made (Kaplan 2011).

Organizations can also design evaluation practices to affect outcomes by gender. One common organizational process, similar to investment decisions, is hiring and promotion decisions – where evaluators make decisions about people under conditions of uncertainty, over a fairly short time frame, with ramifications for organizational funds and reputational outcomes over the long term. Organizations have made many efforts to reduce gender disparities in promotion and hiring, but many have been ineffective or had negative effects (e.g., Kalev et al. 2006, Dobbin et al. 2015, Stephens et al. 2020). Some interventions such as affirmative action policies, can lead to unintended consequences if they unintentionally increase the saliency of stereotypes that target groups lack competence, which can decrease target groups' performance and increase disparities between groups (Leibbrandt et al. 2018, Leslie 2019).

Focusing on evaluation processes and the content of evaluation has proven more successful in these contexts (Stephens et al. 2020), perhaps because these efforts tackle the organizational processes that could unknowingly reproduce inequality (Amis et al. 2020). Successful interventions include limiting employee discretion when making decisions (Castilla 2008), or shortening the evaluation scales employees can use to increase equity in evaluation (Rivera and Tilcsik 2019). Organizations could also focus on developing evaluation criteria that are not exclusionary, such as moving away from “cultural fit” (Rivera 2012), and instead rewarding performance on tasks (Stephens

et al. 2020). However, these recommendations do not perfectly apply to the context of evaluating innovation and venture potential.

Implementing strict decision-making rules in changing environments could theoretically limit an organization's ability to learn and adapt (March 1991, Canales 2014). This adaptation may be necessary, given that startup strategies are subject to change (e.g., Siggelkow 2002, Kirtley and O'Mahony 2020), and startups often operate in rapidly changing environments (Eisenhardt and Tabrizi 1995). Structuring clear rules to evaluate performance may also be difficult, as startups have little history of organizational performance (Stinchcombe 1965, Aldrich and Fiol 1994, Cohen et al. 2019a) and the potential value of an idea is difficult to ascertain before dedicating some resources to testing it (Gans et al. 2019). Using data on team performance could reward founders' elite connections, which would reinforce inequities in the status quo (e.g., Higgins and Gulati 2003, Hallen 2008). Given these difficulties in assessing static information, in investment, interacting with founders is seen as a fundamental part of evaluation (Petty and Gruber 2011). Investors pride themselves on seeking information beyond the business plan, and using their "gut feel" to source and evaluate investment opportunities rather than relying on data (Kirsch et al. 2009, Huang 2018).

3. The Role of Inquiry in Evaluation

Investors evaluate through inquiry – assessing the potential of an innovator and their idea during interactions. During these interactions with founders, investors ask questions to gather information on the venture (Kanze et al. 2018, Miller et al. 2023), but also to assess founders' potential to scale their venture (Huang 2018). Interactions are part of evaluation in many hiring contexts too. Managers typically use job interviews to hire new workers (Macan 2009), but these can introduce disparities in outcomes for both racial minorities and women (e.g., Rivera 2012). To reduce disparities, scholars have theorized that organizations could add structure to interaction by asking evaluators to

use structured interviews – asking the same open-ended questions to all applicants (Huffcutt 2011), or task-based interviews – asking applicants to complete a task or set of tasks that are similar to those required in the job (e.g., Ployhart et al. 2006). However, not all task-based interviews reduce disparities, and there is little causal evidence on the efficacy of structured interviewing (Stephens et al. 2020). This may be because the content of inquiry is important. For example, asking applicants to management consulting roles to evaluate a case on an industry heavily dominated by men was biased against women because they had less background knowledge about the industry (Rivera 2015). Given that inquiry is a crucial part of evaluation in innovation contexts, we examine how organizations might apply structure to inquiry processes, by focusing on two disparities in how investors inquire.

Inquiry about Risk and Reward. During interactions, investors tend to spend more time considering risk for startups with female founders than those led by male founders. Docsend, a platform that allows founders to share pitch decks with investors, has found that investors spend more time assessing traction and product slides for startups with female founders (to assess their current assets). In contrast, investors spend more time on fundraising request slides (to assess what founders might do in the future) for all male teams. In short, investors scrutinize startups with female founders differently than those with all male teams (Frost 2020). Similarly, in pitch competitions, investors typically ask prevention-focused questions – focused on maintaining non-losses and not changing to a worse state – to females to e.g., “How many monthly active users do you have?” They tend to ask promotion-focused questions to males to understand rewards or growth e.g., “How do you plan to acquire customers?” Investors’ patterns of inquiry produce conversations that differ by founder gender, and could cause investors to evaluate ventures with female founders as less valuable (Kanze et al. 2017, 2018). We hypothesize that prompting investors to systematically inquire about risk and reward could result in more consistent investor inquiry across startups, prompting investors to pay attention to both risk and reward for all startups, and reducing gender disparities in evaluation.

Inquiry about Progress. Investment organizations provide evaluation templates to investors to facilitate their assessment of competence by assessing the growth potential of early-stage startups in the absence of a history of startup performance (Cohen et al. 2019b, Gompers et al. 2020). However, evaluating potential can disadvantage female candidates. For example, in a retail organization, evaluating “potential” for leadership did not result in promotions for equally-performing female candidates. If the organization had promoted based on current job performance, it would have reduced disparities (Benson et al. 2022). Whereas performance ratings are backward-looking and based on demonstrable achievements, potential ratings are based on an evaluator’s forecast of a worker’s future performance and contribution. This makes rating “potential” fundamentally more subjective and uncertain, which could increase reliance on ascriptive characteristics such as gender (Ridgeway and Correll 2004, Correll and Benard 2006, Botelho and Abraham 2017, Snellman and Solal 2023).

How investment organizations might assess performance is complicated, given the lack of organizational performance data on startups (e.g., Stinchcombe 1965, Aldrich and Fiol 1994, Cohen et al. 2019a). However, some organizations assess short-term signals of performance during hiring processes. For example, the classic Goldin and Rouse study (2000) demonstrates that when orchestra hiring managers evaluated candidates’ performance through blind auditions, this resulted in hiring more female performers. This suggests that investors may be able to assess competence differently – inquiring about and assessing short-term signs of performance during the selection process.

In fact, some investors already do so. For example, program managers in Vilcap shared that progress made during the three-month program is important to investment decisions: “[We] invest in people that make the most progress during the program”. A VC investor, Mark Suster, (2010) blogged:

“The first time I meet you, you are a single data point... Because I have no observation points from the past, I have no sense for where you will be in the future. Thus, it is very hard to make a commitment to fund you.”

This suggests that some individual investors value signs of progress when making their investment decisions. However, most investment organizations design evaluation templates to assess static elements of a startup including team and venture characteristics (e.g., Tybjee and Bruno 1984, Gompers et al. 2020), which they use to assess future potential. This could disadvantage startups with female founders. We hypothesize that if organizations prompt investors to inquire about startups' progress as well as potential, this would focus investors' attention on demonstrable achievements – dynamic progress during the selection process. This could reduce gender disparities in investment decisions. Overall, we hypothesize that organizations can systematize both the consistency and content of inquiry to reduce gender disparities in investment outcomes.

H1: Investment organizations that systematize inquiry by prompting investors to inquire about risk, reward and progress will reduce gender disparities in investments.

How Organizations Affect Inquiry. Our main hypothesis suggests that organizations can create systems-level change in evaluation by systematizing the consistency and content of inquiry. By changing prompts in evaluation templates, organizations can reduce gender disparities in investors' decisions, which feed into collective investment outcomes. However, this hypothesis assumes a mechanism – that changing organizations' evaluation templates will cause individual investors to change their processes of inquiry, which will in turn, affect their evaluation of startups. This assumption may not hold, as investors pride themselves on using their intuition or gut feel to evaluate investment opportunities and do not follow templates (Kirsch et al. 2009, Huang 2018).

We test each part of the mechanism in the next two hypotheses. We theorize that if investors were to inquire about promotion and prevention for all startups, this could reduce gender disparities (Kanze et al. 2017), due to increasing the consistency of inquiry and investor attention to prevention concerns for all startups. To test the mechanism driving a possible change in gender disparities in

investments, we test whether organizations can affect the consistency of inquiry – how investors inquire about risk and pay attention to it – in hypothesis 2:

H2: Prompting investors to inquire about risk and reward will increase the consistency of inquiry across startups.

We test whether organizations can affect how investors inquire about progress as well as potential in hypothesis 3. There is evidence that if organizations change the content of inquiry, it can backfire. For example, when organizations positioned their hiring and promotion practices as meritocratic, hiring managers were even more likely to favor a male employee over an equally qualified female employee in pay increase decisions (Castilla and Benard 2010). To overcome this type of effect, organizations can create more transparency in evaluation processes and their effects (Castilla 2015). One way to do so is to set criteria in advance to reduce opportunities for retroactive criteria construction – for example by requiring evaluators to weight evaluation criteria before assessing applications (Uhlmann and Cohen 2005). We hypothesize that organizations can change the content of investor inquiry if they change evaluation templates to include new criteria and create transparency around evaluation practices in those templates. We test whether organizations can affect the content of how investors inquire – to prompt them to pay attention to progress – in hypothesis 3:

H3: Prompting investors to inquire about progress will increase investors' attention to startup improvement when evaluating competence across startups.

4. Research Approach and Setting

Understanding how investment organizations evaluate startups requires field research to examine organization and individual-level evaluation practices and link them to the outcomes produced. Following Yang and Aldrich (2014), we conceptualize organization-level evaluation frameworks as an input to decision-making. We designed and ran a two-stage field experiment to test whether systematizing inquiry would affect individual investors' behavior, and investment outcomes. By using a field experimental setting, we demonstrate how effective interventions are under “real-

world” conditions, overcoming concerns of generalizing from experiments in laboratory settings with students or online survey participants (Hsu et al. 2017, Czibor et al. 2019). This is particularly important in this setting, because investors that are trained to assess startups often evaluate startups differently to an average individual (e.g., Kirsch et al. 2009, Clingingsmith and Shane 2018).

These experiments were only possible due to access to a unique field site – Vilcap. Vilcap is a global investment organization with investor training programs in Africa, India, the Middle East, and Latin America. Vilcap is the:

“largest organization in the world supporting impact-driven, seed-stage startups. Since 2009 our team has directly worked with more than 1,100 entrepreneurs in 28 countries, and our affiliated fund, Vilcap Investments, has **invested in 110 startups** that have gone on to raise more than \$4 billion in follow-on capital.” – *Vilcap website*

Vilcap is an appropriate field site for this intervention as it provides access to two types of investor evaluation. Vilcap uses professional investor evaluations to facilitate introductions between startups and investors, and it trains local investors to invest Vilcap funds. This provided researchers with access to a setting using explicit evaluation templates to ensure effective communication between professional investors, Vilcap and trainees, and which facilitated discussions about evaluation with these stakeholders. Vilcap was also open to both field research and experimental methods with real investment funds to resolve the issue it faced: startups with all-male teams formed 70 percent of its portfolio, and it aimed to increase the number of startups with female founders in its portfolio.

In addition, Vilcap provided access to a similarly qualified set of startups. Vilcap uses a competitive process to identify startups with high growth potential to enter their program – with between 200 and 400 applicants for ten places. All startups have a product, are aiming to improve their product-market fit, and are currently seeking investment. Each investment program is focused on one industry or problem statement, so all startups within each program are working in a common

industry (but are not direct competitors). All startups were deemed by Vilcap to be of high quality and at similar stages. Baseline studies suggest that all startups accepted into the Vilcap program have similar observable characteristics, which does not differ by founder gender (Burns et al. 2019). Vilcap also provided access to a curated set of investors that had expressed an interest in startups at an early-stage and in a specific industry, which limits the sample to investors that have the potential to be interested in this stage of startup. Any differences in results should not be driven by differences in startup quality, nor investor interest.

Vilcap's evaluation process is typical of the average investment organization in several ways. The average investment organization employs a collective evaluation process to decide whether to invest their organization's funds into startups (Tyebjee and Bruno 1984, Fried and Hisrich 1994), which typically takes about 90 days (Gompers et al. 2020). Investment organizations assess startups using organization-level criteria which typically include assessments of the founding team, market size, product and business model (Cohen et al. 2019b, Gompers et al. 2020). As shown in a simplistic model in Figure 1, when evaluating, investors typically receive information such as a pitch deck or overview, meet founders, and then make a decision on whether to continue to diligence (or inquire about) the startup. This process repeats, as investors progress through deeper stages of diligence before they invest in a startup.

In Vilcap, this process unfolds with two types of investors. Professional investors,¹ embedded in a range of investment organizations, meet startups once. They receive a venture overview, written by Vilcap, meet approximately three founders for thirty minutes, and evaluate startups using a Vilcap survey where they decide if they wish to conduct due diligence, and ask for additional information

¹ Professional investors were invited by Vilcap and included other accelerator managers, investors from angel groups, and early-stage venture capital funders.

from the startup. If investors say they wish to continue due diligence on a startup, Vilcap facilitates an introduction. In the first stage of the field experiment, we tested the effect of prompting this diverse set of investors to systematically inquire about risk and reward, and assessed the effect on the additional information requested from startups, and the likelihood of continued diligence. This models the beginning of the selection process, and this type of cross-sectional design is common to research attempting to unpack demographic disparities in investment evaluation in the field (e.g., Younkin and Kuppuswamy 2018, Ewens and Townsend 2020).

Vilcap also trains local investors to evaluate startups for Vilcap and to allocate their own investment capital.² Since Vilcap is training investors from the region and market to evaluate startups on its behalf, it requires investors to provide scores on specific elements of the venture, including team, problem and vision, product, market and business model – typical criteria used by other investors. Trainee investors evaluate over multiple periods and are required to explain their reasons for scoring and to provide transparent feedback to startups. This provides a unique setting not only to observe how evaluation is conducted, but also to experiment with heterogeneous organization-level evaluation frameworks over time. In Vilcap, trainee investors typically receive a venture overview before meeting startups. To assess whether trainee and professional investors made similar decisions in the experimental program, all trainee investors were also asked to fill out the survey after initially

² Vilcap trains founders that qualify for its program to be investors and to allocate Vilcap funds. Their website explains the rationale for this decision: “What if, instead of relying on investors to “pick winners”, we chose to rely on entrepreneurs themselves? That hypothesis led to the creation of a collaborative due-diligence model ... to shift decision-making power away from investors... and instead, give that power to entrepreneurs to forecast which ventures are most promising.” Vilcap has run collaborative due diligence more than 70 times. We model entrepreneurs as “trainee investors” as they are trained to evaluate startups, conduct due diligence, and invest money on behalf of the organization. Vilcap’s investment decisions since 2009 are highly correlated with follow-on investment outcomes, suggesting that entrepreneur-investors make similar decisions to “real” investors. We assess the validity of this finding in the first experiment, where we leverage a pooled sample of trainee and professional investors.

meeting startups. As part of Vilcap's normal training program, trainee investors then continue to evaluate startups three more times, using standardized criteria.

In the second stage, we leveraged the panel dataset of trainee investors' investment decisions and tested the effect of an additional treatment, added to the first treatment. We prompted treated investors to inquire about startups' progress during the selection period. Using this panel dataset allowed assessment of whether gender disparities appeared at specific stages of the selection process (e.g., Botelho and Abraham 2017, Bohren et al. 2019).

To conduct the two-stage field experiment, we worked with Vilcap in eight of its investment training programs (two each in four regions – Africa, India, Middle East, and Latin America – allowing for one treatment and one control group in each region). Trainee and professional investors evaluated startups in these eight Vilcap programs. This resulted in a dataset of 31,714 evaluation scores. Our first stage leveraged the cross-section sample of 1,341 dyadic investor-startup decisions by investors made after the investor met a startup founder. This sample included both professional investors as well as trainee investors who were trained by Vilcap to allocate \$320,000 to 16 out of 87 startups. The second stage leveraged the panel nature of the trainee investor dataset. We randomized investors into treatment and control groups, with a panel dataset of 1,530 decisions (from 510 investor-startup dyads over three time periods after the initial analysis we observed in the first stage). Both stages of the field experiment were pre-registered.

5. First Stage: Systematizing Inquiry about Risk and Reward

5.1. Setting and Design. We systematized how Vilcap prompted investors to inquire about risk and reward in their evaluation template and assessed the impact on reducing gender disparities in continuing due diligence (H1), and the consistency of investor inquiry across startups (H2). Trainee investors met startups in a 90-minute welcome meeting where each startup founder was encouraged

to share a little about themselves and their startup. Professional investors met startups in 20-to-30-minute sessions where startups shared an overview of their business and then sought advice from the investor: either on their target market, product growth map, or fundraising strategy, depending on investor expertise. Vilcap shared a venture overview document with all investors with one page on each startup that outlined team members, market, product and the funds the startup was aiming to raise. We randomized professional investors into a treatment or control condition after they met a startup and began to evaluate them. We randomized trainee investors into a treatment or control condition after they were selected for the Vilcap program, stratifying by region, gender and subsector.³ In both cases, after meeting startups, treated investors received a slightly different evaluation form to the control group, which we outline below (see Appendix A).

5.2. Dependent Variable. After meeting startups, Vilcap’s evaluation form asked all investors to evaluate startups on a scale of 1 to 6: *“I would initiate due diligence on this venture.”* This variable is part of a dependent variable made up of four questions used in previous research by Clingingsmith and Shane (2018) – the part closest to a real investment decision.⁴

5.3. Intervention. After they met startups, Vilcap’s evaluation form asked control group investors: *“what additional information would you want on this venture?”* For the treatment group, Vilcap’s form prompted treatment investors to systematize inquiry about risk and reward: *“what additional information would you want on this venture’s potential for growth?”*; AND *“what additional information would you want on how this venture will mitigate risks?”*

5.4. Empirical Design. We ran the following pre-registered regression:

³ Given that all startups were operating in the same industry, at Vilcap’s request, we stratified by subsector to ensure no competing startups appeared in the same cohort.

⁴ We saw the most variance in this part of the variable in exploratory studies / pre-tests, which we ran on different investor populations before the experiment.

$$Y_{is} = \beta_1 F_s + \beta_2 T_i + \beta_3 F_s T_i + X_s + R + \varepsilon_{is}$$

The unit of analysis is the investor decision per startup. The dependent variable is Y_{is} – the propensity to invest in a startup s by investor i , measured using a six-point scale for professional investors. F is a binary variable which equals one when a female founder represented the startup and zero when solely male founders represented the startup. T is a binary variable which equals one when investors were prompted to inquire about risk and reward systematically (treated) and zero otherwise. The coefficient of interest is β_3 , on the interaction of inquiry about risk/reward and female founder. We included fixed effects for the region R . Although we randomized startups into treatment and control, given the relatively small number of startups we assessed (87), it is possible that startup characteristics could affect the size of the estimates. Therefore, we also controlled for observable startup characteristics X_s – the number of employees and the log of funds raised at selection into Vilcap’s program.⁵ We ran an ordinary least squares regression, which was pre-registered, but added an ordered logit because the dependent variable was ordinal. We report the ordered logit, as results were the same in both models. We clustered errors in all models by investor – the level at which the treatment was implemented.

Mechanism. To assess the mechanism, we assessed if prompting investors to systematize inquiry about risk and reward increased the consistency of questions investors asked across founders, Y is equal to one if a prevention question was asked to a startup, and zero otherwise. Two research assistants (one for Spanish, one for English) coded all investor questions posed to founders with prevention/risk or promotion/reward, following Kanze and coauthors (2018).⁶ Any disagreements were discussed in a group with the first author so that codes were applied consistently. The final

⁵ Vilcap also collected funds raised by startups, but we did not include this as a control, as many startups had not yet raised funds, and there were many zeros. We also opted not to use Vilcap’s own evaluation score as a control, as it was highly correlated with employees.

⁶ We constructed the prevention variable following Kanze and coauthors (2018) but included some additional variables to better suit the setting – a series of dyadic entrepreneur-investor interactions, rather than a one-off pitch where investors asked questions to entrepreneurs in a group setting.

decision on whether a question was coded by promotion or prevention was made by the first author, who did not have access to the startup's founder gender when making these decisions. We constructed a binary measure at the investor-dyad level to measure the incidence of a prevention focus, and used the same type of binary incidence measure for a promotion focus. Simply put, if an investor question to a startup did not have a promotion nor prevention focus, both variables would equal zero. If an investor question to a startup included both a promotion focus and a prevention focus, both variables would equal one.⁷

5.5. Results. We analyzed 1,341 decisions taken by 278 investors – combining 198 professional investors and 80 trainee investors – on 87 startups. As shown in Table 2, we assessed differences across all investor characteristics across treatment and control groups. As expected, given our randomized treatment assignment, there were no significant observable differences – using raw numbers, percentages or the p-value taken when regressing each characteristic on treatment.⁸ As a result, any differences in evaluation practices between treatment and control groups are likely to be caused by our randomized treatments – not by the types of investors in treatment and control groups.

Investors in the control group scored startups with female founders significantly lower than those with all male teams. Startups with female founders received an average score of 3.7 out of 6, while startups with all male teams received 4.1. When including startup controls, investors in the control group gave startups with female founders significantly lower scores than startups with all male

⁷ Following Kanze et al. (2018), we also used a computerized method on English responses, utilizing a dictionary of 27 promotion and 25 prevention words developed and validated by Gamache et al. (2015), and uploaded these dictionaries into Linguistic Inquiry and Word Count (LIWC) software to determine their frequencies. Similarly to Kanze et al., this was not our preferred method, as the dictionary approach leaves LIWC vulnerable to a low detection rate; the software is not sensitive enough to capture intentions that do not directly overlap with the very specific 52 words in the regulatory focus dictionary (Gamache et al. 2015). We use this LIWC method as a robustness check for all responses in English, to verify the direction of results from the research assistants' qualitative coding. We find the same directional results.

⁸ A small minority of professional investors met startups in multiple programs i.e., in the Middle East and Africa. As we randomized the investor according to the survey they received, 18 of the 276 investors encountered the treatment condition in one program, and the control condition in another program.

teams (0.6 on a 6 point scale – equivalent to 10 percentage points), as shown in Appendix B. This disparity held across trainee investors and professional investors, as well as male and female investors. This difference in scores was correlated with the fact that investors were more likely to focus on prevention when inquiring from startups with female founders than startups with all-male teams. As shown in Appendix B, investors in the control group were 15% more likely to ask a prevention-focused question to a startup with a female founder than a startup with an all-male team. This difference was directional, but not significant. Together, this suggests that when investors evaluate startups, they score startups with female founders lower than startups with all-male teams, and ask systematically different questions by founder gender. These differences in inquiry are directionally similar to those observed in US-based pitch contexts (e.g., Kanze et al. 2018).

We tested whether changing an evaluation framework to systematize inquiry around risk and reward would affect gender disparities in outcomes (H1) and whether it could change investor behavior (H2). Figure 2 shows the effect of systematizing inquiry by prompting investors to ask about risk and reward on an investor’s decision to continue diligence on the startup, by founder gender. While control investors scored female founders significantly lower than startups with all male teams, treated investors did not. This provides some evidence in support of H1, suggesting that if investment organizations systematize inquiry by prompting investors to ask about risk and reward, they can reduce gender disparities in evaluation outcomes.

Figure 2 also shows that investors in the control group were directionally more likely to ask prevention-focused questions to startups with female founders than those with all-male teams and that treated investors were more likely to inquire consistently across startups. This provides some suggestive evidence in support of H2; if investment organizations systematize inquiry by prompting investors to ask about risk and reward, investors will inquire more consistently across startups. This

difference between treated and control investors is driven by treated investors being significantly more likely to ask prevention-focused questions to *all* startups – directionally, even more so for those with all-male teams. This provides some evidence in support of H1 and H2, suggesting that if investment organizations systematize inquiry by prompting investors to ask startups about risk and reward, investors will inquire more consistently across all startups, and reduce gender disparities in evaluation outcomes.

As shown in Table 3, regression analysis including startup controls showed similar results. Control investors were less likely to take a startup with female founders through due diligence compared to a startup with an all-male team, while treated investors were equally likely to take a startup through due diligence, regardless of founder gender. Control investors were only 65 percent as likely to increase their score by one unit (i.e., from agree to strongly agree to take the startup through due diligence) for startups with female founders compared to all male teams. However, treated investors were equally likely to do so (0.65 main effect multiplied by the 1.63 interaction).⁹ This provides more evidence in support of H1. Regression analysis suggests that systematizing inquiry by prompting investors to ask about risk and reward increased the likelihood that investors would ask a prevention-focused question to all startups (by 265%), but that the likelihood increased less for startups with female founders.

These results provide evidence that simply prompting an investor to ask founders about “potential for growth” and “how this venture will mitigate risks” meaningfully affected the types of questions that investors posed to all startups, but particularly those with male founders. For example, the male founder of a platform startup that used mobile technology to connect handymen with work

⁹ Common to many experiments, we focused our design for this experiment on isolating the effects of gender on investment decisions (score), and the effect of our treatments. As an investment decision, we expect other unobservable preferences – including the weather – to add noise (e.g., Dushnitsky & Sarkar 2022).

opportunities received different questions from control and treated investors. A control investor asked: Can I see a “marketing plan clearly highlighting the marketing strategies?”. In contrast, a treated investor asked the same founder: How will the company “manage delayed payments (Risk) in case the company decides to partner with county or National Government”? Both questions were about scaling, but treated investors were more likely to use a prevention-focus frame – similar to the frames all investors used when assessing startups with a female founder. Treated investors were equally likely to agree or strongly agree to the statement “I will conduct due diligence” whether the startup had a female founder or all-male team. A simple change to an evaluation template affected investors’ use of prevention-framing, and whether they wanted to take the startup through due diligence. Together, this suggests that if organizations systematize inquiry by prompting investors to inquire about risk and reward in evaluation templates, this causes investors to inquire consistently across all founders, paying attention to risk and reward regardless of founder gender, which reduces gender disparities in investment evaluation practices.

5.6. Alternative Explanations. Differences in outcomes were robust to alternative measures of the dependent variable (score – using OLS or a weighted score by investor), the independent variable (analysis of prevention-focused questions by LIWC) and the female binary variable (presence in venture overview – see Appendix C). The results held for heterogeneous investor types (male and female, and trainee and professional). We do not have a large enough sample to conduct heterogeneity analysis by other investor characteristics, but we observed no directional difference on the relationship of treatments with the score.¹⁰ This suggests that results could hold for a diverse range of investors.

5.7. Investment Outcomes. This experiment revealed that organizations can change evaluation frameworks to prompt investors to inquire about risk and reward systematically, and that

¹⁰ Heterogeneity analyses are available on request.

this affected the questions investors asked, and reduced gender disparities in evaluation. This effect was also meaningful. Investors that agreed or strongly agreed to the statement “I will conduct due diligence” were likely to actually do so in this context. In this sample, 31 control investors selected “strongly agree” and 73 “agree” for startups with female founders. In the treatment group, 37 investors selected “strongly agree” and 82 selected “agree”. This suggests that startups with female founders would have entered into 15 more due diligence processes in the treatment group than the control, which could have meaningful implications for future investment. If these results hold more broadly, this type of intervention could reduce disparities in investment outcomes by founder gender.

If investment organizations prompt investors to think about prevention and promotion, investors inquire more consistently from startups with female founders and those with all male teams. This results not only in more rigorous due diligence on startups with all-male teams, but also an assessment that startups with female founders and all male team pose similar risks. This assessment produces consistent evaluation across founder gender, and changes the number of startups with female founders that enter due diligence processes.

One important limitation to this experiment is that we did not observe final investment decisions, so findings can only generalize to early stages of the investment selection process – the decision to begin a due diligence process on a startup. For this reason, we conducted Stage 2, to evaluate the effects of systematizing inquiry on real investment decisions.

6. Second Stage: The Effect of Systematizing Inquiry on Risk, Reward and Progress

6.1. Setting and Design. Vilcap trains local entrepreneurs to evaluate startups, conduct due diligence, and invest \$320,000 of Vilcap’s money into 16 early-stage startups over three months. The first author observed the entire process in the same Vilcap programs as Stage 1, to further assess whether cross-section results applied over time (e.g., Bohren et al. 2019). We changed Vilcap’s

evaluation framework by systematizing how Vilcap prompted investors to inquire from startups during interactions. As a bundled treatment, we systematized both how Vilcap prompted investors to inquire about risk and reward, AND about progress. We assessed the impact on reducing gender disparities in investment decisions (H1), the consistency of investor inquiry across startups (H2), and whether investors assessed startup improvement (H3).

6.2. Dependent Variable. After baseline, three more times over the course of its 90-day program, Vilcap asked investors to complete due diligence again, and rank startups. Vilcap’s normal set of evaluation questions are focused on assessing potential: *“what is the company’s growth opportunity? And what is the company’s investment opportunity?”* across eight categories (i.e., team, value proposition, market, scale).¹¹ Investors use a four-point scale per category, resulting in a 24-point scale overall – from 8 to 32. The final evaluation scores result in two investments into the most highly-scored startups.

6.3. Intervention. After Vilcap recruited entrepreneurs, we randomized startups/investors into treatment and control groups, stratifying by region and gender. Treatment investors received a bundled intervention using Vilcap’s changed evaluation template: 1) The same treatment that prompted investors to systematically inquire about risk/reward in Stage 1; and 2) Vilcap prompted investors to inquire about startup progress as well as potential during the program. For the treatment group, Vilcap added four questions (each on a 4-point scale, weighted to equal 1/3 of the overall evaluation score) to assess startups’ progress: *“Since the beginning of the program, how much has this company improved in...”*: i) *“understanding its path to growth?”*; ii) *“executing its path to growth?”*; iii) *“understanding its risks?”* iv) *“executing on risk mitigation?”*

¹¹ Vilcap is unwilling to share its proprietary evaluation templates openly, but will share it with academic reviewers to demonstrate the similarities between its due diligence questions, and the additional questions created as part of the intervention.

6.4. Empirical Model. We ran the following ANCOVA regression, to increase statistical power, following McKenzie (2012). By including the baseline score from Stage 1 as a control variable, we assessed the change in scores after the additional bundled treatment was applied:

$$Y_{ist} = \beta_1 Y_{is0} + \beta_2 F_s + \beta_3 T_i + \beta_4 F_s T_i + X_s + R + t + \varepsilon_{ist}$$

Our unit of analysis is the investor decision per round. The dependent variable is Y_{ist} – the propensity to invest in a startup s by an investor i at a time period t . t is the stage of measurement – a scale variable collected over three time periods. Y_{is0} is the baseline measure of evaluation of a startups by an investor, also evaluated on a scale. All scale evaluations are normalized by Vilcap, using a zscore.¹² F is a binary variable which equals one when a female founder represented the startup and zero when solely male founders represented the startup, which varies over time. T is a binary variable which equals one when investors were in the changed evaluation framework treatment group and zero when investors were not. The coefficient of interest is β_4 , on the interaction of treatment and female. We included fixed effects for the region R and time t and controlled for the same observable startup characteristics X_s as in Stage 1. We clustered errors in all models by investor – the level at which the treatment was implemented.

Mechanism. We added mechanism variables to assess if systematizing inquiry at the organization level changed how investors assessed startups. First, we examined whether investors inquired consistently across all founders using the same mechanism as Stage 1. Second, we examined whether investors assessed startups’ dynamic improvement by asking all investors to weight the criteria they used when evaluating: *“Please think about how you made your decisions and weight the criteria below with*

¹² In Vilcap, each startup receives a zscore per round. The inputs are the average score and the standard deviation per judge per round. Then for each judge's score for each startup Vilcap creates a $z_score = (score - avg_score)/sd_score$. Vilcap then joins the ranks together by taking an average across all rankers. Vilcap’s zscore weights scores according to a judge’s baseline score. This type of weighted score can help to avoid heterogeneity in judges’ baseline scores driving results (Gonzalez-Urbe & Reyes 2021).

percentages of how much weight you placed on each criterion. (Please make sure it adds up to 100%!) – [Growth opportunity, Investment opportunity, Improvement made during program]. Given that investors do not accurately explain the criteria that are important to them (Petty and Gruber 2011), we followed recent field experiment research and conducted semi-structured interviews to provide more insight into mechanisms driving results (e.g., Dimitriadis and Koning 2022). The first author interviewed all 15 trainee investors in one region to further examine how they decided to give a startup a high or low score – what investors were paying attention to during evaluation.

To assess the effect of systematizing inquiry on individual investors' evaluation, we compared how investors in the treatment and control group evaluated startups using a set of criteria, by founder gender, across three phases of evaluation. We used the same regression model, but changed the dependent variable to each criterion score i.e., score for “business model”. We then reran the regression to assess how investors re-evaluated startups from their first impressions at baseline.

6.5. Results: The balanced panel dataset after attrition was 1,530 investment decisions made by 65 investors on 69 startups over three time periods (510 decisions per period). Attrition was largely due to Covid-related absences. As shown in Table 4, there was no significant difference in observable characteristics between treatment and control groups – using raw numbers, percentages, or the p-value taken when regressing each characteristic on treatment.

To assess the baseline, we used trainee investors in the control group. Similarly to Stage 1, investors asked startups represented by females more prevention-focused questions and awarded lower scores than startups represented by all males. However, during the program, we observed this effect lessening, observing only directionally different effects by gender (see Appendix D). This suggests that the Vilcap program itself reduced gender disparities in the control group. The “Vilcap effect” could be driven either by the fact that all startups continue to be evaluated over time, or Vilcap

programming in which investors were provided with a standardized scoring system and questions when assessing each other. Overall, this suggests that VilCap is a conservative setting to test whether systematizing inquiry can reduce gender disparities in evaluation.

Investors in the control group evaluated startups represented by all male founders lower over the course of the program, from a baseline zscore of 0.12, which was reduced to 0.07. As demonstrated in Figure 3, investors very slightly increased their evaluation of startups represented by females on average from -0.09 to -0.07 on the zscore. This suggests that the Vilcap program itself acted to reduce gender disparities, largely by reducing zscores for startups with all male teams. Even so, startups with female founders received directionally lower zscores than startups with all male teams at the end of the program. By contrast, in the treatment group, the baseline score was already closer for startups represented by male founders (0.04) and female founders (0.01) because investors had been treated at baseline (prompted to ask about risk and reward). Even given this baseline difference, there were additional effects of the systematizing inquiry bundled treatment. Treated investors evaluated startups represented by all males lower over the course of the program (from 0.04 to -0.07), and evaluated startups represented by female founders higher over the course of the program (from 0.01 to 0.11). This effect was driven by changes in how startups with female founders were evaluated. Effects in the treatment group were larger than those observed in the control group, and were driven by increases in scores to startups with female founders.

As shown in Table 5, regression analysis showed similar results. Treated investors using Vilcap's systematizing inquiry evaluation template scored startups represented by females lower than startups represented by males at the beginning of the program but not significantly so. Investors treated systematizing inquiry scored startups with all male teams lower than control group investors,

but the effect on females was positive, increasing their scores by 0.31 in the pre-registered model. This effect was significant at $\alpha=0.05$, providing support for H1.

Organizational prompts to investors to inquire about “potential for growth”, “how this venture will mitigate risks”, and “startup progress” during the selection process, affected how investors evaluated and scored startups with female founders. Systematizing inquiry in these ways caused treated investors to evaluate startups with female founders as higher than those with all-male teams. The bundled treatment not only reduced gender disparities in scores, but reversed them. This suggests that organizations can change evaluation templates to reduce gender disparities in evaluation outcomes, even with real investments made over time. It also suggests that systematizing inquiry by prompting investors to inquire about progress may create a different effect than prompting investors to inquire about risk and reward.

6.6. Investment Outcomes. We next examined the effects of reversing gender disparities in evaluation scores, on gender disparities in investment outcomes. In the eight Vilcap programs in the sample, only 16 investments were made (to the two startups in each program with the highest scores). We cannot assess the effects of changing evaluation outcomes on this rare outcome. However, we conducted a simple calculation to assess if increasing a zscore by 0.31 – the difference in score for a startup with a female founder in the treatment group versus the control group – could affect the likelihood of investment for a startup. The average zscore for a startup that was ranked second and received investment was 0.76, compared to the average zscore of 0.43 for a startup that was ranked third and did not (Appendix E). The average difference was 0.33, close to the effect of the treatment for female startups. This suggests that the size of the increase in score could meaningfully change investment outcomes and reduce the gender disparities in Vilcap’s overall portfolio.

6.7. Mechanism Analysis: We conducted exploratory analysis to examine the mechanisms behind the effect of the bundled treatment that systematized inquiry on risk, reward and progress. We examined whether the differential effect in score by treatment and gender was driven by two mechanisms: the consistency of inquiry – by likelihood of asking a prevention-focused question to a startup (H2); or the investors’ assessment of dynamic improvement when evaluating startups (H3).

We reran the analysis from Stage 1 on the panel dataset and found that during the VilCap program (when VilCap provided due diligence questions for trainee investors to use), female founders were equally likely to receive prevention-focused questions as males, so the mechanism could not work as predicted (see Appendix F). Prompting investors to inquire about risk and reward had similar results to Stage 1, where males were asked more prevention-focused questions, and this effect was lower for females, but this did not affect the score. These results did not provide evidence in support for H2 in this setting, where investors and founders interacted repeatedly over a long period of time. Combining these results with results from Stage 1, we suggest two explanations. First, prompting systematic inquiry around risk and reward may be most valuable to retain startups presented by female founders at early stages in the evaluation pipeline, to keep them in consideration for the further selection process. Second, any prompt to systematically inquire (either from Vilcap during its program, or from the intervention we designed) may have a substitution effect. Both prompts increased scrutiny for startups led by all founders, but particularly for startups with all-male teams.

We next assessed whether investors’ assessment of dynamic improvement differed by treatment and control groups (H3). As shown in Appendix G, treated investors weighted the criterion of “improvement” as a higher part of their evaluation criteria (20.8%) than control investors (18.5%). This difference was close but not statistically significant at $p = 0.108$, on a sample of 65 investors, providing some directional evidence in support of H3. However, it is difficult to interpret the import

of this small percentage difference on how investors actually assessed startups. To provide more insight, the first author conducted 45-minute semi-structured interviews with investors in one region, to examine how investors conducted evaluation, and how they decided to give startups a high or low score. After interviewing all investors in the treatment and control program in one region, we found that many investors evaluated static startup elements of the venture to assess its potential. For example, a male investor explained:

“I gave AA a four [top score] in most of the categories... I really like their solution... it has a lot of potential for scaling... I went through their website... I was quite impressed with the profiles of people that work in the team... their business model too.”

A female investor explained that she rated a startup well when the problem was convincing: “I rated BB high... because I think the business idea is really necessary... I see its use and purpose.” These investors used similar criteria to that elaborated by Gompers and coauthors in their survey of early-stage investors (2020). In the control group, all nine investors evaluated static criteria (100%). As shown in Table 6, in the treatment group, three of the seven treated investors (43%) also described how they gathered data on static criteria in at least one of their responses.

Some investors took a more dynamic lens, focusing on progress or improvement when scoring ventures. A male investor described how he had ranked a startup well because he had observed improvement over time:

“The question is, have you seen improvement in them during this business program? ... I’ve been in a breakout room with YY twice. And the suggestion I had noted to her in the first breakout room ... was repeated with a group of mentors... they asked a similar question.... I noted that when she’s answering the question, she’s answering it differently... it sounded much better than the first time... once you see those things from people you can see that they are improving and changing.”

In the treatment group, six of the seven investors (83%) explained that they valued startups that made progress over time. Only two of nine control group investors (22%) explained that they assessed progress. These results provide evidence that when Vilcap asked investors to evaluate progress as well as potential, investors assessed startups dynamically. This shifted the focus of evaluation from startup attributes and a forward-looking assessment of potential, to a backward-looking assessment of what startups had actually accomplished over a short period of time. Taken together with the finding on changes in the score, this suggests that when investors assessed startups dynamically, it positively affected their evaluation of female founders' startups, which reduced gender disparities in investment outcomes. This provides evidence for H3.

To further examine how investors changed their scores of startups, we analyzed the effects of systematizing inquiry on assessments of startup characteristics. As shown in Table 7, systematizing inquiry affected how investors assessed startups that were represented by female founders. Over time, treated investors assessed startups with female founders significantly higher on “product” and “investor exit” than control investors. Treated investors also assessed startups with female founders directionally higher on “business model” and “scale”, compared to control investors. There was no significant difference in scores given to startups with male founders.

By systematizing inquiry (prompting investors to systematically inquire about risk, reward, and startup progress), investors changed how they evaluated startups with female founders. This was driven by changes in how they evaluated venture attributes like product, or potential for exit.

Organizational prompts to inquire about growth, risk and startup progress during the selection process affected how investors evaluated startups with female founders. In particular, asking investors to assess “since the beginning of the program, how much has this company improved?” in an evaluation template, significantly affected how investors scored startups on a set of criteria. Treated

investors were more likely to assess startups dynamically, paying attention to improvement during a selection process. In doing so, they evaluated startups with female founders more positively than those with all-male teams.

6.8. Alternative Explanations. We ran similar robustness checks to stage 1 on the panel data in stage 2. Differences in outcomes were robust to alternative measures of the dependent variable, the independent variable, and the female binary variable (as shown in Appendix H). It is possible that progress would be more salient to judges, or easier for startups, when startups entered the evaluation process at a less mature stage, but we found no statistically significant difference between startups with female founders and startups with male founders when they entered the evaluation process (as shown in Appendix I). This result is unsurprising because all startups passed through Vilcap’s common selection process prior to evaluation. It also suggests that this study’s results were not driven by differences in the stage at which startups entered the program.

7. Limitations

This research is an early contribution on the effects of organization-level evaluation systems on investment decisions, and has several limitations. Similar to other research in this area, including research on crowdfunding or angel platforms (e.g., Younkin and Kuppuswamy 2018, Ewens and Townsend 2020, Bapna and Ganco 2021, Gornall and Strebulaev 2020), the results in Stage 1 are limited to the first stage of investors’ selection processes. To mitigate this limitation, we tested the effects of treatments in Stage 2 and leveraged a setting where real investments were made, over time. Vilcap provided a unique opportunity to design and test multiple evaluation templates, and observe their effects on evaluation practices in a transparent process. However, similarly to most field studies, we do not trace the long-term effects of treatments. Following investors over time will subject results

to noise stemming from a fading effect of the intervention (e.g., Ridgeway and Correll 2003), an investor's follow-on experiences, or changes in the environment, so cannot fully mitigate this issue.

Also, since Vilcap observed persistent gender disparities over time, we bundled the treatment to structure inquiry, to create a strong intervention that would have a stronger chance of having a meaningful effect on a small sample in a noisy field setting. We find some evidence for both hypothesized interventions but are unable to distinguish their relative importance. We were also unable to determine a statistical effect of prompting investors to ask about risk and reward on consistency of investor inquiry. While not being able to separate out the exact mechanisms driving effects is a limitation, it is balanced by the benefit of identifying a cost-effective organization-level treatment, which affected investment outcomes and that was field-tested with real investors and entrepreneurs.

Like any field experiment, these experiments are limited by their context. In this paper, we cannot separate out the effects of the treatment by investor type (barring investor gender, where it did not change the main results). We tested the treatment in one organization's templates. This organization had created an evaluation template, used it, mined its own data, identified a disparity in evaluation by gender, and was willing to change its evaluation templates to attempt to redress gender disparities. This suggests that these findings will only hold for organizations that identify gender disparities and which are willing to redress them by changing templates. Promisingly, the effects of the organization's efforts held with investors from multiple organizations and geographic contexts, which suggests some generalizability. However, more research is needed in different contexts and with larger samples to assess the conditions under which this treatment, or other interventions to structure inquiry, can reduce gender disparities in investments.

Finally, because we tested this treatment with one dependent variable in mind – gender – we cannot assess how these or other evaluation practices might affect other disparities in investment outcomes. This treatment may have the potential to affect disparities driven by founder race (e.g.,

Younkin and Kuppuswamy 2018), or other characteristics of a founder or venture. Future research is needed to assess the effect of systematizing inquiry in these contexts. Similarly, we designed this research to isolate the effects of gender and treatments on investment decisions. As such we do not theorize other inputs to investment decision-making such as organizational preferences (Tyebjee and Bruno 1984, Cohen et al. 2019), individual investor preferences (e.g., Huang 2018), or extraneous conditions such as weather (e.g., Dushnitsky and Sarkar 2022).

8. Discussion

What is the effect of investment organizations' evaluation practices on gender disparities in funding innovation?

We hypothesized that investment organizations could change their evaluation systems to invest into female-founded startups and designed and tested interventions in a two-stage field experiment. In the first stage, we tested whether prompting investors to systematically inquire about risk and reward would reduce gender disparities in investment decisions early in an investor-founder relationship. An analysis of 1,341 decisions taken by 278 investors suggested that prompting investors to inquire about both prevention (risk) and promotion (growth) in an evaluation template resulted in more consistent inquiry – investors asked more prevention-focused questions to all startups, but particularly those with all-male founding teams. This reduced gender disparities in evaluation in the treated group compared to a control.

In the second stage, we tested whether systematizing inquiry could affect collective investment decisions that allocated \$320,000 to 16 of 69 startups over three months. Leveraging a panel dataset of 1,530 decisions, we tested the effects of a bundled treatment of prompting investors to systematically inquire about: 1) risk and reward; and 2) startup progress. Treated investors more positively assessed startups with female founders than the control. This resulted in meaningfully higher scores that affected the likelihood of investment for startups with female founders. Differential effects between treatment and control appeared to be driven by how the investors assessed startups' venture

characteristics: control investors assessed static characteristics to assess competence, while treated investors were more likely to assess ventures dynamically. This focus on what startups had accomplished in a short period of time, rather than in assessing their potential, benefitted startups with female founders. Treated investors assessed that these startups had shown competence, and demonstrated future growth potential.

Together, these results suggest that organizations can change their evaluation systems to reduce gender disparities in investment outcomes. Using a global sample with real investment decisions, our paper makes three main contributions to our understanding of the effects of evaluation systems on investment outcomes: 1) changing the system – through organizational evaluation practices – can affect funding for female founders; 2) the promise of systematizing inquiry; 3) how status biases around competence can be counteracted.

8.1. Changing the System Can Affect Funding for Female Founders. Research on investment processes in entrepreneurship has explicated the importance of gatekeepers. Because investors source deals through their homophilous networks (e.g., Saxenian 1990, 1996; Sorenson and Stuart 2001, Hallen Davis and Murray 2020), the lack of female investors results in lower evaluation for startups with female founders (Gompers and Wang 2017, Greenberg and Mollick 2017, Howell and Nanda 2019, Ewens and Townsend 2020). However, research on the effects of having female investors is mixed (e.g., Bapna and Ganco 2021, Snellman and Solal 2023, Gornall and Strebulaev 2020). Other scholars have theorized the importance of founder narratives in accessing resources (e.g., Lounsbury and Glynn 2001, Martens Jennings and Jennings 2007, Hallen and Eisenhardt 2012), or combating investors' biases through founder pitches (e.g., Kanze et al. 2018, Lee and Huang 2018, Balachandra et al. 2019, Huang et al. 2021). However, pitching tactics may not have the same effect with every investor (Panhke et al. 2015, Clough et al. 2019). Rather than focusing on networks or

interventions that put the onus on the founder, we instead theorized how investment organizations could engender changes in investment outcomes.

We considered evaluation not as a dyadic process between an investor and a founder, but as part of a collective process, designed by investment organizations, in which investors are nested. To structure evaluation processes, organizations design and use tools, templates, and evaluation frameworks (March and Simon 1958, Fayard and Metiu 2014, Anthony 2021). Scholars have found that the templates organizations use affect decisions that are made (Kaplan 2011), and the types of knowledge that organizations create (Anthony 2021). Similarly, in our research, we explain the importance of the structured evaluation templates that investors use. By designing and testing organization-level interventions to change evaluation templates, we reveal a novel organization-level mechanism that affects gender disparities in investments. Prompting investors to systematically inquire about risk and reward resulted in treated investors asking all startups more prevention-focused questions, particularly those with all male teams. Treated investors gave startups with male founders lower scores. This reduced gender disparities in both the types of questions asked by investors and the scores given – to startups with all male teams and those with female founders. Prompting investors to systematically inquire about progress, meaningfully affected how they evaluated startups with female founders. This was driven not by an improved assessment of their team, but by a dynamic consideration of their venture. Startups with female founders were considered to have demonstrated competence, and shown future potential.

By identifying how investment organizations' evaluation practices can reduce gender disparities in investment decisions, we raise a broader question of how investment organizations' evaluation tools, templates, evaluation frameworks and processes, often designed early in an organization's development, might affect investors' later assessments of startups. We suggest that

tracing a link between organizational systems of evaluation, and examining the effects on investor decisions, may be fruitful for future research that aims to explain how investment decisions are made.

8.2. The Promise of Systematizing Inquiry. When theorizing how to mitigate gender biases about people under conditions of uncertainty in a short time frame, scholars have examined the effects of blinding applications (e.g., Goldin and Rouse 2000), limiting employee discretion (Castilla 2008), shortening evaluation scales (Rivera and Tilcsik 2019), structuring interviews so that evaluators ask the same questions to all candidates (Huffcutt 2011), or rewarding proven past performance on a task, rather than allowing evaluators to assess potential (Stephens et al. 2020, Benson et al. 2021). Each of these efforts focus on creating stricter regulations and processes around evaluation, and limiting individual evaluator discretion. We hypothesized how this might be applied in innovation contexts, where investors consider not only a person, but also the potential of their startup. Past data on the startup is insufficient to produce decision-making (e.g., Stinchcombe 1965, Aldrich and Fiol 1994, Cohen et al. 2019a), and investors value the ability to use their discretion: using processes of inquiry during interactions with applicants to assess potential (Kirsch et al. 2009, Petty and Gruber 2011, Huang 2018, Miller et al. 2023). These processes of inquiry in investment can engender disparities in evaluation in innovation contexts (Kanze et al. 2018), and more broadly (Rivera 2012a, b, Rivera 2015, Stephens et al. 2020), yet their causal effects on economic outcomes are understudied.

In this context, we theorized that interventions that allowed investors to inquire freely – assessing the potential of an innovator and their idea during interactions – would be important in this context. We simultaneously hypothesized that systematizing how organizations prompt investors to inquire freely (on risk, reward and performance), can result in fewer gender disparities in individual evaluation and collective investment decisions. We theorized that organizations can create evaluation templates to reduce disparities in investment outcomes, by reducing gender disparities in investors’

processes of inquiry. We intervened in two aspects of the inquiry process where we identified where investors typically evaluated startups differently by founder gender: (1) how risk and reward were assessed; and (2) assessing progress over time. We explain how structuring processes of inquiry in this manner caused treated investors to (1) inquire consistently across all founders and (2) assess startup competence dynamically. This eliminated, even reversed, the gender gap in investment decisions. These effects held across investors' characteristics, including their organization and geography. In short, these interventions significantly changed how investors inquired, and their eventual decisions. Our research demonstrates the importance of processes of inquiry for investment outcomes by founder gender. Data collected through inquiry, particularly data collected over the selection process, using dynamic assessment, has the potential to increase the information used by investors in making collective decisions – which could outweigh the benefits of reducing individual evaluator discretion.

Extrapolating from this finding, we theorize a broader implication for many investment organizations – where evaluators assess multiple candidates, allocate funds and use processes of inquiry to gather data on potential, which is not readily available in static form. For example, such as when university hiring committees are attempting to assess the potential of an early-stage candidate's research pipeline. Structuring processes of inquiry by prompting evaluators to inquire consistently about the same content to all candidates and to record it in templates, has the potential to reduce disparities in inquiry and outcomes across these contexts.

8.3. How Status Biases around Competence Can Be Counteracted. Most scholars have theorized that disparities are caused by disadvantage – mechanisms such as prejudice, stereotypes and structural barriers (Phillips et al. 2022). When evaluators use ascriptive characteristics like gender in evaluation, they undervalue females' competence (England et al. 1994, Ridgeway and Correll 2004, Correll and Benard 2006, Botelho and Abraham 2017, Snellman & Solal 2023). Similarly to Botelho

and Abraham (2017), we focused on the processes of evaluation, and theorized disparities in the evaluation practices used by investors. This did not require an assumption that female founders' competence was necessarily undervalued by investors – and we examined this question.

In the first stage of the experiment, we found that disparities in investor inquiry around risk and reward were mitigated by treating male founders more like female founders, which resulted in less disparate investment outcomes by gender. This suggests that at least some of the discrepancy in investment outcomes is not driven by undervaluing startups with female founders, but by underquestioning, and overvaluing all-male teams. It appears that startups with all-male teams may receive advantages that female founders do not – the benefit of the doubt – in early-stage evaluation processes. In the second stage of the experiment, we found that disparities in investment outcomes reversed when investors evaluated competence dynamically, driven both by increases in evaluation outcomes for startups with female teams, and decreases for all male teams. This finding is congruent with the previous, and suggests that startups led by male founders may receive a boost when competence is assessed statically, but are less valued when investors pay attention to demonstrated competence.

Together, these findings suggest that male founders may benefit from advantaging mechanisms such as permissiveness (Phillips et al. 2022), which can affect their ability to enter further investor due diligence. This can have implications for not only female founders, but also for investment organizations, who may spend more time evaluating startups with all-male teams and less time evaluating startups led by a more diverse set of founders than is warranted. Perhaps one way to engender more equitable evaluation would be to inquire more about risk and progress, to prompt male innovators to prove their competence. This has broader implications for how status-based biases around competence can be counteracted – by focusing on advantaging mechanisms occurring within organizational evaluation processes in innovation, and perhaps beyond.

8.4. Policy Implications. Policy makers and investors are increasingly recognizing the importance of improving gender diversity in investing. Development Finance Institutions (DFIs) like the International Finance Corporation (IFC), Development Finance Corporation (DFC) and British International Investment (BII) are committing billions of dollars to invest in female-founded ventures (DFC, 2021), as are private investment funds like Fidelity and Nia Impact Capital, and venture capital firms such as Andreessen Horowitz. In this context, our findings have a number of important policy implications for reducing gender disparities in investment organizations and systems by changing the way investors evaluate startups.

First, changing investment evaluation processes without explicitly focusing on gender reduces the chance of backlash, compared to affirmative action policies (e.g., Leslie 2019). Indeed, this “levelling the assessment field” might increase efficiency in markets as investors make decisions based on more complete information. Second, these interventions are designed to retain investment organizations’ focus on identifying the most promising ventures – investors are not expected to focus on multiple goals (i.e., diversity and investment potential together).

Third, while other interventions to improve gender outcomes in investment have focused on the actions of investment seekers, we show that systems-level change is possible by changing organizational processes. At first glance, it may seem difficult for investment organizations to implement an assessment of progress in their selection processes – part of the problem in selecting early-stage startups is that they do not have a history of performance to analyze. The Vilcap investment team initially worried that their own selection processes would not allow dynamic assessment. However, after further discussion, Vilcap identified that startups could make progress between filling out their application form and having an interview with a local investor. Vilcap added a question to their interview template to assess startup progress: “Do you have any updates for us since you filled out the application form? (Has there been any change in how you think about your business or how

you execute your strategy?)” If more investment organizations prompt investors to inquire about progress and assess it, rather than only assessing potential, they could detect startups that are able to make rapid improvement – a capability important to many investors. A more structured approach to assessing signs of demonstrated progress could change how investors evaluate a startup’s potential, which seems particularly important for startups with female founders, who tend to be less rewarded for their potential. Thus, a shift to focusing on short-term signs of performance (progress) might help!

Fourth, these interventions are relatively inexpensive changes to processes, compared to costly training programs, investment guarantees, or providing supplementary funding for female-founded ventures. For DFIs, these findings offer a promising avenue to improve gender diversity in their portfolios in a cost-effective and efficient way.

Finally, although the number and types of investment organizations have increased in the US and globally (Cohen et al. 2019b, Lerner and Nanda 2020, Guttentag et al. 2021), gender disparities persist across organizational type (e.g., Ewens and Townsend 2020, Bapna and Ganco 2021) and global regions (e.g., Lall et al. 2020). Yet much of extant research has focused on investment decisions made by VC firms in the US (Drover et al. 2017, Clough et al. 2019). Our field experiment takes place in four global regions – Africa, India, Latin America, and the Middle East and North Africa – bringing together investors and startups from over 30 countries. We tested the same treatments across regions with a range of investors and am confident that the key insights from this study can be applied in many contexts, within and outside these regions.

8.5. Conclusion. Rather than considering how to improve signals that a female resource-seeker can provide to an investor, either through changing their storytelling narratives or providing female-focused funds, we suggest a novel mechanism to reduce disparities in investment outcomes. Through a two-stage field experiment with real investment decisions, we found that organizations can reduce gender disparities in investments by changing their evaluation practices. Prompting investors

to systematically inquire about risk and reward reduced disparities in scores, largely by lowering male scores. This effect appeared to be driven by investors inquiring more consistently across all founders, and conducting more rigorous diligence on startups with all-male founding teams. Prompting investors to systematically inquire about startup progress increased scores allocated to startups with female founders. This appeared to be driven by prompting a dynamic assessment of the venture, which informed investors' assessments of competence, and convinced treated investors that startups with female founders could make rapid progress and scale. Systematizing inquiry can meaningfully affect investment outcomes, which has implications for entrepreneurship theory, provides insight on the role of inquiry in evaluation, and how organizations can systematize inquiry. It has policy implications for any organization interested in investing in innovation.

References

- Aldrich, H. E., & Fiol, C. M. (1994). Fools rush in? The institutional context of industry creation. *Academy of management review*, 19(4), 645-670.
- Amis, J. M., Mair, J., & Munir, K. A. (2020). The organizational reproduction of inequality. *Academy of Management Annals*, 14(1), 195-230.
- Balachandra, L., Briggs, T., Eddleston, K., & Brush, C. (2017). Don't Pitch Like a Girl!. *Entrepreneurship Theory and Practice*, 1-22.
- Bapna, S., & Ganco, M. (2021). Gender gaps in equity crowdfunding: Evidence from a randomized field experiment. *Management Science*, 67(5), 2679-2710.
- Bartel, C. A., & Garud, R. (2009). The role of narratives in sustaining organizational innovation. *Organization science*, 20(1), 107-117.
- Benson, A., Li, D., & Shue, K. (2021). "Potential" and the Gender Promotion Gap. Working paper. SSRN.
- Bernstein, S., Korteweg, A., & Laws, K. (2017). Attracting early-stage investors: Evidence from a randomized field experiment. *The Journal of Finance*, 72(2), 509-538.
- Biegel, A., Hunt, S. M., Matteucci, A. (2020). Project SAGE 3.0: Tracking venture capital, private equity, and private debt with a gender lens. *Wharton Social Impact Initiative*. Accessed at <https://socialimpact.wharton.upenn.edu/research-reports/reports-2/project-sage-3/>
- Bingham, C. B., & Eisenhardt, K. M. (2011). Rational heuristics: the 'simple rules' that strategists learn from process experience. *Strategic management journal*, 32(13), 1437-1464.
- Bohren, J. A., Imas, A., & Rosenberg, M. (2019). The dynamics of discrimination: Theory and evidence. *American economic review*, 109(10), 3395-3436.
- Botelho, T. L., & Abraham, M. (2017). Pursuing quality: How search costs and uncertainty magnify gender-based double standards in a multistage evaluation process. *Administrative Science Quarterly*, 62(4), 698-730.
- Brooks, A. W., Huang, L., Kearney, S. W., & Murray, F. E. (2014). Investors prefer entrepreneurial ventures pitched by attractive men. *Proceedings of the National Academy of Sciences*, 111(12), 4427-4431.
- Burns, A., Tashima, R., Matranga, H. (2019). Flipping the Power Dynamics: Can Entrepreneurs Make Successful Investment Decisions?. *Village Capital*.
- Canales, R. (2014). Weaving straw into gold: Managing organizational tensions between standardization and flexibility in microfinance. *Organization Science*, 25(1), 1-28.
- Castilla, E. J. (2008). Gender, race, and meritocracy in organizational careers. *American journal of sociology*, 113(6), 1479-1526.
- Castilla, E. J. (2015). Accounting for the gap: A firm study manipulating organizational accountability and transparency in pay decisions. *Organization Science*, 26(2), 311-333.
- Castilla, E. J., & Benard, S. (2010). The paradox of meritocracy in organizations. *Administrative science quarterly*, 55(4), 543-676.
- Clingingsmith, D., & Shane, S. (2018). Training aspiring entrepreneurs to pitch experienced investors: Evidence from a field experiment in the United States. *Management Science*, 64(11), 5164-5179.
- Clough, D. R., Fang, T. P., Vissa, B., & Wu, A. (2019). Turning lead into gold: How do entrepreneurs mobilize resources to exploit opportunities?. *Academy of Management Annals*, 13(1), 240-271.
- Cohen, S. L., Bingham, C. B., & Hallen, B. L. (2019a). The role of accelerator designs in mitigating bounded rationality in new ventures. *Administrative Science Quarterly*, 64(4), 810-854.
- Cohen, S., Fehder, D. C., Hochberg, Y. V., & Murray, F. (2019b). The design of startup accelerators. *Research Policy*, 48(7), 1781-1797.
- Correll, S. J., & Benard, S. (2006). Biased estimators? Comparing status and statistical theories of gender discrimination. In *Advances in group processes* (Vol. 23, pp. 89-116). Emerald Group Publishing Limited.
- Correll, S. J., Ridgeway, C. L., Zuckerman, E. W., Jank, S., Jordan-Bloch, S., & Nakagawa, S. (2017). It's the conventional thought that counts: How third-order inference produces status advantage. *American Sociological Review*, 82(2), 297-327.
- Cortes, E. (2019). Gender-Lens Investing Strategies for 2019. *Stanford Social Innovation Review*.
- Czibor, E., Jimenez-Gomez, D., & List, J. A. (2019). The dozen things experimental economists should do (more of). *Southern Economic Journal*, 86(2), 371-432.
- Da Rin, M., Hellmann, T., & Puri, M. (2013). A survey of venture capital research. In *Handbook of the Economics of Finance* (Vol. 2, pp. 573-648). Elsevier.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American psychologist*, 34(7), 571.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243(4899), 1668-1674.

- Delecourt, S., & Ng, O. (2021). Does gender matter for small business performance? Experimental evidence from India. *Experimental Evidence from India (April 21, 2021)*.
- DFC. (2021). Global Gender Finance Initiative Sets Ambitious New \$15 Billion Fundraising Goal After Securing More Than Double Its Original \$3 Billion Target. Accessed at: <https://www.dfc.gov/media/press-releases/global-gender-finance-initiative-sets-ambitious-new-15-billion-fundraising>
- Dimitriadis, S., & Koning, R. (2022). Social skills improve business performance: evidence from a randomized control trial with entrepreneurs in Togo. *Management Science*.
- Dobbin, F., Kim, S., & Kalev, A. (2011). You can't always get what you need: Organizational determinants of diversity programs. *American Sociological Review*, 76(3), 386-411.
- Drover, W., Busenitz, L., Matusik, S., Townsend, D., Anglin, A., & Dushnitsky, G. (2017). A review and road map of entrepreneurial equity financing research: venture capital, corporate venture capital, angel investment, crowdfunding, and accelerators. *Journal of management*, 43(6), 1820-1853.
- Dutt, N., & Kaplan, S. (2018, July). Acceleration as Mitigation: Whether & When Processes Can Address Gender Bias in Entrepreneurship. In *Academy of Management Proceedings* (Vol. 2018, No. 1, p. 16160). Briarcliff Manor, NY 10510: Academy of Management.
- Eisenhardt, K. M., & Tabrizi, B. N. (1995). Accelerating adaptive processes: Product innovation in the global computer industry. *Administrative science quarterly*, 84-110.
- Elsbach, K. D., & Kramer, R. M. (2003). Assessing creativity in Hollywood pitch meetings: Evidence for a dual-process model of creativity judgments. *Academy of Management journal*, 46(3), 283-301.
- England, P., Herbert, M. S., Kilbourne, B. S., Reid, L. L., & Megdal, L. M. (1994). The gendered valuation of occupations and skills: Earnings in 1980 census occupations. *Social Forces*, 73(1), 65-100.
- Ewens, M., & Townsend, R. R. (2020). Are early stage investors biased against women?. *Journal of Financial Economics*, 135(3), 653-677.
- Fried, V. H., & Hisrich, R. D. (1994). Toward a model of venture capital investment decision making. *Financial management*, 28-37.
- Frost, (2020). How VC bias in viewing pitch decks can affect fundraising success. *Docsend*. Accessed at: <https://www.docsend.com/blog/how-vc-bias-in-viewing-pitch-decks-can-affect-fundraising-success/>
- Gamache, D. L., McNamara, G., Mannor, M. J., & Johnson, R. E. (2015). Motivated to acquire? The impact of CEO regulatory focus on firm acquisitions. *Academy of Management Journal*, 58(4), 1261-1282.
- Gans, J. S., Stern, S., & Wu, J. (2019). Foundations of entrepreneurial strategy. *Strategic Management Journal*, 40(5), 736-756.
- Garud, R., Schildt, H. A., & Lant, T. K. (2014). Entrepreneurial storytelling, future expectations, and the paradox of legitimacy. *Organization Science*, 25(5), 1479-1492.
- Gavetti, G., Levinthal, D., & Ocasio, W. (2007). Perspective—Neo-Carnegie: The Carnegie school's past, present, and reconstructing for the future. *Organization Science*, 18(3), 523-536.
- Gavetti, G., Greve, H. R., Levinthal, D. A., & Ocasio, W. (2012). The behavioral theory of the firm: Assessment and prospects. *Academy of Management Annals*, 6(1), 1-40.
- Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American economic review*, 90(4), 715-741.
- Gompers, P. A., & Wang, S. Q. (2017). *And the children shall lead: Gender diversity and performance in venture capital* (No. w23454). National Bureau of Economic Research.
- Gompers, P. A., Gornall, W., Kaplan, S. N., & Strebulaev, I. A. (2020). How do venture capitalists make decisions?. *Journal of Financial Economics*, 135(1), 169-190.
- González-Urbe, J., & Reyes, S. (2021). Identifying and boosting "Gazelles": Evidence from business accelerators. *Journal of Financial Economics*, 139(1), 260-287.
- Greenberg, J., & Mollick, E. (2017). Activist choice homophily and the crowdfunding of female founders. *Administrative Science Quarterly*, 62(2), 341-374.
- Guttentag, M., Davidson, A., & Hume, V. (2021). Does Acceleration Work? Five years of evidence from the Global Accelerator Learning Initiative. *Aspen Network of Development Entrepreneurs & The Roberto C. Goizueta Business & Society Institute*. Accessed at: <https://www.galidata.org/publications/does-acceleration-work/>
- Guzman, J., & Kacperczyk, A. O. (2019). Gender gap in entrepreneurship. *Research Policy*, 48(7), 1666-1680.
- Hallen, B. L. (2008). The causes and consequences of the initial network positions of new organizations: From whom do entrepreneurs receive investments?. *Administrative Science Quarterly*, 53(4), 685-718.
- Hallen, B. L., Davis, J. P., & Murray, A. (2020). Entrepreneurial network evolution: Explicating the structural localism and agentic network change distinction. *Academy of Management Annals*, 14(2), 1067-1102.
- Hallen, B. L., & Eisenhardt, K. M. (2012). Catalyzing strategies and efficient tie formation: How entrepreneurial firms obtain investment ties. *Academy of Management Journal*, 55(1), 35-70.

- Higgins, M. C., & Gulati, R. (2003). Getting off to a good start: The effects of upper echelon affiliations on underwriter prestige. *Organization science*, 14(3), 244-263.
- Higgins, M. C., & Gulati, R. (2006). Stacking the deck: The effects of top management backgrounds on investor decisions. *Strategic Management Journal*, 27(1), 1-25.
- Howell, S. T., & Nanda, R. (2019). *Networking frictions in venture capital, and the gender gap in entrepreneurship* (No. w26449). National Bureau of Economic Research.
- Hsieh, C. T., Hurst, E., Jones, C. I., & Klenow, P. J. (2019). The allocation of talent and US economic growth. *Econometrica*, 87(5), 1439-1474.
- Hsu, D. K., Simmons, S. A., & Wieland, A. M. (2017). Designing entrepreneurship experiments: A review, typology, and research agenda. *Organizational Research Methods*, 20(3), 379-412.
- Huang, L. (2018). The role of investor gut feel in managing complexity and extreme risk. *Academy of Management Journal*, 61(5), 1821-1847.
- Huang, L., & Pearce, J. L. (2015). Managing the unknowable: The effectiveness of early-stage investor gut feel in entrepreneurial investment decisions. *Administrative Science Quarterly*, 60(4), 634-670.
- Huang, L., Joshi, P., Wakslak, C., & Wu, A. (2021). Sizing up entrepreneurial potential: Gender differences in communication and investor perceptions of long-term growth and scalability. *Academy of Management Journal*, 64(3), 716-740.
- Jeppesen, L. B., & Lakhani, K. R. (2010). Marginality and problem-solving effectiveness in broadcast search. *Organization science*, 21(5), 1016-1033.
- Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American sociological review*, 71(4), 589-617.
- Kanze, D., Huang, L., Conley, M. A., & Higgins, E. T. (2017). Male and female entrepreneurs get asked different questions by VCs—and it affects how much funding they get. *Harvard Business Review*, 27.
- Kanze, D., Huang, L., Conley, M. A., & Higgins, E. T. (2018). We ask men to win and women not to lose: Closing the gender gap in startup funding. *Academy of Management Journal*, 61(2), 586-614.
- Kaplan, S. N., Sensoy, B. A., & Strömberg, P. (2009). Should investors bet on the jockey or the horse? Evidence from the evolution of firms from early business plans to public companies. *The Journal of Finance*, 64(1), 75-115.
- Kaplan, S. N., Klebanov, M. M., & Sorensen, M. (2012). Which CEO characteristics and abilities matter?. *The journal of finance*, 67(3), 973-1007.
- Kirsch, D., Goldfarb, B., & Gera, A. (2009). Form or substance: the role of business plans in venture capital decision making. *Strategic Management Journal*, 30(5), 487-515.
- Kirtley, J., & O'Mahony, S. (2020). What is a pivot? Explaining when and how entrepreneurial firms decide to make strategic change and pivot. *Strategic Management Journal*.
- Koning, R., Samila, S., & Ferguson, J. P. (2020, May). Inventor Gender and the Direction of Invention. In *AEA Papers and Proceedings* (Vol. 110, pp. 250-54).
- Lall, S. A., Chen, L. W., & Roberts, P. W. (2020). Are we accelerating equity investment into impact-oriented ventures?. *World Development*, 131, 104952.
- Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*.
- Lee, M., & Huang, L. (2018). Gender bias, social impact framing, and evaluation of entrepreneurial ventures. *Organization Science*, 29(1), 1-16.
- Leibbrandt, A., Wang, L. C., & Foo, C. (2018). Gender quotas, competitions, and peer review: Experimental evidence on the backlash against women. *Management Science*, 64(8), 3501-3516.
- Lerner, J., & Nanda, R. (2020). Venture Capital's Role in Financing Innovation: What We Know and How Much We Still Need to Learn. *Journal of Economic Perspectives*, 34(3), 237-61.
- Leslie, L. M. (2019). Diversity initiative effectiveness: A typological theory of unintended consequences. *Academy of Management Review*, 44(3), 538-563.
- Lounsbury, M., & Glynn, M. A. (2001). Cultural entrepreneurship: Stories, legitimacy, and the acquisition of resources. *Strategic management journal*, 22(6-7), 545-564.
- Luo, H., & Zhang, L. (2022). Scandal, social movement, and change: Evidence from# metoo in 44ollywood. *Management Science*, 68(2), 1278-1296.
- Macan, T. (2009). The employment interview: A review of current studies and directions for future research. *Human Resource Management Review*, 19(3), 203-218.
- March, J.G., & Simon, H.A. (1958). *Organizations*. Wiley.
- Martens, M. L., Jennings, J. E., & Jennings, P. D. (2007). Do the stories they tell get them the money they need? The role of entrepreneurial narratives in resource acquisition. *Academy of management journal*, 50(5), 1107-1132.

- McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of development Economics*, 99(2), 210-221.
- Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of business venturing*, 29(1), 1-16.
- National Venture Capital Association (2020). *Yearbook*. Accessed at: <https://nvca.org/research/nvca-yearbook/>
- Pahnke, E. C., Katila, R., & Eisenhardt, K. M. (2015). Who takes you to the dance? How partners' institutional logics influence innovation in young firms. *Administrative science quarterly*, 60(4), 596-633.
- Petty, J. S., & Gruber, M. (2011). "In pursuit of the real deal": A longitudinal study of VC decision making. *Journal of Business Venturing*, 26(2), 172-188.
- Phillips, L. T., Jun, S., & Shakeri, A. (2022). Barriers and boosts: Using inequity frames theory to expand understanding of mechanisms of race and gender inequity. *Academy of Management Annals*, 16(2), 547-587.
- Ployhart, R. E., Schneider, B., & Schmitt, N. (2005). *Staffing organizations: Contemporary practice and theory*. CRC Press.
- Ridgeway, C. L., & Correll, S. J. (2004). Unpacking the gender system: A theoretical perspective on gender beliefs and social relations. *Gender & Society*, 18(4), 510-531.
- Rivera, L. A. (2012). Hiring as cultural matching: The case of elite professional service firms. *American sociological review*, 77(6), 999-1022.
- Rivera, L. A. (2015). Go with your gut: Emotion and evaluation in job interviews. *American journal of sociology*, 120(5), 1339-1389.
- Rivera, L. A., & Tilcsik, A. (2019). Scaling down inequality: Rating scales, gender bias, and the architecture of evaluation. *American Sociological Review*, 84(2), 248-274.
- Roberts, P. W., & Lall, S. A. (2018). *Observing acceleration: Uncovering the effects of accelerators on impact-oriented entrepreneurs*. Springer.
- Saxenian, A. (1990). Regional networks and the resurgence of Silicon Valley. *California management review*, 33(1), 89-112.
- Saxenian, A. (1996). *Regional advantage*. Harvard University Press.
- Siggelkow, N. (2002). Evolution toward fit. *Administrative science quarterly*, 47(1), 125-159.
- Snellman, K., & Solal, I. (2023). Does investor gender matter for the success of female entrepreneurs? Gender homophily and the stigma of incompetence in entrepreneurial finance. *Organization Science*, 34(2), 680-699.
- Sorenson, O., & Stuart, T. E. (2001). Syndication networks and the spatial distribution of venture capital investments. *American journal of sociology*, 106(6), 1546-1588.
- Stephens, N. M., Rivera, L. A., & Townsend, S. S. (2020). What works to increase diversity? A multi-level approach. *Research in Organizational Behavior*, 1-51.
- Stinchcombe A. L., (1965) *Social structure and organizations*. (Emerald Group Publishing Limited).
- Tyebjee, T. T., & Bruno, A. V. (1984). A model of venture capitalist investment activity. *Management Science*, 30(9), 1051-1066.
- Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science*, 16(6), 474-480.
- Yang, T., & Aldrich, H. E. (2014). Who's the boss? Explaining gender inequality in entrepreneurial teams. *American Sociological Review*, 79(2), 303-327.
- Younkin, P., & Kuppaswamy, V. (2018). The colorblind crowd? Founder race and performance in crowdfunding. *Management Science*, 64(7), 3269-3287.
- Zacharakis, A. L., & Meyer, G. D. (2000). The potential of actuarial decision models: can they improve the venture capital investment decision?. *Journal of Business venturing*, 15(4), 323-346.
- Zott, C., & Huy, Q. N. (2007). How entrepreneurs use symbolic management to acquire resources. *Administrative science quarterly*, 52(1), 70-105.
- Zuckerman, E. W. (2012). Construction, concentration, and (dis) continuities in social valuations. *Annual Review of Sociology*. 38(1), 223-245.

Tables and Figures

Table 1: Whole Sample

		Whole sample (scores)	Cross-section	Balanced Panel (3 rounds)
Investment decisions	Per investor*startup*round* criterion	33,541	n/a	n/a
	Per investor*startup*round	3,127	n/a	1,530
	Per investor*startup	1,342	1,341	510
Female founder	Female founder present	16,024	614	726
Female investor	Female decision-maker (a)	8,948	402	409
Systematizing inquiry	On risk/reward and progress (b)	17,920	n/a	717
	On risk/ reward (c)	660	653	n/a
Region	Africa	5,779	385	276
	India	6,626	210	294
	Latin America	8,929	460	486
	MENA	12,180	286	474
Round (in panel)	Round 0	1,354	1,341	Baseline
	Round 1	10,464	n/a	510
	Round 2	10,160	n/a	510
	Round 3	9,736	n/a	510
Investors		278	278	65
Startups		87	87	69

Pre-registered analyses on cross-section and panel. Exploratory research on whole sample.

- (0) Full sample = 31,680 – some investors did not specify their gender. (b) Panel sample = 30,373. (c) Full sample only Cross-section Round 0 = 1,341.

Table 2: Cross-Section Sample

Panel A) Investors		All		Control		Treated: Risk/Reward		Both	Treatment and Control Check*
		#	%	#	%	#	%	#	%
Investors		278		133		127		18	
Region**	Africa	82	29%	36	27%	42	33%	4	22%
	India	36	13%	19	14%	16	13%	1	6%
	Latin America	73	26%	40	30%	33	26%	0	0%
	MENA	87	31%	38	29%	36	28%	13	72%
Role	Trainee	80	29%	41	31%	39	31%	0	0%
	Professional	198	71%	92	69%	88	69%	18	100%
Gender (N=262)	Female	87	33%	42	33%	40	34%	5	29%
Type (N=179)	Investor***	68	38%	29	35%	29	37%	10	59%
	Local organization	141	84%	66	85%	62	86%	13	76%
Investment organization (N=154)	Diversity mandate	45	29%	19	26%	22	33%	4	27%
	Impact mandate	93	60%	43	59%	39	59%	11	73%
B) Investor Decisions (1,341)		All		Control		Treated: Risk/Reward		Treatment and Control Check*	
		#	%	#	%	#	%		
Decisions	Investor x startup	1,341		689		652			
Sample	Trainee	795	59%	412	60%	383	59%		
	Professional	546	41%	276	40%	270	41%		
Region	Africa	385	29%	188	27%	197	30%	0.261	
	India	210	16%	110	16%	100	15%		
	Latin America	286	21%	147	21%	139	21%		
	MENA	460	34%	243	35%	217	33%		
Female investor (N=1,320)		402	30%	196	28%	206	32%	0.235	
C) Professional Investor Decisions (546)		All		Control		Treated: Risk/Reward		Treatment and Control Check*	
		#	%	#	%	#	%		
Decisions	Investor x startup	546							
Type (N=500)	Self-identify as Investor	203	41%	97	36%	106	46%	0.299	
	Local organization	393	84%	200	82%	193	85%	0.365	
Investment organization (N=425)	Diversity mandate	120	28%	55	26%	65	31%	0.295	
	Impact mandate	259	61%	128	60%	131	62%	0.632	

* Regression of each variable on Treated to assess differences across treatment groups. P-value reported.

**Two investors evaluated firms in Africa and MENA

***From investment organization (VC firm, angel group, accelerator, venture studio) or angel investor
Number of startups evaluated = 87. (78 by control investors, 80 by treatment investors).

16 missing investors selected to remain anonymous.

Table 3: Effect of Systematizing Inquiry by Prompting Risk and Reward (Cross-section)

Round 0, Treatment and Control, Cross-sectional

	Investment Decision: DD Score (Scale 1-6). Ordered Logit, or					Prevention Question Asked (Binary). Logit, or			
	1	2	3	4	5	1	2	3	4
Female founder	0.809* (0.0727)	0.666**	0.645**	0.611**	0.641**	1.086 (0.118)	1.216 (0.209)	1.124 (0.218)	1.096 (0.216)
Inquiry on Risk/Reward		0.775 (0.158)	0.784 (0.162)	0.761 (0.161)	0.724 (0.147)		2.648*** (0.544)	2.229*** (0.481)	2.277*** (0.493)
FF*Inquiry		1.491* (0.288)	1.626* (0.348)	1.750* (0.382)	1.614* (0.344)		0.778 (0.181)	0.950 (0.233)	0.961 (0.241)
<i>Inquiry on Risk/Reward</i>					1.571***_p (0.222)				
Clustered errors (Investor)	x	x	x	x	x	x	x	x	x
FE region	x	x	x	x	x	x	x	x	x
Startup controls			x	x	x			x	x
Female investor				x					x
N*	1,341	1,341	1,162	1,133	1,162	1,341	1,341	1,162	1,133
R-sq / Pseudo R-Sq	0.0023	0.0034	0.0040	0.0053	0.0084	0.0056	0.0381	0.0319	0.0347
Investors	278	278	276	260	276	278	278	276	260

Odds Ratio reported

Models 1-4 provide evidence for similar relationships between the variables

Model 3 was pre-registered, and all other models provide similar directional results

Model 4 suggests that the gender of the investor does not meaningfully change the main relationships between variables

Model 5 suggests partial mediation

Table 4: Panel Sample

Investor and Startup Characteristics Across T&C		All		Control	Treated: Systematizing Inquiry	Treatment & Control Check*
		#	%	#	#	
Investors		65	100.0%	34	31	.
Region	Africa	14	21.5%	9	5	0.325
	India	14	21.5%	7	7	.
	MENA	19	29.2%	10	9	.
	LatAm	18	27.7%	8	10	.
Female investor		19	29.2%	9	10	0.615
Startups		69	100.0%	36	33	.
Female founder		32	46.4%	15	17	0.420
Region	Africa	14	20.3%	9	5	0.334
	India	16	23.2%	8	8	.
	MENA	19	27.5%	10	9	.
	LatAm	20	29.0%	9	11	.
Employees (mean)**		67	10.9	9.71	12.22	0.346
Funds raised (mean)**		64	\$237,896	\$179,152	\$296,640	0.161
Log funds raised (mean)**		64	10.12	9.89	10.44	0.568

All investors in the Panel Sample are Vilcap trainees.

* Regression of each variable on Treated to assess differences across treatment groups. P-value reported.

** Data unavailable for all startups

Table 5: Effect of Systematizing Inquiry on Investment Decisions (Panel)

	ANCOVA – DV Zscore (Rounds 1-3)			
	1	2	3	4
Female founder	0.075 (0.070)	-0.058 (0.096)	-0.058 (0.115)	-0.056 (0.116)
Systematized Inquiry		-0.119 (0.070)	-0.157* (0.076)	-0.156* (0.077)
FF*Inquiry		0.277* (0.132)	0.306* (0.151)	0.304* (0.152)
Baseline score	0.232*** (0.037)	0.229*** (0.037)	0.175*** (0.040)	0.175*** (0.040)
Systematized Inquiry + FF*Inquiry = 0		0.158* (0.068)	0.148 (0.088)	0.147 (0.089)
FF + FF*Inquiry = 0		0.219* (0.093)	0.248* (0.099)	0.248* (0.015)
Clustered errors (Investor)	x	x	x	x
FE region and round	x	x	x	x
Startup controls			x	x
Female judge				x
N	1,530	1,530	1,395	1,395
R-sq	0.0496	0.0544	0.0751	0.0752
Investors	65	65	65	65

Startup controls = number of employees, log (funds raised)

Table 6: Effect of Systematizing Inquiry on How Startups Were Assessed

Investor	Data	Evaluation Criteria	How Competence Assessed	Usage	
				Control	Treated
Control, Female	I rated XX highly...because I think their business idea is just really necessary... I see its use and purpose.	Value proposition	Static startup elements	9 investors (100%)	3 investors (43%)
Control, Female	People that I rated highly... [I though] “Oh, I like this idea, it’s fantastic” and you will just have to overlook every other thing.	Product			
Control, Male	I really like their solution... I think it’s relevant. It has a lot of potential for scaling ...I was quite impressed with the profiles of the people that work in the team...their business model too.	Product, scale, team, business model			
Control, Male	I scored XX highly on their tech, ‘cause I do understand that tech is a game changer in this space... I’ve interacted with their product before, so I had no doubt when giving them the biggest score.	Product, market			
Treated, Female	XX’s business model is really clear, and they have this differentiation... he makes progress... He collected data to understand that how people are working... there are a lot of people doing like freelancer platforms, so I tried to make them realize that the differentiation part was more important.	Business model, progress	Considered dynamic progress in improving startup elements	2 investors (22%)	6 investors (83%)
Treated, Female	I had a great discussion, maybe two times with both [companies]...they have a huge market... YY partnered with the telecom [company] , which is even better... when I see the partnerships, that’s where you can scale... and their team is so strong... They know the next steps in policy rules, regulation.	Market, partnerships, team, next steps			
Treated, Male	YY are trying to create a community of people who can democratize that access to content and also make a living at the same time and challenge one another... the challenge I had was in their business model... if they’re able to fix that bit through this program, they will really do incredible things ... if they get the advice they need and they get the talent to do their growth hacking and processes.	Business model, team, progress			
Treated, Male	They have a solid platform and a solid go to market that is going to have a high chance of success, with not only their customers, but with investors... they were also... getting clients and XX mentioned they just gotta deal with Partner.	Business model, team, progress			

Table 7: Effect of Systematizing Inquiry on Score Elements (Panel)

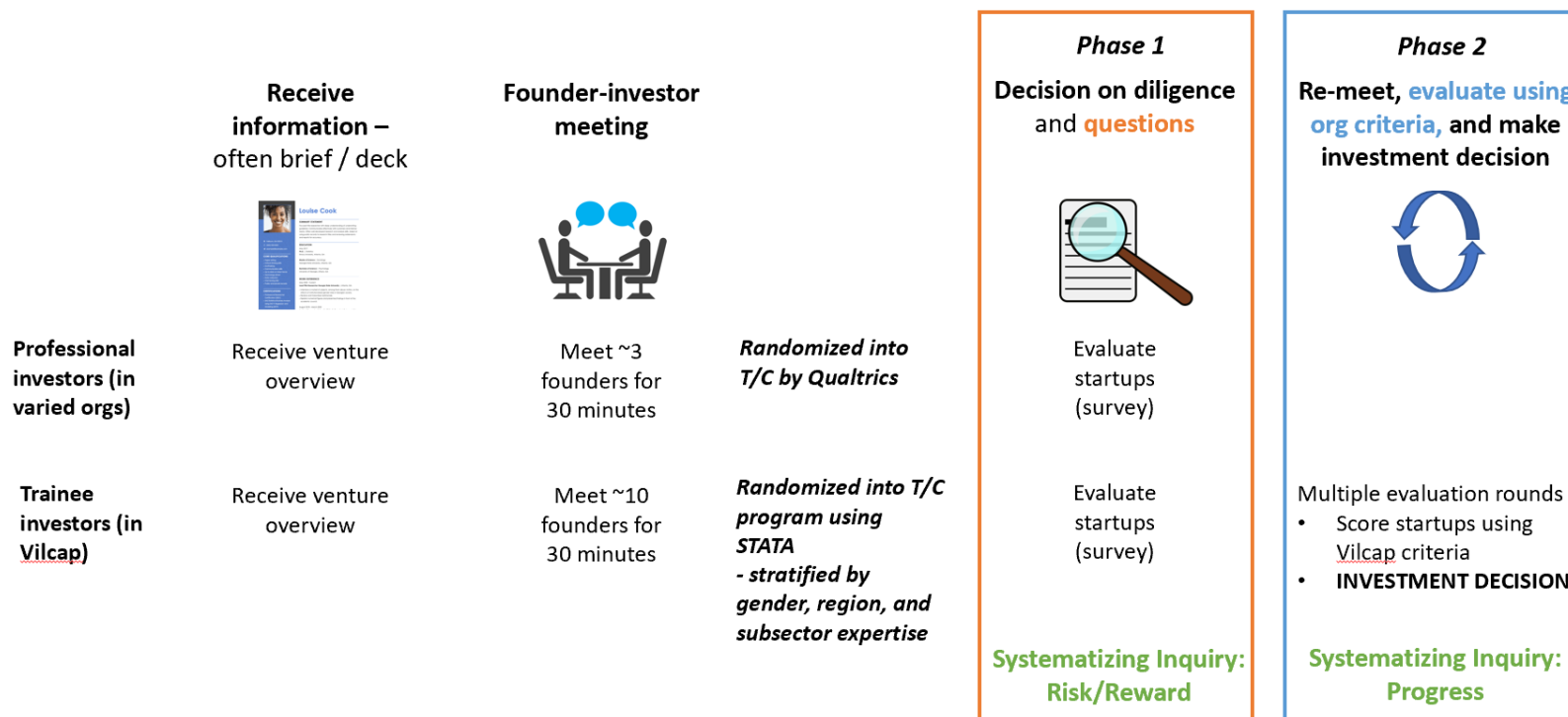
	DV Score Elements on scales of 1-4 (Rounds 1-3)								
	All	Business Model	Investor Exit	Market	Problem and Vision	Product	Scale	Team	Value Proposition
Female founder	-0.012 (0.016)	-0.021 (0.017)	-0.035 (0.024)	-0.009 (0.015)	0.002 (0.020)	-0.018 (0.019)	-0.017 (0.018)	0.005 (0.017)	-0.008 (0.018)
Systematizing Inquiry	-0.002 (0.019)	0.000 (0.022)	-0.006 (0.023)	-0.004 (0.020)	-0.009 (0.025)	-0.002 (0.021)	-0.007 (0.020)	0.010 (0.022)	-0.002 (0.023)
FF*Inquiry	0.040 (0.021)	0.046 (0.023)	0.061* (0.028)	0.032 (0.020)	0.032 (0.027)	0.055* (0.026)	0.044 (0.024)	0.036 (0.023)	0.033 (0.024)
Clustered errors (Investor)	x	x	x	x	x	x	x	x	x
FE region and round	x	x	x	x	x	x	x	x	x
Baseline score	x	x	x	x	x	x	x	x	x
Startup controls	x	x	x	x	x	x	x	x	x
N	22,320	2,790	2,790	2,790	2,790	2,790	2,790	2,790	2,790
R-sq	0.0260	0.0321	0.0543	0.0248	0.0307	0.0364	0.0375	0.0495	0.0323
Investors	65	65	65	65	65	65	65	65	65

Startup controls = number of employees, funds raised (log)

Criteria “All”, “Business model”, and “Scale” were significant at $p < 0.1$

Figures

Figure 1. Processes in Seeking Information and Evaluating Criteria



For more details see: <https://www.socialscienceregistry.org/trials/7685>.

Figure 2. Effect of Systematizing Inquiry on Investor Evaluation (Cross-Section)

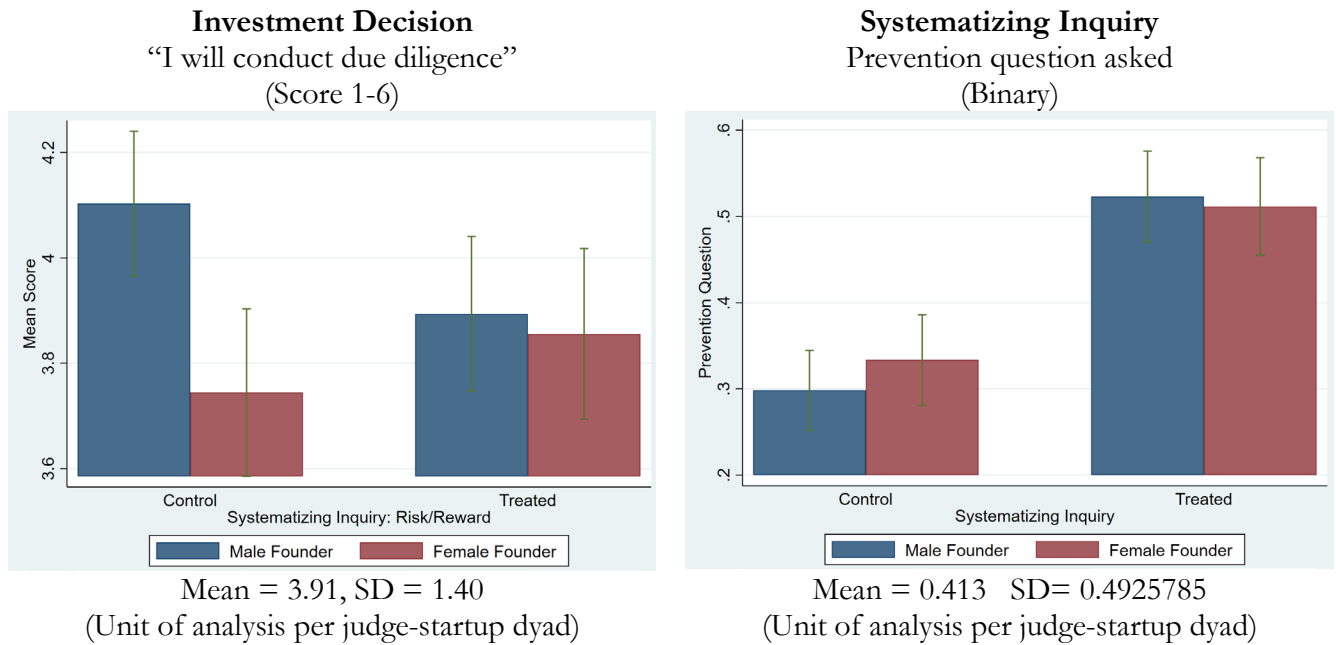
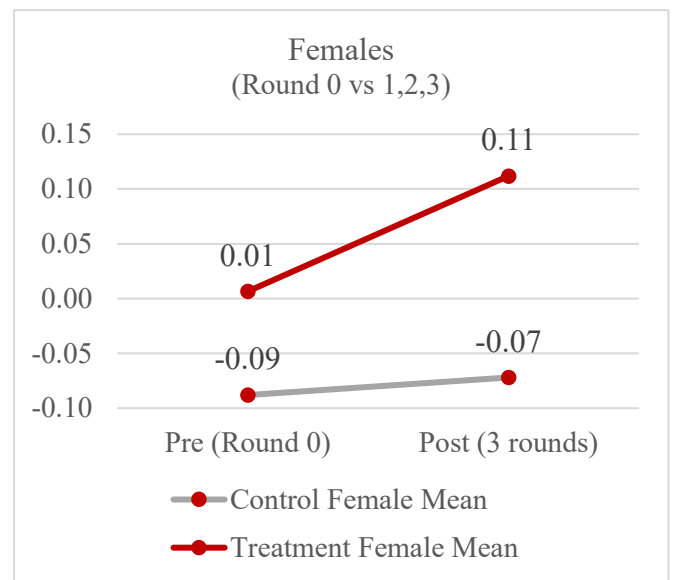
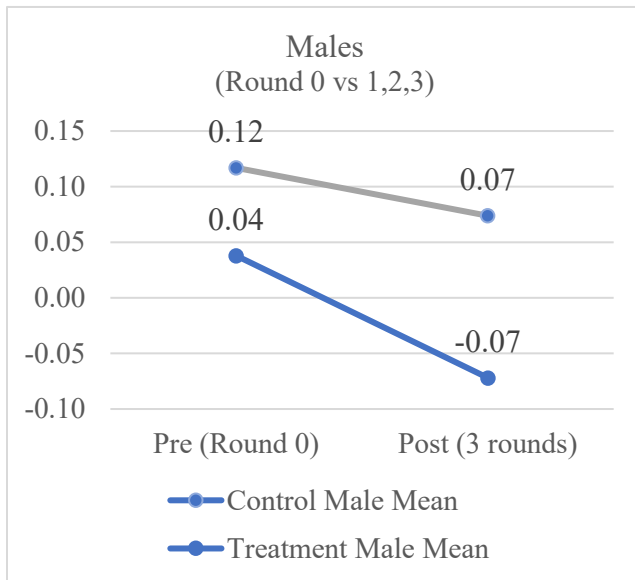


Figure 3. Effect of Systematizing Inquiry on Investment Decisions (Zscore Over Time – Panel)

Panel A: Mean Zscore Over Time – Males

Panel B: Mean Zscore Over Time – Females



Appendices

Appendix A1: Survey for Cross-Section with Treatment

SELECTED COMPANIES

Please select the companies you met today. [Multiple choice]

[Company 1 – 12]

RATING

For [selected company 1], do you agree or disagree with the following statements?

	Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly agree
I will pursue a follow-up meeting to learn more about the venture	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would be interested in seeing a business plan for this venture	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I will recommend this opportunity to a co-investor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I will initiate due diligence on this venture	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

INQUIRY CONTROL

What additional information would you want on this venture? (We will share this answer with the entrepreneurs.) [Open text]

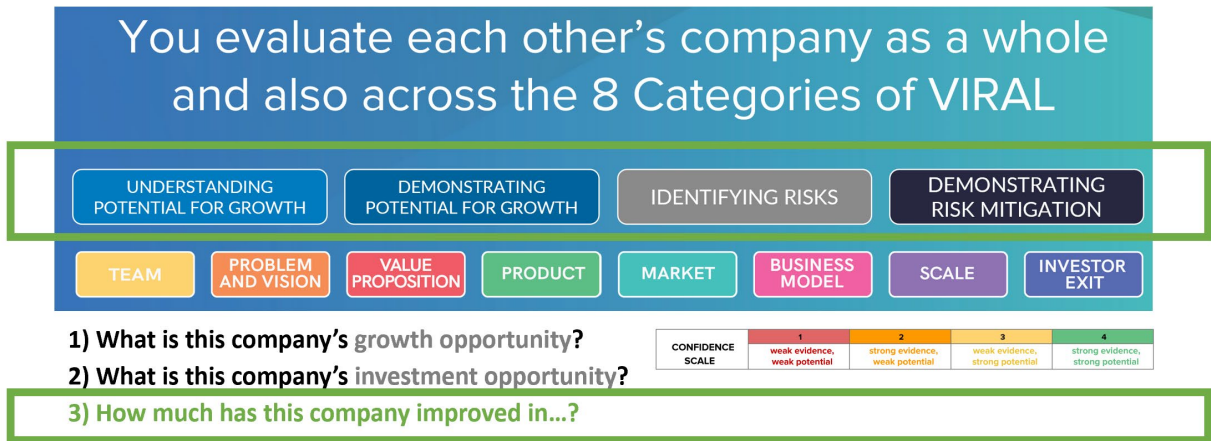
SYSTEMATIZING INQUIRY – RISK/REWARD TREATMENT

What additional information would you want on this venture's potential for growth? (We will share this answer with the entrepreneurs.) [Open text]

What additional information would you want on how this venture will mitigate risks? (We will share this answer with the entrepreneurs.) [Open text]

[Repeat for all companies they met]

Appendix A2: Panel Evaluation Template



Appendix A3: Panel Evaluation Template – Question Examples

Team Questions:

	Team	Vision	Value Prop	Product
GROWTH OPPORTUNITY	How confident are you that this team will deliver results and can make the right hires as it grows?			
INVESTMENT OPPORTUNITY	How confident are you that the team has the tenacity and drive to navigate risks and roadblocks and do what it takes to provide investors with a return?			

Progress Questions:

We ask four questions in this part of the assessment:

1. Since the beginning of the program, how much has this company improved in **understanding its path to growth**?
2. Since the beginning of the program, how much has this company improved in **demonstrating its path to growth**?
3. Since the beginning of the program, how much has this company improved in **identifying its risks**?
4. Since the beginning of the program, how much has this company improved in **demonstrating its risk mitigation**?

Appendix B: Baseline Control Group Scores by Founder Gender (Cross-section)

	Round 0, Control - Cross-Sectional			
	Prevention Question (Binary)		DD Score (Scale 1-6)	
	Logit odds ratio		Ordered logit odds ratio	
Female founder	0.240 (0.159)	0.150 (0.206)	0.635*** (0.082)	0.589** (0.097)
Clustered errors (Investor)	x	x	x	x
FE region	x	x	x	x
Startup controls		x		x
N	688	581	688	581
R-sq / Pseudo	0.0116	0.0080	0.0058	0.0079
Investors	151	151	151	151

Startup controls = number of employees, funds raised.

Appendix C: Robustness Check – Replaced Female Variable with Venture Overview

	Round 0, Treatment and Control, Cross-sectional			
	Prevention Question (Binary)		DD Score (Scale 1-6)	
	Logit odds ratio		Ordered logit odds ratio	
	3	4	3	4
Female venture overview	1.418 (0.375)	1.347 (0.367)	0.571* (0.136)	0.513** (0.125)
Inquiry on Risk/Reward	2.524*** (0.540)	2.575*** (0.553)	0.788 (0.167)	0.770 (0.166)
FF* Inquiry Risk/Reward	0.635 (0.220)	0.646 (0.230)	1.977* (0.602)	2.166* (0.670)
<i>% prevention questions</i>				
Clustered errors (Investor)	x	x	x	x
FE region	x	x	x	x
Startup controls	x	x	x	x
Female investor		x		x
N	1,162	1,133	1,162	1,133
R-sq	0.0328	0.0356	0.0037	0.0050
Investors	276	260	276	260

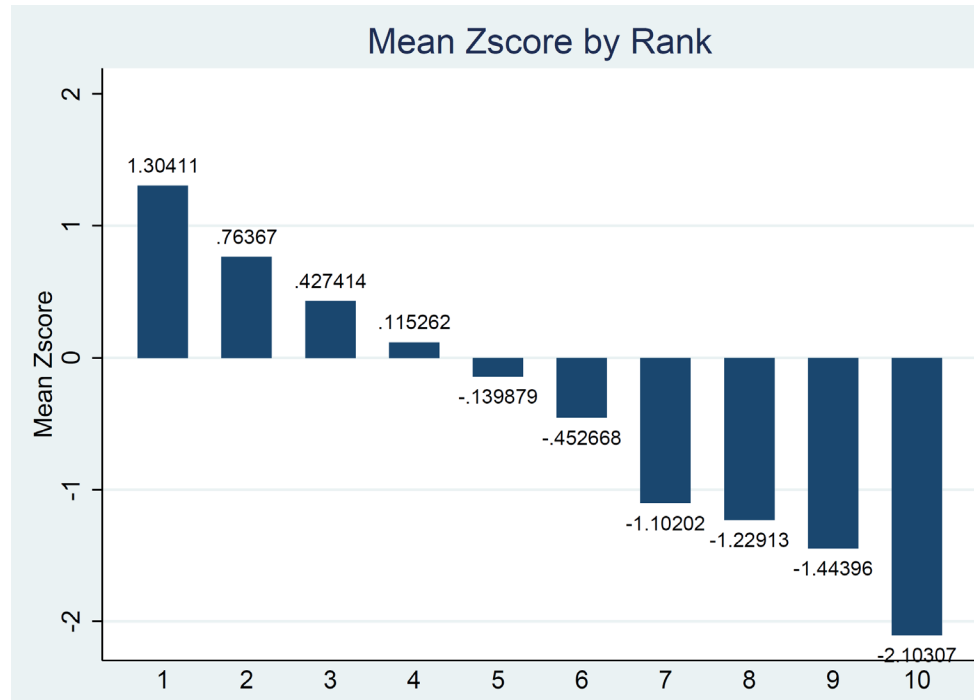
Startup controls = number of employees, funds raised.

Appendix D: Baseline Control Group Scores by Founder Gender (Panel)

	Control Group Baseline (Round 0)				Control Group Program (Balanced, Rounds 1-3)			
	Prevention Question (Binary) Logit odds ratio		DV Zscore (continuous) OLS		Prevention Question (Binary) Logit odds ratio		DV Zscore (continuous) OLS	
Female founder	1.301 (0.249)	1.13 (0.284)	-0.246* (0.109)	-0.341* (0.125)	1.017 (0.215)	1.195 (0.262)	-0.123 (0.116)	-0.082 (0.144)
Clustered errors (Investor)	x	x	x	x	x	x	x	x
FE region	x	x	x	x	x	x	x	x
FE round					x	x	x	x
Startup controls		x		x		x		x
N	412	323	412	323	813	723	813	723
Pseudo R-sq / R-sq	0.0293	0.0264	0.0151	0.0882	0.1539	0.1641	0.0054	0.0482
Investors	41	41	41	41	34	34	34	34

Startup controls = number of employees, log of funds raised

Appendix E: Mean Zscore by Rank



Each startup received a zscore per round. The inputs are the average score and the standard deviation per judge per round. Then for each judge's score for each startup Vilcap creates a zscore = (score - avg_score)/sd_score. Vilcap then joins the zscores together by taking an average across all judges. The highest zscore becomes rank 1, followed by rank 2, and so on. The lowest zscore is ranked 10 of 10 startups. Only startups ranked 1 and 2 received \$20,000 investment.

Appendix F: Effect of Systematizing Inquiry by Prompting Risk and Reward (Panel)

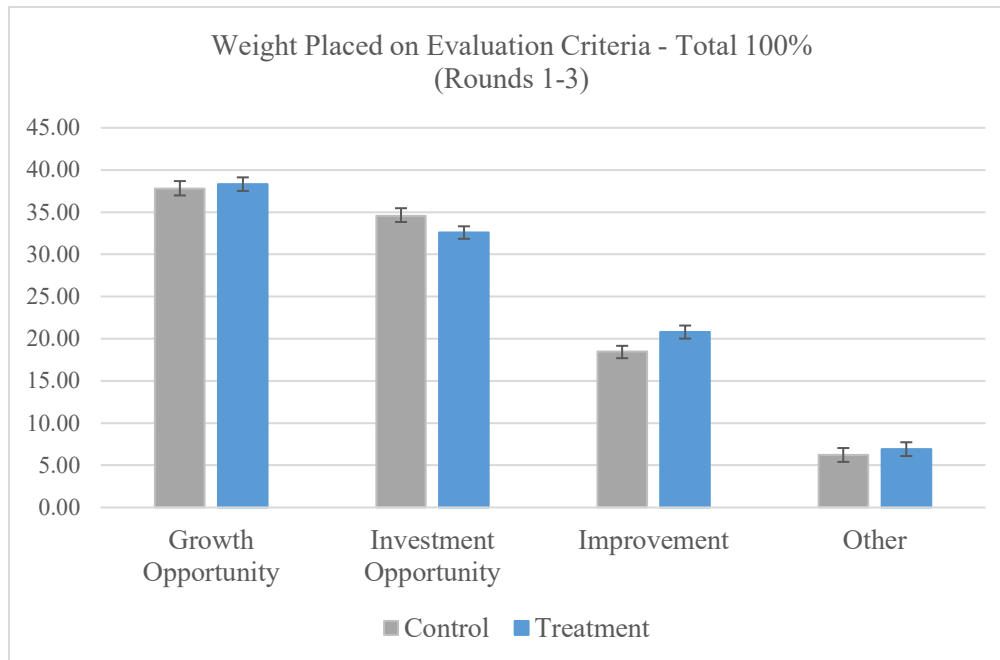
	Prevention Questions (Binary) Logit Odds Ratio (Rounds 1-3)			DV Zscore Ancova (Rounds 1-3)	
	1	2	3	4	5**
Female founder	1.019 (0.123)	1.004 (0.217)	1.218 (0.288)	-0.059 (0.115)	-0.055 (0.115)
Systematized Inquiry		4.432*** (1.154)	4.769*** (0.409)	-0.157* (0.077)	-0.146 (0.080)
FF*Inquiry		0.788 (0.207)	0.633 (0.179)	0.306* (0.151)	0.302* (0.151)
Baseline score	0.981 (0.065)	0.991 (0.067)	1.006 (0.075)	0.175*** (0.040)	0.175*** (0.040)
<i># prevention questions</i>					<i>-0.041</i> (0.060)
Systematized Inquiry + FF*Inquiry = 0		1.251*** (0.316)	1.105** (0.340)	0.148 (0.088)	0.156 (0.089)
FF + FF*Inquiry = 0		-0.234 (0.143)	-0.260 (0.075)	0.248* (0.099)	0.246* (0.099)
Clustered errors (Investor)	x	x	x	x	x
FE region	x	x	x	x	x
FE round	x	x	x	x	x
Startup controls			x	x	x
N	1,530	1,530	1,395	1,395	1,395
Pseudo R-sq / R-sq	0.1537	0.2117	0.2155	0.0751	0.0754
Investors	65	65	65	65	65

Startup controls = number of employees, log of funds raised

Model 3 was pre-registered, and all other models provide similar directional results

**Model 5 is mediation analysis - shows little evidence for moderation

Appendix G: Effect of Systematizing Inquiry on Evaluation Criteria Considered by Investors



Appendix H: Female Variable Robustness Check (Panel)

	ANCOVA - DV Zscore (Rounds 1-3)		
	1	2	3
Female Application	0.124 (0.066)	-0.017 (0.092)	-0.033 (0.101)
Systematizing Inquiry		-0.136 (0.070)	-0.183*** (0.074)
FF*Inquiry		0.296* (0.126)	0.340* (0.135)
Baseline score	0.235*** (0.064)	0.239*** (0.037)	0.184*** (0.040)
Systematizing Inquiry + FF*Inquiry = 0		0.160* (0.061)	0.156* (0.074)
FF + FF*Inquiry = 0		0.280** (0.086)	0.307** (0.089)
Clustered errors (Investor)	x	x	x
FE region	x	x	x
FE round	x	x	x
Startup controls			x
N	1,530	1,530	1,395
R-sq	0.0522	0.0579	0.0791
Investors	65	65	65

Appendix I: Startup Characteristics by Founder Gender

	DV - Proxies for Startup Stage / Quality			
	Total Employees	Funds Raised	Funds Raised (log)	VilCap Score
Female founder	-2.849 (2.507)	\$47,506 (\$85,043)	0.366 (1.137)	0.003 (0.099)
FE region and round	x	x	x	x
N	67	64	64	68
R-Sq	0.2050	0.1278	0.0763	0.0408