

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

DRAFT – July 28, 2023

Amisha Miller, Saurabh A. Lall, Markus Goldstein and Joao Montalvao

Abstract. Female innovators raise less resources from investors than male innovators, even when their ventures are similar or identical. In uncertain contexts, evaluators systematically value women's competence or leadership potential as lower than men's, and investors are more likely to inquire about risks when facing female founders than males. However, efforts to mitigate disparities have typically focused on changing how individual founders seek investment. Without examining investors' evaluation processes and the practices used by the organizations in which they are embedded, we cannot fully explain how systemic gender disparities in investment outcomes are produced, nor how they might be reduced. *What is the effect of investment organizations' evaluation practices on gender disparities in funding innovation?* We ran a two-stage global field experiment with investors making 1,871 investment decisions on early-stage startups that resulted in \$320,000 invested in 16 startups. We aimed to systematize investor inquiry across all ventures by changing the organization's evaluation framework to include prompts to assess (1) risk and reward and (2) progress during the evaluation period. Treated investors assessed startups more consistently and assessed start-up competence more dynamically than control investors. Our interventions eliminated, and even reversed, the gender gap in investment outcomes. We demonstrate the causal effect of organizations' evaluation frameworks on gender disparities in investment and identify a novel approach to tackle disparities. More broadly, we theorize a link between micro-level processes of inquiry and evaluation outcomes, with implications for organizations evaluating innovation in uncertain contexts and those aiming to reduce gender disparities.

Word Count: 248

We thank our implementation partner Village Capital. We also thank the Umbrella Facility for Gender Equality, the We-Fi initiative, the International Finance Corporation, Aspen Network of Development Entrepreneurs, and Small and Growing Business group for funding. Finally, we thank David McKenzie, Christopher Woodruff, the NESTA Innovation Growth Lab, the Boston University Strategy and Innovation group, and the University of Oregon Institute for Policy Research & Engagement for helpful comments on early research designs and drafts.

1. Introduction

Women are underrepresented in leadership positions in innovation and entrepreneurship; among the population of funded ventures, less than 12% 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 to those of all-male teams (e.g., Brooks et al. 2014, Roberts and Lall 2018, Guzman and Kacperczyk 2019, Ewens and Townsend 2020). This situation impedes the ability of female-founded ventures to grow (e.g., 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 misallocation of resources within economies (e.g., Hsieh et al. 2019).

Gender disparities in economic outcomes are often the result of choices made by individuals or organizations that together, create and perpetuate a system of inequity (e.g., Fernandez-Mateo and Kaplan 2018). For example, in male-dominated networks, such as those in start-up investment, male evaluators typically socialize with, hire, and invest in people who share their gender (e.g., Ibarra 1993, Gompers and Wang 2017, Greenberg and Mollick 2017, Howell and Nanda 2019, Bapna and Ganco 2021). Beyond homophily, in uncertain contexts, all evaluators rely on easily accessible indicators of status, including gender (e.g., Ridgeway and Correll 2004, Botelho and Abraham 2017), and typically undervalue women's competence or their leadership potential (e.g., Benson et al. 2022, Snellman and Solal 2023). These gendered differences are also embedded in investors' behaviors when evaluating a startup. In the absence of information about prior organizational performance (Aldrich and Fiol 1994, Cohen et al. 2019a), evaluating a startup's potential involves making decisions based on partial knowledge, rather than complete information (Knight 1921, Rindova and Courtney 2020). To form an opinion about the potential of the team and their ideas, investors interact with founders (Kirsch et al. 2009, Petty and Gruber 2011, Miller et al. 2023). They pattern-match innovators' behaviors –

typically men's – to previous successful investments (e.g., Elsbach and Kramer 2003, Huang 2018) which can reinforce investors' preference for the status quo. Investors assess risks and rewards differently: pressing female founders about risk while focusing on rewards for male founders (Kanze et al. 2018). Given these behavioral patterns, investors' processes of inquiry – how evaluators assess the potential of innovators and their ideas during interactions – can disadvantage female innovators.

Scholars have examined the causes of gender disparities in investment outcomes more than their mitigation (e.g., Jennings and Brush 2013). An exception is research on investor–founder interactions that examines the effects of female founders' pitches on investors' decisions (e.g., Kanze et al. 2018, Lee and Huang 2018, Balachandra et al. 2019, Huang et al. 2021) but holds investors' evaluation processes constant to study this question. However, a focus on pitching tactics puts the onus for change on the founder and may not have the same effect for every investor (Pahnke et al. 2015, Clough et al. 2019). Therefore, we focus on investors, who are often embedded in organizations that invest collective funds as part of a designed evaluation process (Tyebjee and Bruno 1984, Drover et al. 2017, Lerner and Nanda 2020). Investment organizations and their funders have invested \$4.8 billion into diversity strategies since 2018 (Cortes 2019, Biegel et al. 2020, International Development Finance Corporation [DFC] 2021), but the strategies they employ and their effects on investment outcomes are understudied. Without examining organizational evaluation practices, which could include agreeing on referents, negotiating criteria, and establishing value by comparing entities (Lamont 2012) – we cannot fully understand how disparities in investment outcomes are produced, nor how they can 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 frameworks, investors' processes of inquiry, and investment decisions by founder gender. In investment organizations, evaluation of startups typically unfolds over three months (e.g., Tyebjee and Bruno

1984, Cohen et al. 2019b, Gompers et al. 2020). Investors receive a pitch deck or overview, meet founders, and then make a decision on whether to conduct due diligence on (and inquire more about) the start-up. This process repeats through deeper stages of diligence before investment organizations decide to invest in a start-up. Examining this process at scale is difficult because investment organizations often do not share their evaluation processes and outcomes publicly (Da Rin et al. 2013). To do so, 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 introduces startups to a broader set of potential follow-on investors. We designed a two-part intervention to reduce gender disparities in Vilcap’s investment outcomes. We changed their evaluation frameworks to systematize inquiry and examined the effects on investors’ processes of assessment and investment decisions, analyzing more than 31,000 startup evaluation scores during the study period.

In Stage 1 of the field experiment, we used a cross-sectional design to examine investors’ decisions to conduct further diligence on a startup or to exclude it from further consideration. We randomized 278 investors (both Vilcap and potential follow-on investors) into a treatment group, which Vilcap prompted to systematically inquire about risk and reward, and a control group, in which investors evaluated startups as normal – using status quo processes. We analyzed 1,341 evaluator–start-up investment decisions on whether to continue diligence for 87 startups. We found that during interactions, control investors were more likely to assess risk for startups with a female founder while treated investors assessed startups more consistently. This intervention significantly reduced gender disparities in investment decisions – in this Stage, entry into due diligence.

In Stage 2, we leveraged a unique facet of Vilcap’s evaluation process – it trained local investors to invest \$320,000 in 16 startups over three months. We built on recent research suggesting that when evaluators focus on demonstrable achievements over time, rather than assessing potential, gender

disparities are reduced (Benson et al. 2022). In an investment context in which dynamic, backward-looking information is scarce, we prompted Vilcap investors to systematically inquire whether startup had made *progress* on their growth and risk-mitigation strategies during the three-month evaluation period. We examined the effect on Vilcap's actual investment decisions over time using a panel dataset of 1,530 decisions. We found that treated investors assessed startups using dynamic evidence of demonstrated progress, rather than static data, which served as a signal of potential by control investors. This was followed by significant differences in their decisions to invest in a startup. These small but significant changes to Vilcap's evaluation framework produced changes in investors' assessment, which spurred differences in evaluation that eliminated, and even reversed, the gender gap in investment decisions.

A growing body of literature in entrepreneurship considers how to level the playing field for female entrepreneurs and has typically focused on how to prepare startups for investors' evaluation. This provides insight into individual investor biases and how to circumvent them. By contrast, we theorize evaluation not as a dyadic process between an investor and a founder but as part of a collective evaluation process, designed by investment organizations, in which investors are nested. By focusing on evaluation processes, we identify a novel means to tackle systemic disparities in investments in a systematic manner: by prompting consistent risk assessment and dynamic progress assessment. We provide insight into how organizations' evaluation frameworks produce gender disparities in evaluating innovator's potential. We also theorize the importance of processes of inquiry for evaluation in contexts where evaluators lack complete information; take action based on only partial knowledge, and where evaluator discretion is valued. We contribute to a growing literature on how organizations make decisions under uncertainty, by examining the inquiry processes that investors use to make decisions and explaining the effect of these processes on gender disparities. We extend prior research on mitigating demographic disparities through organizational practices, which has typically

focused on adding structure to evaluation processes in hiring (see a review in Stephens et al. 2020). We explain the promise of systematizing inquiry, making small changes to evaluation frameworks that preserve individual evaluators' discretion to ask questions but that affect their subsequent processes of assessment and evaluation. Doing so could inform policy for a broader set of organizations interested in reducing demographic disparities.

2. The Role of Organizations in Evaluation

Gender disparities replicate through social interactions (e.g., Ridgeway and Correll 2000, Ridgeway 2014) and are pervasive in evaluation processes and outcomes in organizations (for a review, see Fernandez-Mateo and Kaplan 2018). These disparities are particularly prevalent in contexts in which quality is uncertain, such as funding innovation. In these contexts, individual 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 or more competent than women, this (often unconscious) reliance on gender can produce lower evaluations for women (Correll and Benard 2006, Botelho and Abraham 2017, Snellman and Solal 2023). In addition, early evaluators often consider the preferences of others, which pushes them to make more conventional choices, as encoded in status beliefs (Correll et al. 2017).

However, organizations can design their evaluation frameworks to influence how individual evaluators behave, which could affect outcomes by gender. Organizations design templates and tools that shape how decisions are made (e.g., March and Simon 1958), and these evaluation practices can shape evaluation outcomes (e.g., Lamont 2012, Zuckerman 2012). In other innovation contexts, organizations can specify how data that inform decision-making is shared or analyzed – in letters, PowerPoint decks, or excel spreadsheets (e.g., Fayard and Metiu 2014, Anthony 2021), which can affect how decisions are made and shape the content of decisions. For example, using PowerPoint documents enabled an organization to create spaces for discussion and idea evolution, which shaped

how strategy was made (Kaplan 2011). Organizations' evaluation frameworks may even affect whether decisions are gendered, but this causal link is understudied.

An adjacent organizational evaluation process, similar to investment decisions, is hiring and promotion decisions, in which evaluators make decisions about people under conditions of uncertainty, in a short time frame, with ramifications for organizational funds and reputational outcomes over time. Organizations have made 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. 2011, Stephens et al. 2020). For example, affirmative action policies can lead to unintended consequences if they unintentionally increase the saliency of stereotypes that target groups lack competence. This can decrease target groups' performance and increase disparities (Leibbrandt et al. 2018, Leslie 2019).

Focusing on evaluation processes has proved more successful (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 start-up strategies are subject to change (e.g., Siggelkow 2002, Kirtley and O'Mahony 2023) 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 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, 2006; Hallen 2008). Given these difficulties in assessing static information, investors deem interacting with founders a fundamental part of evaluation (Petty and Gruber 2011). Investors pride themselves on seeking information beyond the business plan and using their "gut feeling" to source and evaluate investment opportunities (Kirsch et al. 2009, Huang and Pearce 2015, Huang 2018). Investors typically evaluate through processes of inquiry, assessing the potential of an innovator and their idea during interactions.

3. The Role of Inquiry in Evaluation

During processes of inquiry, investors ask questions not only 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). Processes of inquiry are also part of evaluation in many hiring contexts. 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 structure interactions such that evaluators ask the same open-ended questions to all applicants (Huffcutt 2011) or use task-based interviews in which applicants complete a task or set of tasks similar to those required in the job (e.g., Ployhart et al. 2005). However, not all task-based interviews reduce disparities, and causal evidence of the efficacy of structured interviewing is lacking (Stephens et al. 2020). This lack may be because the type of inquiry is important. For example, asking management consulting applicants to evaluate a case on a male-dominated industry is biased against women, who have less background knowledge about the industry (Rivera 2015). Given that inquiry is a crucial part of evaluation in innovation contexts and can produce gender disparities in other contexts, we focus on identifying differences in investors' processes of inquiry by

founder gender. We then theorize how investment organizations might mitigate these differences to affect gender disparities in investment outcomes.

3.1. Inquiry about Risk and Reward

During interactions, investors often spend more time assessing risk for startups with female founders than those with male founders. Docsend, a platform that allows founders to share pitch decks with investors, indicates that investors spend more time assessing traction and product slides for startups with female founders (to judge their current assets), while they spend more time on fundraising request slides (to assess what founders might do in the future) for all-male teams startups (Frost 2020). Similarly, in pitch competitions, investors typically ask women prevention-focused questions related to maintaining nonlosses and avoiding a worse future state e.g., “How many monthly active users do you have?” while they ask men promotion-focused questions to understand rewards or growth e.g., “How do you plan to acquire customers?” Investors’ processes of inquiry produce conversations that differ by founder gender and could cause them to evaluate ventures with female founders as less valuable (Kanze et al. 2017, 2018). We hypothesize that systematically prompting investors to inquire about risk and reward could prompt them to more consistently assess both promotion and prevention across startups, reducing gender disparities in evaluation.

3.2. Inquiry about Progress

Investment organizations provide evaluation frameworks to investors to assess the growth potential of early-stage startups in the absence of a history of 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 by current job performance, it would have reduced disparities (Benson et al. 2022). Whereas performance ratings are backward-looking and based on demonstrable competence, 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 assess demonstrable competence is complicated, given the lack of a history of organizational performance data for startups (e.g., Stinchcombe 1965, Aldrich and Fiol 1994, Cohen et al. 2019a). However, in other contexts, some organizations assess short-term signals of performance during hiring processes. For example, Goldin and Rouse’s (2000) classic study demonstrates that when orchestra-hiring managers evaluated candidates’ performance through blind auditions, they hired more female performers. This suggests that investors may be able to assess competence by evaluating performance during the selection process. Some investors already do so. For example, Vilcap program managers shared that progress made during the three-month program is important to its decisions: “[We] invest in people that make the most progress during the program.”. Similarly, 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.”

Thus, some individual investors value dynamic evidence of progress as a signal of competence when assessing startups. However, most investment organizations design evaluation frameworks to assess static elements of a start-up, including team and venture characteristics (e.g., Tyejee 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’ assessment on dynamic evidence of progress during the selection process, and could reduce gender disparities in investment decisions. Together, we created one hypothesis:

Hypothesis 1. *Investment organizations that systematize inquiry by prompting investors to inquire about risk, growth, and progress will reduce gender disparities in investments.*

3.3. How Organizations Affect Inquiry

Our main hypothesis suggests that organizations can create systems-level change in evaluation by changing prompts in evaluation frameworks to reduce gender disparities in investors' decisions, which feed into collective investment outcomes. However, this hypothesis assumes a mechanism – that changing organizations' evaluation frameworks will affect individual investors' assessment during processes of inquiry, which in turn will affect their evaluation of startups. This assumption may not hold, as investors pride themselves on using their intuition or gut feeling to evaluate investment opportunities and do not always follow evaluation frameworks (Kirsch et al. 2009, Huang 2018). We test the first part of the mechanism with the following:

Hypothesis 2. *Prompting investors to inquire about risk and reward will increase the consistency of their assessment across startups.*

Research in adjacent contexts shows that if organizations change the content of assessment, it can backfire (Leibbrandt et al. 2018, Leslie 2019). For example, when organizations positioned their hiring and promotion practices as meritocratic, hiring managers were even more likely to favor male employees over equally qualified female employees 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, 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 frameworks to include new criteria and create transparency around evaluation practices. We test whether organizations can affect investors' assessment of dynamic evidence of progress during the inquiry processes with the following:

Hypothesis 3. *Prompting investors to inquire about progress will increase their dynamic assessment of competence.*

4. Research Approach

4.1. 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 design a two-stage field experiment to test whether systematizing inquiry affects individual investors' assessment processes and investment outcomes. By using a field experimental setting, we demonstrate how effective interventions are under real-world conditions, overcoming concerns about generalizing from experiments in laboratory settings with students or online surveys (Hsu et al. 2017, Czibor et al. 2019). This is particularly important in an investment setting, because trained investors often evaluate startups differently than an average individual (e.g., Kirsch et al. 2009, Clingingsmith and Shane 2018).

These experiments were possible given our 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 an appropriate field site for this intervention as it provides access to two types of investor evaluation. It uses professional investor evaluations to facilitate introductions between startups and investors, and it trains local investors to invest Vilcap funds. Since 2009, Vilcap has used explicit evaluation frameworks (templates) to ensure effective communication among professional investors, Vilcap, and trainees, and it 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% of its portfolio, and it wanted to increase the number of startups with female founders in its portfolio of over 100 startups.

In addition, Vilcap provided access to a similarly qualified set of startups. Vilcap used a competitive process to identify startups with high growth potential to enter its program, with between

200 and 400 applicants for 10 places. All startups had a product, were aiming to improve their product-market fit, and were seeking investment. Each investment program recruited for one industry problem statement, so startups within each program were working in a common industry (but were not direct competitors). Vilcap had deemed all startups of high quality and at similar stages. Baseline studies suggest that all startups accepted into the Vilcap program had similar observable characteristics, which did not differ by founder gender (Burns et al. 2019). Vilcap also provided access to a curated set of investors who had expressed interest in startups at an early-stage and in a specific industry. Thus, any differences in results should not be driven by differences in start-up quality nor investor interest.

Vilcap's evaluation process is typical of the average investment organization. Investment organizations employ a collective evaluation process to decide whether to invest their funds into startups (Tyebjee and Bruno 1984, Fried and Hisrich 1994), which typically takes approximately 90 days (Gompers et al. 2020). 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 the simple model in Figure 1 shows, 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 conduct diligence on the start-up (and inquire further). This process repeats, as investors advance through deeper stages of diligence before they invest in a startup.

In Vilcap, this process unfolded with two types of investors. Professional investors,¹ embedded in a range of investment organizations, met startups once. They received a venture overview written by Vilcap, met approximately three founders for 30 minutes, and evaluated startups using a Vilcap survey to decide whether they wished to conduct due diligence by receiving additional information from the start-up. If investors wished to continue due diligence on a start-up, Vilcap

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

facilitated an introduction. In Stage 1 of the field experiment, we tested the effect of prompting this diverse set of investors to systematically inquire about risk and reward and examined the effect on the processes of investor assessment and the likelihood of continued diligence. This models the beginning of the selection process, and this type of cross-sectional design is common in research attempting to unpack demographic disparities in investment evaluation in the field (e.g., Younkin and Kuppuswamy 2018, Ewens and Townsend 2020).

Vilcap also trained local investors to evaluate startups for Vilcap and to allocate their own investment capital.² In Vilcap, trainee investors typically received 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 meeting startups. As part of Vilcap’s normal training program, trainee investors then continued to evaluate startups three more times, using standardized criteria.

In Stage 2, 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. Use of 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). Because Vilcap trained investors from the region and market to evaluate startups on its behalf, it required investors to provide scores on specific elements of the venture, including team, problem and vision, product, market, and business model –

² Vilcap trains founders who qualify for its program to be investors and to allocate Vilcap funds. Its 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 start-ups, 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 examine the validity of this assumption in the first experiment, where we leverage a pooled sample of trainee and professional investors.

typical criteria used by other investors. Trainee investors evaluated over multiple periods and were required to explain their reasons for scoring and to provide transparent feedback to startups. This provided a unique setting not only to experiment with organization-level evaluation frameworks but also to observe how evaluation was conducted over time.

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, which resulted in a dataset of 31,714 evaluation scores. Stage 1 leveraged the cross-section sample of 1,341 dyadic investor–start-up decisions by investors made after the investor met a founder. This sample included both professional investors and trainee investors who Vilcap trained to allocate \$320,000 to 16 of 87 startups (see Table 1). Stage 2 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–start-up dyads over three periods after the initial analysis we observed in Stage 1). We preregistered both stages of the field experiment.

[Insert Table 1 about here]

4.2 Stage 1: Systematizing Inquiry about Risk and Reward

4.2.1 Setting and Design. We systematized how Vilcap prompted investors to inquire about risk and reward in their evaluation templates and assessed the impact on reducing gender disparities in continuing due diligence (Hypothesis 1) and the consistency of investor assessment across startups (Hypothesis 2). Trainee investors met startups in a 90-minute welcome meeting during which each start-up founder was encouraged to share a little about themselves and their startups. Professional investors met startups in 20- to 30-minute sessions in which startups shared an overview of their businesses and then sought advice from the investor: 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 start-up that outlined team members, market, product, and the funds the start-up wanted to raise. We randomized professional investors into a treatment or control condition after they met the startups and began evaluating 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, startup-treated investors received a slightly different evaluation form than the control group (see Appendix A).

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

4.2.3. 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?”

4.2.4. Empirical Design. We ran the following preregistered regression:

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

where the unit of analysis is the investor decision per start-up; the dependent variable Y_{is} is the propensity to invest in start-up s by investor i , measured with a 6-point scale for professional investors; F is a binary variable that equals 1 if a female founder represented the start-up and 0 otherwise; and T is a binary variable that equals 1 if investors were prompted to inquire about risk and reward

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

⁴ The most variance appeared in this part of the variable in exploratory studies/pretests, which we ran on different investor populations before the experiment.

systematically (treated) and 0 otherwise. The coefficient of interest β_3 is the interaction of a risk/reward inquiry treatment with a female founder. We included fixed effects for the region R . Although we randomized startups into treatment and control groups, given the relatively small number of startups we assessed (87), start-up characteristics could affect the size of the estimates. Therefore, we also controlled for observable start-up characteristics X_s , or the number of employees and the log of funds raised at selection into Vilcap's program.⁵ We ran an ordinary least squares (OLS) regression, which was preregistered, but added an ordered logit because the dependent variable was ordinal. We report the ordered logit, as the results were the same in both models. We clustered errors in all models by investor, or the level at which the treatment was implemented.

4.2.5. Mechanism. We examined whether prompting investors to systematize inquiry about risk and reward increased the consistency of their assessment. To examine consistency of assessment, we measured the prevention–promotion focus questions investors asked across founders. Here, Y equals 1 if a prevention question was asked to a start-up and 0 otherwise. Two research assistants (one for Spanish, one for English) coded all investor questions posed to founders with a prevention or promotion focus, following Kanze et al. (2018).⁶ Any disagreements were discussed in a group with the first author so that codes were applied consistently. The first author, who did not have access to the start-up's founder gender, made the final decision on whether a question was coded as promotion or prevention. We constructed a binary measure at the investor-dyad level to measure the incidence of a prevention focus in assessment and used the same type of binary incidence measure for a promotion focus. Simply put, if an investor question to a start-up did not have a promotion or prevention focus, both variables would equal 0. If an investor question to a start-up included both a

⁵ Vilcap also collected funds raised by start-ups, but we did not include these as a control, as many start-ups had not yet raised funds, leading to 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 et al. (2018) but included additional variables to better suit the setting – a series of dyadic entrepreneur-investor interactions, rather than a one-off pitch in which investors asked questions to entrepreneurs in a group setting.

promotion focus and a prevention focus, both variables would equal 1.⁷ We focused our analysis on investors' use of prevention questions and used this measure to examine whether their assessment was consistent across startups.

4.2.6. 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 the 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 the two groups.

[Insert Table 2 about here]

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 start-up controls, investors in the control group gave startups with female founders significantly lower scores than startups with all-male teams (0.6 on a 6-point scale – equivalent to 10 percentage points), as Appendix B shows. This disparity held across trainee investors and professional investors, as well as male and female investors. This difference in scores was correlated with the likelihood of investors to focus more on prevention when inquiring from startups with female founders than startups with all-male teams. As Appendix B shows, investors in the control group were 15% more likely to ask a prevention-focused question to

⁷ Following Kanze et al. (2018), we used a computerized method as a robustness check for all responses in English, to verify the direction of results from the research assistants' qualitative coding. We used 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 software to determine their frequencies. Similar to Kanze et al. (2018), we find the same directional results.

⁸ A small minority of professional investors met start-ups in multiple programs (i.e., in the Middle East and Africa). As we randomized investors 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.

a start-up with a female founder than a start-up with an all-male team. This difference was directional but not significant. Together, this result 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 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 (Hypothesis 1) and whether it could change investor assessment (Hypothesis 2). 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 start-up, by founder gender. While control investors scored female founders significantly lower than startups with all-male teams, treated investors did not. This result provides support for Hypothesis 1, 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 assess prevention for startups with female founders and asked prevention-focused questions to startups with female founders more than those with all-male teams. However, treated investors were more likely to inquire consistently across startups. This result provides suggestive evidence in support of Hypothesis 2. When investment organizations systematized inquiry by prompting investors to inquire about risk and reward, investors assessed prevention and promotion more consistently. This effect is driven by treated investors being significantly more likely to ask prevention-focused questions to *all* startups than the control group and, directionally, even more for those with all-male teams.

[Insert Figure 2 about here]

As Table 3 shows, regression analysis including start-up controls showed similar results. Control investors were less likely to take a start-up with female founders through due diligence than

all-male teams, while treated investors were equally likely to take a start-up through due diligence, regardless of founder gender. Control investors were only 65% likely to increase their score by one unit (i.e., from agree to strongly agree to take the start-up through due diligence) for startups with female founders compared with all-male teams. However, treated investors were equally likely to do so (0.65 main effect multiplied by the 1.63 interaction).⁹ This result provides additional evidence in support of Hypothesis 1. 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.

[Insert Table 3 about here]

These results provide evidence that simply prompting an investor to ask founders about the “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 start-up that used mobile technology to connect handymen with work opportunities received different questions from control and treated investors. A control investor asked: “[Can I see a] marketing plan clearly highlighting the marketing strategies?” By contrast, a treated investor asked the same founder: “[How will the company] manage delayed payments [... 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 with the statement “I will conduct due diligence” on whether the start-up had a female founder or all-male team. A simple change to an evaluation template affected investors’

⁹ As in 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, such as the weather, to add noise (e.g., Dushnitsky and Sarkar 2022).

use of prevention framing and whether they wanted to take the start-up through due diligence. Together, these results lend support to our hypotheses. When Vilcap systematized inquiry by prompting investors to inquire about risk and reward in its evaluation templates, investors assessed prevention and promotion consistently across all founders, regardless of founder gender, and reduced gender disparities in investment decisions and outcomes.

4.2.7. 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 Linguistic Inquiry and Word Count), and a female binary variable (presence in venture overview; see Appendix C). The results held for heterogeneous investor types (male and female, 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 in the relationship of treatments with the score.¹⁰ This suggests that the results would hold for a diverse range of investors.

4.2.8. Investment Outcomes. This experiment revealed that changing an organization’s evaluation framework to prompt investors to inquire about risk and reward systematically influenced their assessment during processes of inquiry and reduced gender disparities in evaluation. This effect was also meaningful. Investors who 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 than the control group, 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.

¹⁰ Heterogeneity analyses are available on request.

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

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

4.3. Stage 2: The Effect of Systematizing Inquiry on Risk, Reward, and Progress

4.3.1. Setting and Design. Vilcap trained 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 examine whether cross-section results applied over time (e.g., Bohren et al. 2019). We changed Vilcap’s evaluation framework by systematizing how it 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 examined the impact on reducing gender disparities in investment decisions (Hypothesis 1), the consistency of investor assessment across startups (Hypothesis 2), and whether investors assessed competence dynamically (Hypothesis 3).

4.3.2. Dependent Variable. After the baseline score, given in Stage 1, three more times over the course of its 90-day program, Vilcap asked investors to complete due diligence, and rank startups. Vilcap’s normal set of evaluation questions were focused on assessing potential: (“What is the company’s growth opportunity, and what is the company’s investment opportunity?”) across

categories (e.g., team, value proposition, market, scale).¹¹ Investors used a 4-point scale per category, resulting in a 24-point scale overall (from 8 to 32). The final evaluation scores resulted in investments in the two most highly scored startups.

4.3.3. 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 framework: (1) the same treatment that prompted investors to systematically inquire about risk/reward in Stage 1 and (2) Vilcap’s prompt to inquire about start-up progress and 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 ...”: (1) “Understanding its path to growth?” (2) “Executing its path to growth?” (3) “Understanding its risks?” and (4) “Executing on risk mitigation?”

4.3.4. 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:

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

where our unit of analysis is the investor decision per round; the dependent variable Y_{ist} is the propensity to invest in start-up s by an investor i at time t ; t is the stage of measurement, that is, a scale variable collected over three periods; and Y_{is0} is the baseline measure of evaluation of a startups by an investor, also evaluated on a scale. We normalize all scale evaluations using a z-score.¹² In addition, F

¹¹ Vilcap is unwilling to openly share its proprietary evaluation templates.

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

is a binary variable that equals 1 if a female founder represented the start-up and 0 if only a male founder represented the start-up, and T is a binary variable that equals 1 if investors were in the changed evaluation framework treatment group and 0 if not. The coefficient of interest β_4 is the interaction of treatment with female founder. We included fixed effects for the region R and time t and controlled for the same observable start-up characteristics X_i as in Stage 1. We clustered errors in all models by investor, or the level at which the treatment was implemented.

4.3.5. Mechanism. We added mechanism variables to test whether systematizing inquiry at the organization level changed how investors assessed startups. First, we examined whether the intervention increased the consistency of investors assessment across startups, using the same mechanism as in Stage 1. Second, we examined whether the intervention increased investors' assessment of dynamic evidence of progress 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 percentages of how much weight you placed on each criterion. (Please make sure it adds up to 100%)"': growth opportunity, investment opportunity, and improvement made during program. Given that many investors do not accurately explain the criteria that are important to them (Petty and Gruber 2011), we followed recent field experiment research and conducted semistructured interviews to glean more insight into mechanisms driving results (e.g., Dimitriadis and Koning 2022). The first author conducted 45-minute semistructured interviews with investors in one region to determine how investors undertook evaluation and came to give startups a high or low score.

To further evaluate the effect of systematizing inquiry on individual investors' assessment, we compared how investors in the treatment and control groups 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 reevaluated startups from their first impressions at baseline.

4.3.6. Results. The balanced panel dataset after attrition was 1,530 investment decisions made by 65 investors on 69 startups over three periods (510 decisions per period). Attrition was largely due to COVID-19-related absences. Table 4 shows no significant difference in observable characteristics between the treatment and control groups using raw numbers, percentages, or the p-value taken when regressing each characteristic on treatment.

[Insert Table 4 about here]

To examine the baseline, we used trainee investors in the control group. Similar to Stage 1, investors asked startups represented by women more prevention-focused questions and awarded lower scores than startups represented by only men. However, during the program, we observed this effect lessening, finding 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 continuous evaluation of all startups over time or by Vilcap’s programming in which investors were provided with a standardized scoring system and questions. Overall, 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 z-score of 0.12 to 0.07. As Figure 2 illustrates, investors only slightly increased their evaluation of startups represented by women, on average from a z-score of -0.09 to -0.07 . This suggests that the Vilcap program itself acted to reduce gender disparities, largely by reducing z-scores for startups with all-male teams. Even so, startups with female founders still received directionally lower z-scores 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 both men (0.04) and women (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 men lower over the course of the program (from 0.04 to -0.07) and evaluated startups represented by women higher over the course of the program (from 0.01 to 0.11). This effect was driven by changes in how they evaluated startups with female founders. Effects in the treatment group were greater than those in the control group and were driven by increases in scores to startups with female founders.

[Insert Figure 2 about here]

Regression analysis showed similar results (see Table 5). Treated investors using Vilcap's systematizing inquiry evaluation framework scored startups with female founders lower than startups with male founders at the beginning of the program, but not significantly so. Treated investors scored startups with all-male teams lower than control group investors, but the effect on female founders was positive, with an increase of scores by 0.31 in the preregistered model. This effect was significant ($\alpha = 0.05$), providing support for Hypothesis 1.

[Insert Table 5 about here]

Organizational prompts to investors to inquire about the “potential for growth,” “how this venture will mitigate risks,” and “start-up progress” during the selection process influenced how investors evaluated and scored startups with female founders. Systematizing inquiry in these ways caused treated investors to evaluate such startups as higher than those with all-male teams. The bundled treatment not only reduced gender disparities in scores but also reversed them. This suggests that organizations can change evaluation frameworks to reduce gender disparities in evaluation outcomes, even with real investments made over time.

4.3.7. 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 (in 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 determine whether increasing the z-score by 0.31 – the difference in score for a start-up with a female founder in the treatment group versus the control group – affected the likelihood of investment in a start-up. The average z-score for a start-up that was ranked second and received investment was 0.76, compared with the average z-score of 0.43 for a start-up that was ranked third and did not receive investment (Appendix E). The average difference was 0.33, close to the effect of the treatment for female-founded startups. This suggests that the size of the increase in score can change investment outcomes and reduce the gender disparities in Vilcap’s overall portfolio.

4.3.8. Mechanism Analysis. We conducted exploratory analysis to illuminate 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 assessment (Hypothesis 2) and investors’ assessment of dynamic improvement when evaluating startups (Hypothesis 3).

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 male founders, so the mechanism could not work as predicted (see Appendix F). Prompting investors to inquire about risk and reward had similar results to those in Stage 1, in which male founders were asked more prevention-focused questions, but this effect was lower for female founders; however, this did not affect the score. These results provide no support for Hypothesis 2 in this setting, in which investors and founders interacted repeatedly over time. Combining these results with those from Stage 1, we suggest two explanations. First, prompting systematic inquiry of risk and reward may be most valuable to retain startups with 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 investor scrutiny of startups represented by all founders, but particularly for startups with all-male teams.

We next assessed whether investors' dynamic assessment of progress differed by treatment and control groups (Hypothesis 3). As Appendix G shows, 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 not statistically significant ($p = 0.108$) in a sample of 65 investors, providing some directional evidence in support of Hypothesis 3. However, interpreting the importance of this small percentage difference on how investors actually assessed startups is difficult. To provide more insight, the first author conducted 45-minute semistructured interviews with all investors in the treatment and control program in one region, to examine how they engaged in evaluation and came to give startups a high or low score. We found that many investors evaluated static elements of the startup to assess its potential. For example, a male investor explained:

I gave [startup] 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 [and]... their business model too.

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

[Insert Table 6 about here]

Some investors assessed startups dynamically, focusing on progress or improvement when scoring them. A male investor described how he had ranked a start-up 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 [startup] 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 assessed how startups made progress over time. Only two of nine control group investors (22%) explained that they assessed progress. These results indicate that when Vilcap asked investors to evaluate progress and potential, investors assessed startups dynamically. This shifted the focus of evaluation from start-up 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. Together with the finding on changes in the score, this suggests that investors' dynamic assessment of startups positively affected their evaluation of female-founded startups, which reduced gender disparities in investment outcomes, in support of Hypothesis 3.

Both types of systematizing inquiry interventions resulted in significantly more favorable investment decisions for female-founded startups in the treated than control group. These mechanisms worked differently. In Stage 1, prompting investors to inquire about risk and reward increased the consistency of assessment by using a higher prevention focus for all-male teams and scrutinizing them more. By contrast, in Stage 2, prompting investors to inquire about progress increased their dynamic assessment of progress and resulted in rewarding female founders for demonstrated competence.

To further examine how investors changed their scores of startups, we analyzed the effects of systematizing inquiry on assessments of start-up characteristics. As Table 7 shows, systematizing inquiry affected how investors assessed startups that were represented by women. Over time, treated investors assessed startups with female founders significantly higher on growth and investment

opportunity in the “product” and “investor exit” categories and directionally higher on “business model” and “scale”, than control investors. There was no significant difference in scores given to startups with male founders over time.

[Insert Table 7 about here]

Organizational prompts to inquire about risk, reward, and progress during the selection process affected how investors evaluated startups with female founders. This was driven by changes in how they assessed static venture attributes (e.g., product) or potential for exit. In particular, asking investors to assess “since the beginning of the program, how much has this company improved?” in an evaluation framework, significantly affected how investors scored startups. Treated investors were more likely to assess startups dynamically, paying attention to improvement during the selection process. This assessment of demonstrated progress also affected how treated investors assessed the growth and investment potential of startups represented by female founders. In doing so, treated investors evaluated startups with female founders more positively than those with all-male teams.

4.3.9. Alternative Explanations. We ran similar robustness checks to those in 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 (see Appendix H). Progress might be easier for startups, and more salient to investors, when startups entered the evaluation process at a less mature stage; however, we found no statistically significant difference between startups with female founders and those with male founders upon entering the evaluation process (see Appendix I). This result is unsurprising because all startups passed through Vilcap’s common selection process before evaluation. Thus, it is unlikely that the results were driven by differences in the stage at which startups entered the program.

5. Limitations

This research is an early contribution regarding the effects of organization-level evaluation systems on investment decisions and thus has several limitations. As with other research in this area, including studies on crowdfunding or angel platforms (e.g., Younkin and Kuppuswamy 2018, Ewens and Townsend 2020, Bapna and Ganco 2021), the results in Stage 1 are limited to investors' first-stage selection processes. To mitigate this limitation, we tested the effects of treatments in Stage 2 and leveraged a setting in which real investments were made over time. Vilcap provided a unique opportunity to design and test multiple evaluation frameworks and to observe their effects on evaluation practices transparently. However, we do not track the long-term effects of treatments. Following investors over time would subject the results to noise stemming from the fading effect of the intervention (e.g., Ridgeway and Correll 2004), investors' follow-on experiences, and changes in the environment and thus cannot fully mitigate this issue.

Moreover, as Vilcap observed persistent gender disparities, we bundled the treatment to systematize inquiry, to create a strong intervention that would have a meaningful effect on a small sample in a noisy field setting. We found some evidence of both hypothesized interventions but were 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 the consistency of investor inquiry. Despite our inability to separate out the exact mechanisms driving the effects, this limitation is balanced by the benefit of identifying a cost-effective organization-level treatment that affected investment outcomes and which we field-tested with real investors and entrepreneurs.

As with any field experiment, our experiments are limited by their context. We cannot separate out the effects of the treatment by investor type (barring investor gender, which did not change the main results). We tested the treatment in one organization's framework; this organization had created an evaluation framework, mined its own data, identified a disparity in evaluation by gender, and was willing to change the framework to attempt to redress gender disparities. This suggests that our

findings will hold only for organizations that are motivated to identify gender disparities and to redress them and that already have evaluation frameworks in place. However, the effects of Vilcap’s efforts in Stage 1 held with investors from multiple organizations and in different geographic locations, which suggests greater generalizability. Further research in different contexts and with larger samples is necessary to assess the conditions under which our treatments, or other interventions to systematize inquiry, can reduce gender disparities in investments.

Finally, because we tested our treatments with one dependent variable – gender – we cannot assess how these or other evaluation practices might affect other disparities in investment outcomes. This treatment may affect disparities driven by founder race (e.g., Younkin and Kuppuswamy 2018) or other founder or venture characteristics. Future research could examine 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 (Tyejee and Bruno 1984, Cohen et al. 2019b) and individual investor preferences (e.g., Huang 2018), or extraneous conditions that affect investor decisions, such as weather (e.g., Dushnitsky and Sarkar 2022).

6. Discussion

6.1. Effect of Investment Organizations’ Evaluation Practices on Gender Disparities in Funding Innovation

We hypothesized that investment organizations could change their evaluation frameworks to reduce gender disparities in their investment decisions. We designed and tested interventions in a two-stage field experiment, using a global sample of 278 investors making 1,341 investment decisions. In Stage 1, we tested whether changing an evaluation template to prompt investors to inquire about both prevention (risk) and promotion (growth) resulted in more consistent assessment – investors asked more prevention-focused questions to all startups, but particularly those with all-male teams. This

intervention significantly reduced gender disparities in entry into further diligence processes in the treated (vs. control) group. These effects held across investors' characteristics, including their organization and geography.

In Stage 2, 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) start-up progress. Treated investors more positively assessed startups with female founders than control investors, resulting in higher scores that affected the likelihood of investment for startups with female founders. Differential effects between the treatment and control were 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 demonstrably accomplished in a short period, rather than in assessing their potential, benefited startups with female founders. Treated investors judged these startups as having shown competence and thus having future growth potential.

Together, these results indicate the causal effect of organizations' evaluation frameworks on gender disparities in investment. We trace a link among organizations' evaluation frameworks, micro-processes of inquiry and assessment, organizations' decisions, and societal outcomes to make three main contributions to theory: (1) the impact of changing the system on funding for female founders, (2) the role of inquiry processes in decision-making in uncertain contexts, and (3) the promise of systematizing inquiry.

6.1.1. Impact of Changing the System on Funding for Female Founders. A growing body of literature on investment processes in entrepreneurship examines gender disparities. The homophilous networks that investors use to source deals, combined with a lack of female investors, result in less investment for startups with female founders (Saxenian 1990, 1996; Sorenson and Stuart

2001, Gompers and Wang 2017, Greenberg and Mollick 2017, Howell and Nanda 2019, Ewens and Townsend 2020, Hallen et al. 2020). However, simply allocating more female investors may not alleviate disparities (e.g., Bapna and Ganco 2021, Snellman and Solal 2023). Founders can construct narratives or pitches to access resources or combat investors bias (e.g., Lounsbury and Glynn 2001, Martens et al. 2007, Hallen and Eisenhardt 2012, Kanze et al. 2018, Lee and Huang 2018, Balachandra et al. 2019, Huang et al. 2021), but pitching tactics may not have the same effect with every investor (Pahnke et al. 2015, Clough et al. 2019). The mixed results of mitigation strategies designed to reduce gender disparities in investment suggest that our theories on investment processes and demographic disparities may be incomplete.

Rather than focusing on networks or interventions that put the onus on founders, we instead theorized the role of investment organizations, whose practices can affect decision-making in other contexts (Lamont 2012, Zuckerman 2012). Organizations design evaluation tools and templates (March and Simon 1958, Fayard and Metiu 2014, Anthony 2021) that affect the decisions made (Kaplan 2011) and the types of knowledge organizations create (Anthony 2021). Building on this literature, we treated 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. We demonstrated that Vilcap's evaluation framework produced gendered investment outcomes, and we tested the effect of two interventions to change this evaluation framework on mitigating gender disparities. Both interventions resulted in substantive differences in investors' processes of assessment—their consistent risk assessment across startups and dynamic assessment of venture competences. We provide evidence of the effect of organizations' evaluation frameworks on gender disparities in investment outcomes: after prompts to inquire about risk, treated investors gave startups with male founders lower scores; after prompts to inquire about progress, treated investors gave startups with female founders higher venture scores. By focusing on organizations' evaluation processes, we

identified a novel means to tackle systemic disparities in investments in a systematic fashion – by changing evaluation frameworks to systematize processes of inquiry.

Although we examined the effects on how investors assessed startups, we can leverage our results to offer implications for investment organizations that aim to mitigate gender disparities. In this setting, Vilcap allowed us to examine its evaluation practices to identify gender disparities, and to tweak small elements of its evaluation templates to systematize investors' assessment during inquiry processes and to reduce gender disparities in investment decisions. Similarly, we suggest that investment organizations can examine their evaluation practices and outcomes, and tweak their evaluation frameworks to refocus investor assessment and reduce gender disparities in their decisions. We theorize an important role for investment organizations in creating systems of inquiry that can produce or reduce societal disparities in investment outcomes.

More broadly, we demonstrate an effect on gender disparities, but other disparities in decision-making could be similarly affected. We raise a question of how investment organizations' evaluation frameworks, often designed early in their development, might more broadly affect investors' later assessments of startups. We suggest that focusing on the role of the organization by examining the effects organizational frameworks on investors' decision-making would be fruitful for future research that aims to explain the causes of investment outcomes.

6.1.2. Inquiry Processes in Decision-Making under Uncertainty. A growing number of scholars have theorized how strategists in entrepreneurial firms make decisions under conditions of Knightian uncertainty, when actions are based on opinion or partial knowledge rather than complete information (Knight 1921, Gans et al. 2019, Rindova and Courtney 2020). Scholars have theorized how strategists in innovation and entrepreneurship contexts address the incomplete knowledge problem they face through pattern recognition and matching (Elsbach and Kramer 2003, Baron and Ensley 2006, Huang 2018), trial-and-error learning (e.g., Bingham and Eisenhardt 2011), or purposeful

experimentation (e.g., Camuffo et al. 2020, McDonald and Eisenhardt 2020) to inform strategic decision-making. We build on this research by examining investment organizations' strategic decisions to invest their limited funds in a start-up and theorize the importance of processes of inquiry as a means to address the incomplete knowledge problem investment organizations face. We explain that investors use processes of inquiry to assess the potential of an innovator and their idea during interactions with founders and trace a link between micro-level processes of inquiry, assessment, and gender disparities in investment outcomes. Our research demonstrates the importance of processes of inquiry for investment outcomes by founder gender.

We did not leverage a specific cognitive theory to explain why an individual investor might make gendered decisions but instead examined decision-making practices, following other scholars in foregrounding how evaluation is conducted (e.g., Kaplan 2011, Botelho and Abraham 2017). We changed the practices used by investment organizations and investors and now use the results of our study to theorize how processes of inquiry can produce gender disparities in individual investors' decisions. In Stage 1 of the experiment, disparities in investor assessment of 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 driven not by investors 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 Stage 2, disparities in investment outcomes reversed when investors evaluated competence dynamically, driven by both increases in evaluation outcomes for startups with female teams and decreases for all-male teams. This finding is congruent with the previous finding and suggests that startups represented by male founders receive a boost when competence is assessed statically but are less valued when investors pay attention to demonstrated competence.

Our findings suggest that male founders benefit from advantaging mechanisms such as permissiveness (Phillips et al. 2022), which can affect their ability to enter due diligence. This can have implications not only for female founders but also for investment organizations, which may spend more time evaluating startups with all-male teams and less time evaluating startups represented by a more diverse set of founders than is warranted. Perhaps one way to elicit more equitable evaluation would be to inquire more about risk and progress, to prompt male innovators to prove their competence. We join a small but growing number of scholars in explaining how strategists in innovation and entrepreneurship contexts address the incomplete knowledge problem, and in tracking the positive effects on startups with male teams. For example, Cao et al. (2021) found that startups address the incomplete knowledge problem by beta testing on platforms and that those with male-focused products benefit from the predominance of male users on these platforms, which results in higher evaluations and more growth for startups with male-focused (vs. female-focused) products. Our theorizing adds the mechanism of organization-wide processes of inquiry as an additional advantaging mechanism – the “benefit of the doubt” – for startups with male teams. We suggest that examining the decision-making processes used to address the incomplete knowledge problem that uncertainty poses could provide more insight into how disparities are produced.

6.1.3. The Promise of Systematizing Inquiry. When theorizing how organizations can mitigate gender biases in evaluation, scholars have examined the effects of issuing stricter regulations and structured processes around evaluation, and limiting individual evaluator discretion in hiring and promotion contexts (e.g., Goldin and Rouse 2000, Castilla 2008, Huffcutt 2011, Rivera and Tilcsik 2019, Stephens et al. 2020, Benson et al. 2022). We hypothesized how this might be applied in innovation contexts, in which past data on startups are insufficient to bring about decision-making (e.g., Stinchcombe 1965, Aldrich and Fiol 1994, Cohen et al. 2019a), and how investors value the ability to use discretion: assessing potential during interactions with founders (Kirsch et al. 2009, Petty

and Gruber 2011, Huang 2018, Miller et al. 2023). These processes of inquiry can themselves engender disparities in evaluation in innovation contexts (Kanze et al. 2018) and more broadly (Rivera 2012a, b, 2015; Stephens et al. 2020), yet their causal effects on economic outcomes are understudied.

In this context, we theorized that interventions that allow investors to inquire freely – assessing the potential of innovators and their ideas during interactions – would be important. We hypothesized that organizations that prompt investors to inquire systematically (on risk, reward, and progress) but *still freely* during interactions could produce less gender disparities in investment outcomes. We show the effect on assessment processes for treated investors who (1) assessed more consistently and (2) assessed start-up competence more dynamically than a control. This eliminated, and even reversed, the gender gap in investment decisions. Data collected through inquiry, particularly data dynamically collected and assessed during multiple interactions over the selection process, can increase the types of data investors use to make collective decisions. This could outweigh the benefits of reducing individual evaluator discretion.

Extrapolating from this finding, we theorize a broader implication for many investment organizations whose evaluators assess multiple candidates, allocate funds, and use processes of inquiry to gather data on potential, which is not readily available in static form (e.g., when university hiring committees attempt to assess the potential of an early-stage candidate’s research pipeline). Systematizing processes of inquiry by prompting evaluators to inquire consistently about dynamic content has the potential to reduce disparities in assessments and outcomes across these contexts.

6.2. Policy Implications

Policy makers and investors are increasingly recognizing the importance of improving gender diversity in investing. Development finance institutions, such as the International Finance Corporation, DFC, and British International Investment, are committing billions of dollars to invest in female-founded ventures (DFC 2021), as are private investment funds, such as Fidelity and Nia Impact Capital, and

venture capital firms, such as Andreessen Horowitz. In this context, our findings have a number of policy implications for reducing gender disparities in investment systems by changing the way investment organizations evaluate early-stage startups.

First, we designed interventions to reduce the chance of backlash by changing investment evaluation processes without explicitly focusing on gender (e.g., Leslie 2019) and to retain investment organizations' focus on identifying the most promising ventures. Systematizing inquiry might increase efficiency, as investors make decisions based on more complete information. This should increase investment organizations' interest in applying these interventions.

Second, these interventions are relatively inexpensive changes to processes, compared with costly training programs, investment guarantees, or provision of supplementary funding for female-founded ventures. These findings offer a promising avenue for development finance institutions to improve gender diversity in their portfolios in a cost-effective and efficient way. 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 a dynamic 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. However, in most investment organizations, startups could make progress between filling out an application form/sending over a pitch deck and having an interview with an investor. For example, Vilcap added a question to its interview template to assess start-up 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?)” Interventions such as this could help investment organizations detect startups that are able to make rapid improvement – a capability both important to many investors and rewarding for startups with female founders.

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

7. Conclusion

Rather than prepare female founders to interact differently with investors, we examined how investment organizations can reduce gender 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. When investment organizations prompted investors to systematically inquire about risk and reward, this resulted in more consistent investor assessment across startups. When investment organizations prompted investors to systematically inquire about start-up progress, investors assessed the venture more dynamically, which convinced treated investors that startups with female founders could make rapid progress and scale. This eliminated, and even reversed, the gender gap in investment decisions. Changing organizational practices to systematize investor inquiry can meaningfully affect investment outcomes. This has implications not only for entrepreneurship theory but also for a broader set of organizations funding innovation in uncertain contexts and those interested in reducing gender disparities.

References

- Aldrich HE, Fiol CM (1994) Fools rush in? The institutional context of industry creation. *Academy of Management Review* 19(4):645-670.
- Amis JM, Mair J, Munir KA (2020) The organizational reproduction of inequality. *Academy of Management Annals* 14(1):195-230.
- Anthony C (2021) When knowledge work and analytical technologies collide: The practices and consequences of black boxing algorithmic technologies. *Administrative Science Quarterly* 66(4):1173-1212.
- Balachandra L, Briggs T, Eddleston K, Brush C (2019) Don't pitch like a girl! How gender stereotypes influence investor decisions. *Entrepreneurship Theory and Practice*, 43(1):116-137.
- Bapna S, Ganco M (2021) Gender gaps in equity crowdfunding: Evidence from a randomized field experiment. *Management Science* 67(5):2679-2710.
- Baron RA, Ensley MD (2006) Opportunity recognition as the detection of meaningful patterns: Evidence from comparisons of novice and experienced entrepreneurs. *Management Science* 52(9):1331-1344.
- Benson A, Li D, Shue K (2022) "Potential" and the gender promotion gap. Working paper, University of Minnesota.
- Bernstein S, Korteweg A, Laws K (2017) Attracting early-stage investors: Evidence from a randomized field experiment. *Journal of Finance* 72(2):509-538.
- Biegel A, Hunt SM, Matteucci A (2020) Project SAGE 3.0: Tracking venture capital, private equity, and private debt with a gender lens. Wharton Social Impact Initiative. <https://socialimpact.wharton.upenn.edu/research-reports/reports-2/project-sage-3/>.
- Bingham CB, Eisenhardt KM (2011) Rational heuristics: The "simple rules" that strategists learn from process experience. *Strategic Management Journal* 32(13):1437-1464.
- Bohren JA, Imas A, Rosenberg M (2019) The dynamics of discrimination: Theory and evidence. *American Economic Review* 109(10):3395-3436.
- Botelho TL, 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 AW, Huang L, Kearney SW, Murray FE (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. Accessed at <https://newsandviews.vilcap.com/reports/flipping-the-power-dynamics-can-entrepreneurs-make-successful-investment-decisions>
- Camuffo A, Cordova A, Gambardella A, Spina C (2020) A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. *Management Science* 66(2):564-586.
- Canales R (2014) Weaving straw into gold: Managing organizational tensions between standardization and flexibility in microfinance. *Organization Science* 25(1):1-28.
- Cao R, Koning RM, Nanda R (2021) Biased sampling of early users and the direction of startup innovation. National Bureau of Economic Research No. 28882.
- Castilla EJ (2008) Gender, race, and meritocracy in organizational careers. *American Journal of Sociology* 113(6):1479-1526.
- Castilla EJ (2015) Accounting for the gap: A firm study manipulating organizational accountability and transparency in pay decisions. *Organization Science* 26(2):311-333.
- Castilla EJ, 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 DR, Fang TP, 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 SL, Bingham CB, Hallen BL (2019a) The role of accelerator designs in mitigating bounded rationality in new ventures. *Administrative Science Quarterly* 64(4):810-854.
- Cohen S, Fehder DC, Hochberg YV, Murray F (2019b) The design of startup accelerators. *Research Policy* 48(7):1781-1797.
- Correll SJ, Benard S (2006) Biased estimators? Comparing status and statistical theories of gender discrimination. Thye SR, Lawler EJ, eds. *Advances in Group Processes*, Vol. 23 (Emerald Group Publishing Limited, Bingley), 89-116.
- Correll SJ, Ridgeway CL, Zuckerman EW, 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*. Accessed at https://ssir.org/articles/entry/gender_lens_investing_strategies_for_2019
- Czibor E, Jimenez-Gomez D, List JA (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. *Handbook of the Economics of Finance*, 2:573-648.
- Delecourt S, Ng O (2021) Does gender matter for small business performance? Experimental evidence from India. Working Paper, SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3830935
- DFC (2021) Global gender finance initiative sets ambitious new \$15 billion fundraising goal after securing more than double its original \$3 billion target. <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* 68(12): 8635-8657.
- 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.
- Dushnitsky G, Sarkar S (2022). Here comes the sun: The impact of incidental contextual factors on entrepreneurial resource acquisition. *Academy of Management Journal* 65(1): 66-92.
- Eisenhardt KM, Tabrizi BN (1995) Accelerating adaptive processes: Product innovation in the global computer industry. *Administrative Science Quarterly* 40(1):84-110.
- Elsbach KD, Kramer RM (2003) Assessing creativity in Hollywood pitch meetings: Evidence for a dual-process model of creativity judgments. *Academy of Management Journal* 46(3):283-301.
- Ewens M, Townsend RR (2020) Are early stage investors biased against women? *Journal of Financial Economics* 135(3):653-677.
- Fayard AL, Metiu A (2014) The role of writing in distributed collaboration. *Organization Science* 25(5): 1391-1413.
- Fernandez-Mateo I, Kaplan S (2018) Gender and organization science: Introduction to a virtual special issue. *Organization Science* 29(6): 1229-1236.
- Fried VH, Hisrich RD (1994) Toward a model of venture capital investment decision making. *Financial Management* 23(3):28-37.
- Frost (2020) How VC bias in viewing pitch decks can affect fundraising success. Docsend. <https://www.docsend.com/blog/how-vc-bias-in-viewing-pitch-decks-can-affect-fundraising-success/>.
- Gamache DL, McNamara G, Mannor MJ, Johnson RE (2015) Motivated to acquire? The impact of CEO regulatory focus on firm acquisitions. *Academy of Management Journal* 58(4):1261-1282.
- Gans JS, Stern S, Wu J (2019) Foundations of entrepreneurial strategy. *Strategic Management Journal* 40(5):736-756.
- Goldin C, Rouse C (2000) Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review* 90(4):715-741.
- Gompers PA, Wang SQ (2017) And the children shall lead: Gender diversity and performance in venture capital. National Bureau of Economic Research No. w23454.
- Gompers PA, Gornall W, Kaplan SN, Strebulaev IA (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. <https://www.galidata.org/publications/does-acceleration-work/>.
- Guzman J, Kacperczyk AO (2019) Gender gap in entrepreneurship. *Research Policy* 48(7):1666-1680.
- Hallen BL (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 BL, Davis JP, Murray A. (2020) Entrepreneurial network evolution: Explicating the structural localism and agentic network change distinction. *Academy of Management Annals* 14(2):1067-1102.
- Hallen BL, Eisenhardt KM (2012) Catalyzing strategies and efficient tie formation: How entrepreneurial firms obtain investment ties. *Academy of Management Journal* 55(1):35-70.
- Higgins MC, 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 ST, Nanda R (2019) Networking frictions in venture capital, and the gender gap in entrepreneurship. National Bureau of Economic Research No. w26449.
- Hsieh CT, Hurst E, Jones CI, Klenow PJ (2019) The allocation of talent and US economic growth. *Econometrica* 87(5):1439-1474.
- Hsu DK, Simmons SA, Wieland AM (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.
- Huffcutt AI (2011) An empirical review of the employment interview construct literature. *International Journal of Selection and Assessment*, 19(1):62-81.
- Ibarra H (1993) Personal networks of women and minorities in management: A conceptual framework. *Academy of Management Review*, 18(1): 56-87.
- Jennings JE, Brush CG (2013) Research on women entrepreneurs: Challenges to (and from) the broader entrepreneurship literature?. *The Academy of Management Annals* 7(1):663-715.
- Jeppesen LB, Lakhani KR (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 MA, Higgins ET (2017) Male and female entrepreneurs get asked different questions by VCs—and it affects how much funding they get. *Harvard Business Review*. <https://hbr.org/2017/06/male-and-female-entrepreneurs-get-asked-different-questions-by-vc-s-and-it-affects-how-much-funding-they-get>.
- Kanze D, Huang L, Conley MA, Higgins ET (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 (2011) Strategy and PowerPoint: An inquiry into the epistemic culture and machinery of strategy making. *Organization Science* 22(2): 320-346.
- Kim JW, King BG (2014) Seeing stars: Matthew effects and status bias in major league baseball umpiring. *Management Science* 60(11):2619-2644.
- 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 (2023) What is a pivot? Explaining when and how entrepreneurial firms decide to make strategic change and pivot. *Strategic Management Journal* 44(1):197-230.
- Knight FH (1921) *Risk, Uncertainty and Profit* (Hart, Schaffner and Marx, New York, NY).
- Koning R, Samila S, Ferguson JP (2020) Inventor gender and the direction of invention. *AEA Papers and Proceedings* 110:250-54.
- Lall SA, Chen LW, Roberts PW (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* 38:201-21
- Lee M, Huang L (2018) Gender bias, social impact framing, and evaluation of entrepreneurial ventures. *Organization Science* 29(1):1-16.
- Leibbrandt A, Wang LC, 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):237-261.
- Leslie LM (2019) Diversity initiative effectiveness: A typological theory of unintended consequences. *Academy of Management Review* 44(3):538-563.
- Lounsbury M, Glynn MA (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 Hollywood. *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 JG, Simon HA (1958) *Organizations* (Wiley, New York, NY).

- March JG (1991). Exploration and exploitation in organizational learning. *Organization Science* 2(1):71-87.
- Martens ML, Jennings JE, Jennings PD (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.
- McDonald RM, Eisenhardt KM (2020) Parallel play: Startups, nascent markets, and effective business-model design. *Administrative Science Quarterly* 65(2):483-523.
- McKenzie D (2012) Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics* 99(2):210-221.
- Miller A, O'Mahony S, Cohen SL (2023) Opening the aperture: Explaining the complementary roles of advice and testing when forming entrepreneurial strategy. *Organization Science* (online).
- Pahnke EC, Katila R, Eisenhardt KM (2015) Who takes you to the dance? How partners' institutional logics influence innovation in young firms. *Administrative Science Quarterly* 60(4):596-633.
- Petty JS, 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 LT, 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 RE, Schneider B, Schmitt N (2005) *Staffing Organizations: Contemporary Practice and Theory* (Lawrence Erlbaum, Mahwah, NJ).
- Podolny JM (1993) A status-based model of market competition. *American Journal of Sociology* 98(4): 829-872.
- Ridgeway CL, Correll SJ (2000) Limiting inequality through interaction: The end (s) of gender. *Contemporary Sociology* 29(1):110-120.
- Ridgeway CL, Correll SJ (2004) Unpacking the gender system: A theoretical perspective on gender beliefs and social relations. *Gender & Society* 18(4):510-531.
- Ridgeway CL (2014) Why status matters for inequality. *American Sociological Review* 79(1): 1-16.
- Rindova V, Courtney H (2020) To shape or adapt: Knowledge problems, epistemologies, and strategic postures under Knightian uncertainty. *Academy of Management Review* 45(4):787-807.
- Rivera LA (2012) Hiring as cultural matching: The case of elite professional service firms. *American Sociological Review* 77(6):999-1022.
- Rivera LA (2015) Go with your gut: Emotion and evaluation in job interviews. *American Journal of Sociology* 120(5):1339-1389.
- Rivera LA, Tilcsik A (2019) Scaling down inequality: Rating scales, gender bias, and the architecture of evaluation. *American Sociological Review* 84(2):248-274.
- Roberts PW, Lall SA (2018) *Observing Acceleration: Uncovering the Effects of Accelerators on Impact-Oriented Entrepreneurs* (Springer, Cham).
- 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, Cambridge, MA).
- Siggelkow N (2002) Evolution toward fit. *Administrative Science Quarterly* 47(1):125-159.
- Simcoe TS, Waguespack DM (2011). Status, quality, and attention: What's in a (missing) name?. *Management Science* 57(2): 274-290.
- 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 TE (2001) Syndication networks and the spatial distribution of venture capital investments. *American Journal of Sociology* 106(6):1546-1588.
- Stephens NM, Rivera LA, Townsend SS (2020) What works to increase diversity? A multi-level approach. Working Paper. www.nicolestephens.com
- Stinchcombe AL (1965) Social Structure and Organizations. In: March JP. Ed. *Handbook of Organizations*, Rand McNally, Chicago, 142-193.
- Suster M (2010) Invest in Lines Not Dots (Posted Nov 16 2010). Accessed at <https://bothsidesofthetable.com/invest-in-lines-not-dots-611f36491d73>
- Tyebjee TT, Bruno AV (1984) A model of venture capitalist investment activity. *Management Science* 30(9):1051-1066.
- Uhlmann EL, Cohen GL(2005) Constructed criteria: Redefining merit to justify discrimination. *Psychological Science* 16(6):474-480.
- Yang T, Aldrich HE (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.

Table 1. Whole Sample

		Whole sample (scores)	Cross-section	Balanced Panel (3 rounds)
Investment decisions	Per investor x start-up x round x criterion	33,541	n/a	n/a
	Per investor x start-up x round	3,127	n/a	1,530
	Per investor x start-up	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

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

- (a) 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

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%	0.477
	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%	0.942
	Professional	198	71%	92	69%	88	69%	18	100%	.
Gender (N=262)	Female	87	33%	42	33%	40	34%	5	29%	0.963
Type (N=179)	Investor***	68	38%	29	35%	29	37%	10	59%	0.816
	Local organization	141	84%	66	85%	62	86%	13	76%	0.798
Investment organization (N=154)	Diversity mandate	45	29%	19	26%	22	33%	4	27%	0.376
	Impact mandate	93	60%	43	59%	39	59%	11	73%	0.982
B. Investor Decisions (1,341)		All		Control		Treated: Risk/reward		Treatment and control check*		
		#	%	#	%	#	%			
Decisions	Investor × start-up	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 × start-up	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 (e.g., venture capital 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 (FF)	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
Start-up 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 ² /pseudo-R ²	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

DD = Due Diligence.

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

Model 3 was preregistered, and all other models provide similar directional results.

Model 4 suggests that the investor's gender does not change the main relationships between variables.

Model 5 suggests partial mediation.

Table 4. Panel Sample

Investor and Start-up Characteristics	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	10	0.615
Startups	69	100.0%	36	33	.	
Female founder	32	46.4%	15	17	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	12.22	0.346
Funds raised (mean)**	64	\$237,896	\$179,152	\$296,640	\$296,640	0.161
Log funds raised (mean)**	64	10.12	9.89	10.44	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 Z-score (rounds 1–3)			
	1	2	3	4
Female founder (FF)	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
Start-up controls			x	x
Female investor				x
N	1,530	1,530	1,395	1,395
R ²	0.0496	0.0544	0.0751	0.0752
Investors	65	65	65	65

Start-up 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 start-up elements	9 investors (100%)	3 investors (43%)
Control, Female	People that I rated highly ... [I thought,] “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 start-up 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			

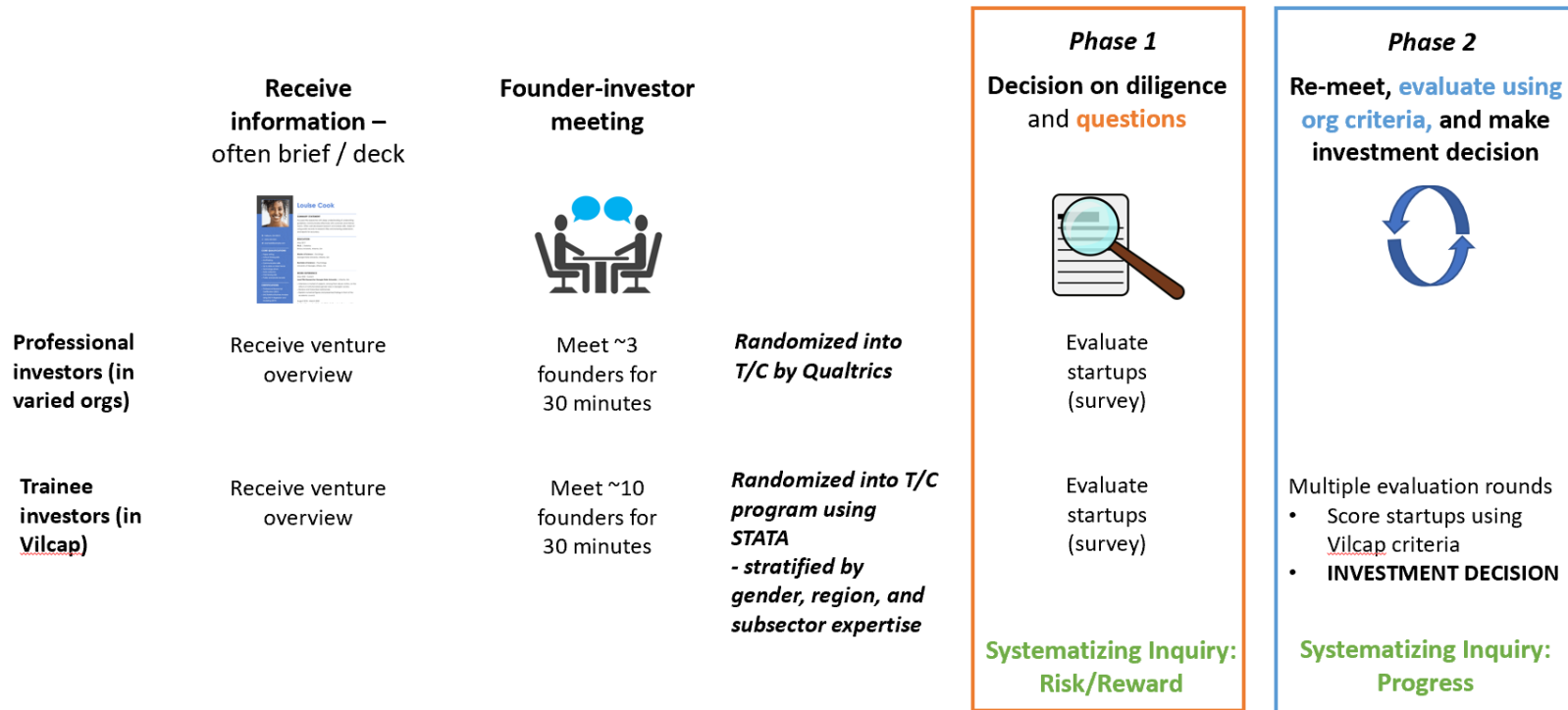
Table 7. Effect of Systematizing Inquiry on Score Elements (Panel)

	DV Score elements (scales of 1–4)								
	All	Business model	Investor exit	Market	Problem and vision <i>(Rounds 1–3)</i>	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
Start-up 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 ²	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

Start-up controls = number of employees, funds raised (log)

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

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)

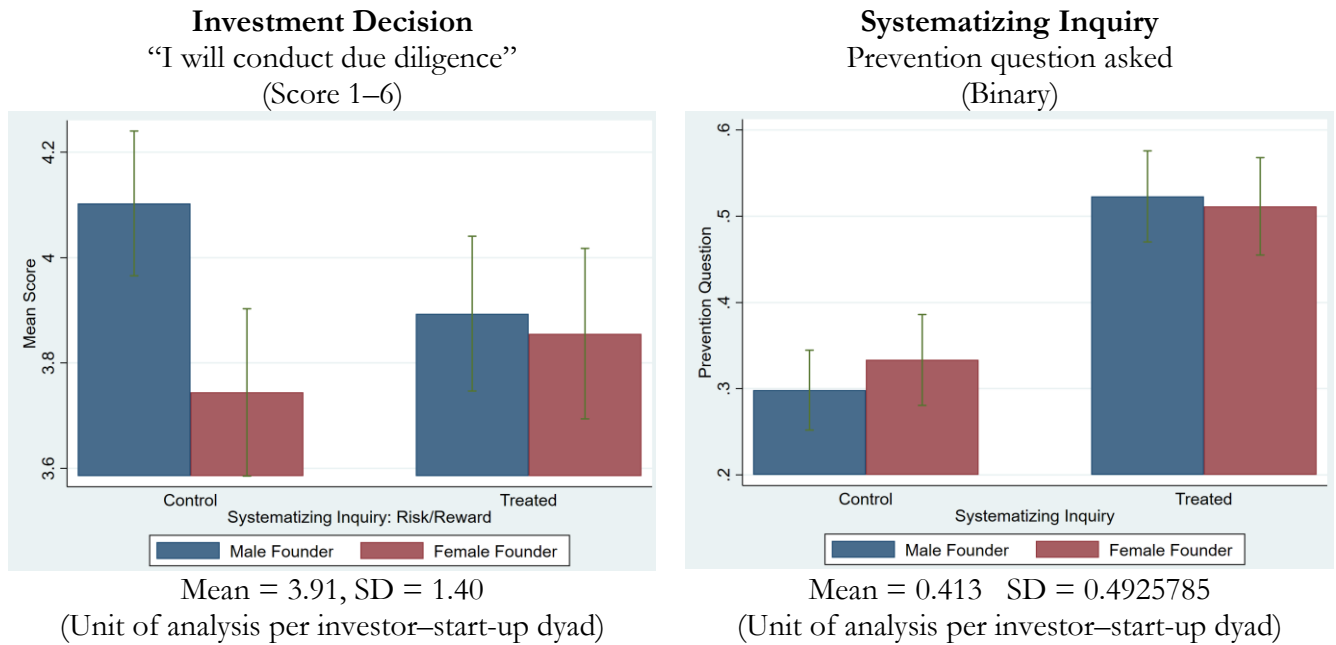
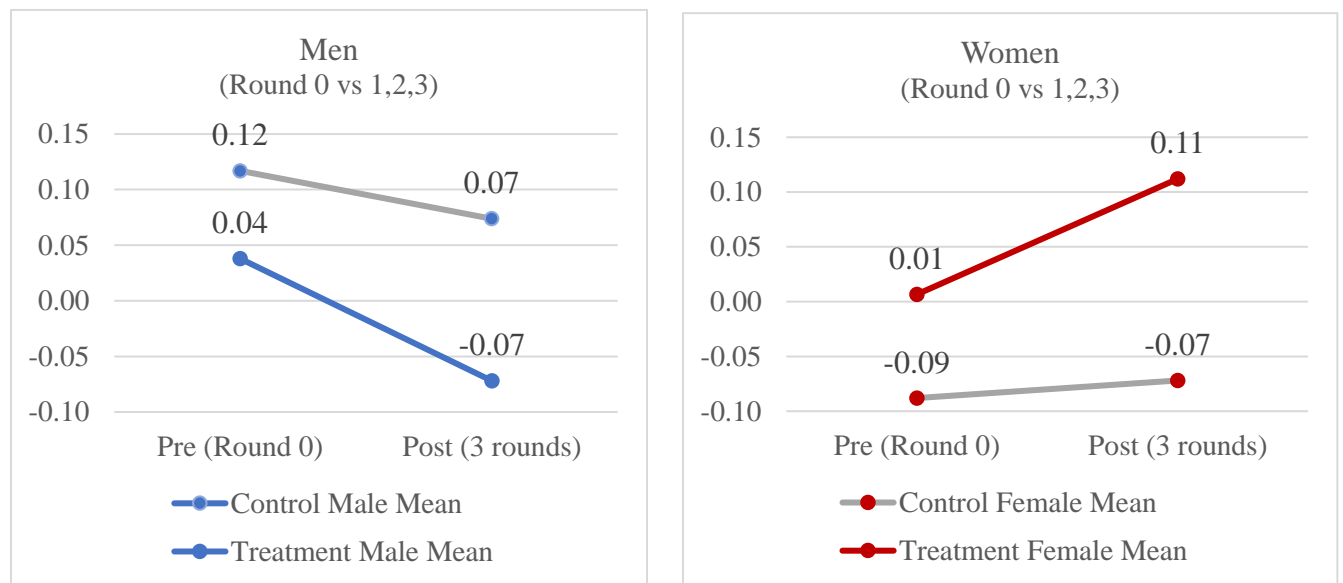


Figure 3. Effect of Systematizing Inquiry on Investment Decisions (Z-score Over Time – Panel)

Panel A: Mean Z-score over Time – Male Founder Panel B: Mean Z-score over Time – Female Founder



Appendix A. Evaluation Surveys

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
Start-up controls		x		x
N	688	581	688	581
R ² /pseudo-R ²	0.0116	0.0080	0.0058	0.0079
Investors	151	151	151	151

Start-up controls = number of employees, funds raised.
DD = Due Diligence.

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
Start-up controls	x	x	x	x
Female investor		x		x
N	1,162	1,133	1,162	1,133
R ²	0.0328	0.0356	0.0037	0.0050
Investors	276	260	276	260

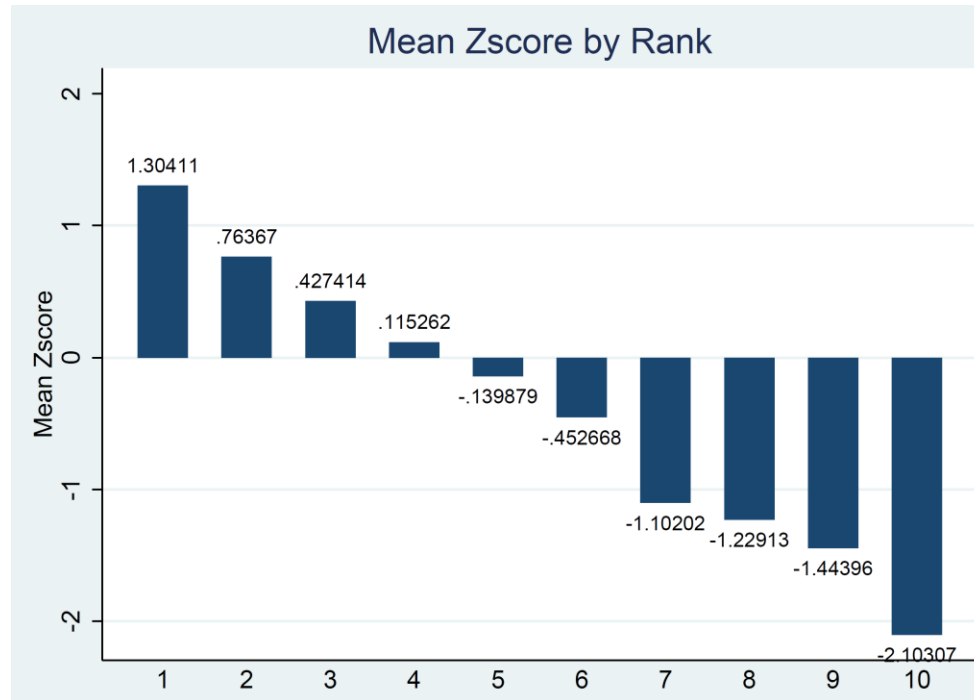
Start-up controls = number of employees, funds raised. FF = female founder.
DD = Due Diligence.

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 Z-score (continuous) OLS		Prevention question (binary) Logit odds ratio		DV Z-score (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
Start-up controls		x		x		x		x
N	412	323	412	323	813	723	813	723
Pseudo-R ² /R ²	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

Start-up controls = number of employees, log of funds raised.

Appendix E. Mean Z-score by Rank



After creating a z-score = (score – avg_score)/sd_score, Vilcap averages the z-scores across all investors. The highest z-score becomes rank 1, followed by rank 2, and so on. The lowest z-score 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)

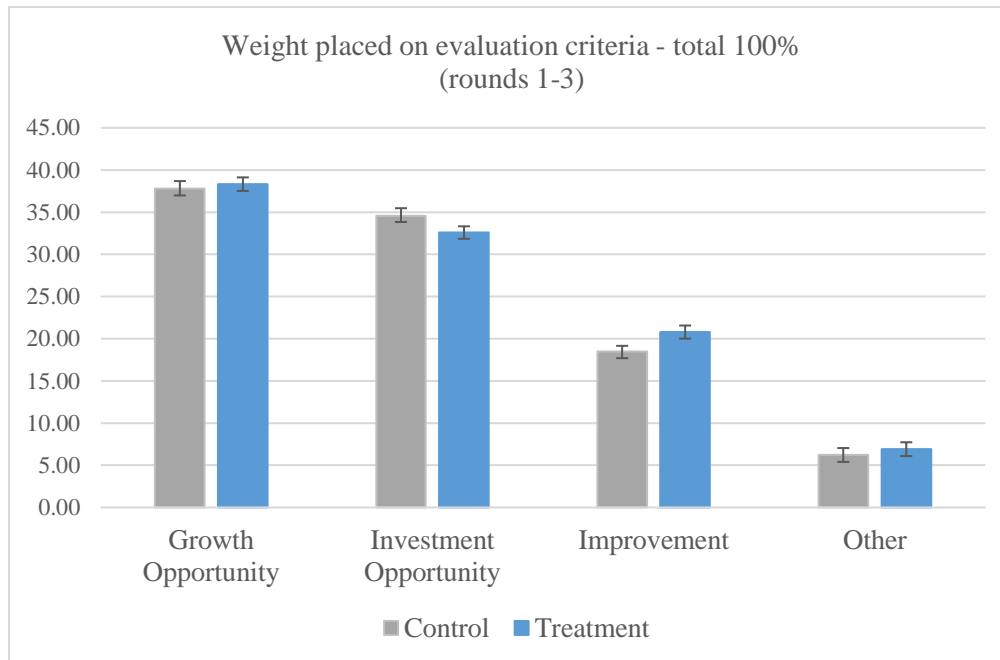
	Prevention questions (binary) logit odds ratio (rounds 1–3)			DV z-score ANCOVA (rounds 1–3)	
	1	2	3	4	5**
Female founder (FF)	1.019 (0.123)	1.004 (0.217)	1.218 (0.288)	-0.059 (0.115)	-0.055 (0.115)
Systematized inquiry		4.432*** (1.154)	4.769*** (0.409)	-0.157* (0.077)	-0.146 (0.080)
FF × Inquiry		0.788 (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
Start-up controls			x	x	x
N	1,530	1,530	1,395	1,395	1,395
Pseudo-R ² /R ²	0.1537	0.2117	0.2155	0.0751	0.0754
Investors	65	65	65	65	65

Start-up controls = number of employees, log of funds raised.

Model 3 was preregistered, and all other models provide similar directional results.

**Model 5 is mediation analysis (shows little evidence of moderation).

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



Appendix H. Female Variable Robustness Check (Panel)

	ANCOVA - DV z-score (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
Start-up controls			x
N	1,530	1,530	1,395
R ²	0.0522	0.0579	0.0791
Investors	65	65	65

FF = female founder.

Appendix I. Start-up Characteristics by Founder Gender

	DV - Proxies for start-up 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 ²	0.2050	0.1278	0.0763	0.0408