Understanding Probabilistic Reasoning in Innovation

Andrea Coali[1], Alfonso Gambardella[1], and Elena Novelli[2]

[1]Bocconi University
[2]Bayes Business School

Preliminary Draft
Please do not circulate without authors' permission

April 2022

Abstract

Organizational decision-making processes are characterized by high uncertainty and prone to decision-making biases. In this paper we explore the mechanisms and implications associated with what we call a *scientific approach to decision making*, based on probabilistic reasoning. We develop a structural model to disentangle and identify two separate but complementary effects of this approach. We estimate our structural model, based on data from two randomized control trials (RCTs) involving early stage start-ups and show that scientific entrepreneurs tend to be more conservative in assessing the value over their ideas, an effect that we call *debiasing effect*. We also show that, conditional on their decision to remain operational, scientific entrepreneurs tend to perform better, an effect that we call *learning effect*. We finally explore the implications associated with the use of this approach and show how the selection induced by the debiasing effect does not significantly increase the rate of false-negatives for treated firms compared to the control group. We discuss the implications for future research and practice.

1 Introduction

An astonishing 90% of newly-born start-ups fail within ten years, with around 21% of them failing already in their first year (NationalBusinessCapital, 2020). Part of the reasons behind this pattern relates to the fact that entrepreneurs - and organizational decision makers involved in innovative projects more in general - face a decision-making process that is characterized by high uncertainty along multiple dimensions (McGrath, 1997, Folta, 1998). In the presence of uncertainty, the assessment on the value of novel ideas, and therefore decision-making, becomes difficult.

One way to deal with uncertainty is to use a probabilistic decision-making process, making decisions on uncertain outcomes based on probabilistic information. However, decision makers often deviate from “rational” or probabilistic reasoning, neglecting systematic decision-making process (Bloom and Van Reenen, 2007) and even ignoring relevant information (Bennett and Chatterji, 2019). Alternatively, they rely on heuristic principles to reduce the complex tasks of probability assessment and value prediction to a simpler task (Tversky and Kahneman, 1974). Whereas this can lead to good results (Bingham and Eisenhardt, 2011), research shows that it can also lead to a plethora of severe biases (Tversky and Kahneman, 1974).

In this paper we aim to understand more about what happens when entrepreneurs are induced to employ probabilistic decision-making processes. We focus in the context of entrepreneurial ventures and address the following research question: *What are the main mechanisms through which probabilistic reasoning supports organizational decision making and what are its implications?*

We focus on an approach to learning and decision making that we refer to as a ”scientific” (Ashraf, Banerjee, and Nourani, 2021; Camuffo et al., 2020), based on probabilistic reasoning. Decision makers following this approach (1) develop a theory of their business idea and the problems it would likely solve, (2) develop hypotheses flowing logically from it, (3) design tests that can provide them with signals regarding the probabilities of those hypotheses being supported with data, (4) evaluate test results in a disciplined way against their prior theoretical expectations. This process resembles the one followed by scientists in developing new knowledge. It supports decision making by combining a cognitive ”off-line” representation of the decision with its ”on-line” validation via the collecting data from the environment.

Prior research that has explored the use of similar approaches to decision making has reported that they are associated with higher project termination rates and superior learning and performance (Camuffo et al., 2020, McDonald and Eisenhardt, 2020). However, research in this area has not yet fully elaborated nor provided direct evidence on the mechanisms that connect the use of such an approach with the observed outcomes. This is perhaps not surprising given the research-design challenges that filling this gap involves. First, answering this question requires an exogenous shock that induces decision makers to reason and make decisions in probabilistic terms. Second, it requires observing the decision-making process steps in detail and not only the outcomes originating from that process, such as the final decisions made and their performance. Third, it requires comparing decision makers using a probabilistic approach to a proper counterfactual. This paper leverages unique data that enable us to respond to these challenges.

Specifically, we first develop a structural model that disentangles and identifies two different effects underlying a scientific approach to decision making. Our model suggests that entrepreneurs following a scientific approach make an earlier and faster downward adjustment of their business’ expected values, ultimately showing a higher rate of project termination, an effect that we call the *debiasing effect*; but they also learn to understand the problem faced better and identify a better solution, an effect that we call the *learning effect*. We then estimate the structural model with data from two randomized control trials (RCTs) conducted in Milan (2017) and Turin (2018), involving 382 early stage start-ups. The RCT design enables us to teach a sample of entrepreneurs to adopt a scientific approach in their decision making. We maintain the other half of the sample in a control condition, where they are delivered equivalent management content but without a scientific approach. We then monitor these entrepreneurs for a variety of months, collecting detailed data on their decision-making process, choices, and performance.

The uniqueness of these data compared to prior studies is that they include the entrepreneurs' own assessment of the potential value of their ideas, which enables us to retrieve all the structural parameters of interest and separate the *debiasing effect* from the *learning effect*. The focus on an entrepreneurial decision making process is ideal as it ensures that individuals undergoing the treatment are key decision makers within the organization and that they can be closely observed at a time in which they are making key decisions on core organizational activities (Siggelkow, 2002)

The co-existence of these two effects leads to a natural follow up question, which is important in the spirit on understanding the implications of the scientific method. Given that a scientific approach to decision making leads both to a more conservative assessment of ideas but also to superior learning and performance, are scientific decision makers excessively cautious when terminating their projects, effectively discarding ideas that could, instead, be successful? In other words, is it possible that while this method leads entrepreneurs to reduce the number of ideas that others would have falsely seen as positive, it also leads entrepreneurs to discard too many (falsely negative) ideas?

Answering this question is no easy task, as it requires the determination of the value of the terminated ideas, were they not terminated, which is clearly not possible. We contribute to this important debate, by presenting four distinct pieces of suggestive evidence. First, we analyze the pattern of funding received, revenue growth over time and survival of treated and control firms. Second, we asked a team of professional to provide an evaluation of the pitches of the ideas developed by treated and control entrepreneurs in the two RCTs. Third, leveraging on the results of the structural estimation, we identify different assumptions and use them to calculate two different scenarios at different end of the spectrum of possibilities. Finally, we also replicate results on termination and on the selection trade-off using a business simulation game. Overall, we do not find evidence in favor of the idea that treated entrepreneurs who decided to terminate were too conservative and that their rate of false negative was higher that their rate of reduction in false positives. This suggests that the selection of firms induced by the scientific training was a positive one.

Overall, this paper makes two unique contributions. First, our model and unique data on entrepreneurs' own assessment of their ideas enable us to provide direct evidence of the mechanisms that connect the use of a scientific approach to its outcomes, providing clarity of the effects of a approaches to decision making that combine cognition with action (Levinthal, 2017, McDonald and Eisenhardt, 2020). Second, this paper explores the trade-off associated with the stricter selection process induced by the scientific approach, providing some original evidence in this direction.

The paper is structured as follows. Section II offers a theoretical background on a scientific approach to organizational decision making. Section III details the structural framework we develop. Section IV describes our methodology and data. Section V reports the estimation results. In Section VI we discuss the trade-off between the false-positive and false-negative rate, while Section VII offer concluding reflections.

2 Theoretical Background

2.1 Organizational Decision Making and a Scientific Approach

Prior literature in management and organization has amply shown how organizational decision making under uncertainty is affected by biases. Decision makers often make decisions following their gut feelings as opposed to using structured approaches that systematically take all available information into account(Bennett and Chatterji, 2019). At the same time, some scholars have documented the benefits that decision makers can derive from the use of structured approaches that support managerial and entrepreneurial decision making, such as the use of structured practices (Bloom and Van Reenen, 2007; Yang et al., 2020; Ott, Eisenhardt, and Bingham, 2017).

Within this context, two main classes of approaches have been emphasized by prior literature (Ott, Eisenhardt, and Bingham, 2017). On the one hand cognition-based approaches to decision making, based on the development of cognitive structures and mental models that help individuals guides an understanding of markets, firms, and strategies (Csaszar and Laureiro-Martínez, 2018, Felin and Zenger, 2009, Gary and Wood, 2011, Walsh, 1995). The core logic supporting the use of cognition-based approaches is that the environment and its future states cannot be predicted or known, but that an understanding of the causal relationship between market characteristics and organizational choices can lead to superior strategies. On the other hand, action-based approaches to decision making hinge upon the idea that - in the face of an environment that cannot be predicted (Eisenhardt, 1989) - taking actions and then adjusting those actions based on the feedback obtained by the environment can lead to valid organizational decisions. Action-based approaches, include, for instance, trial and error learning and heuristics (von Hippel and Tyre, 1995,Bingham and Eisenhardt, 2011, Bingham and Davis, 2012), and experimentation (Ries, 2011; Thomke, 1998).

More recent studies have advanced the idea that these two types of approaches can be successfully combined in a "decision weaving" process that can actually lead to acquiring knowledge about the environment and use that knowledge as a guide for action (Ott, Eisenhardt, and Bingham, 2017, Eisenhardt and Ott, 2017). For instance Gavetti and Rivkin (2007) provide a detailed description of how executive at Lycos developed the company's strategy by combining insights obtained from feedback on their actions, together with the executives' mental representation of the Internet Portal industry. McDonald and Eisenhardt (2020) elaborate on the benefits of deliberate learning, and of testing the cognitive assumptions underlying a business model before committing to it. They show that such an approach leads to superior and faster decision making by reducing the uncertainty faced and the extent to which decisions are based on emotions and opinions.

Camuffo et al. (2020) and Camuffo et al. (2021) focus on a specific way in which a cognition-based approach and an action-based approach can be combined, which they refer to as a "scientific approach" to decision making, and which is the focus of this paper. A scientific approach to decision making starts with a cognitive approach to the problem faced, which includes the definition of assumptions about the environment and a theory on the relationship between the problem and the solution devised. These assumptions and causal relationships are then translated into formal predictions or "hypotheses", which effectively correspond to the definition of prior probabilities. This initial phase is complemented with an action-based approach to the problem, consisting in conducting tests that can support or reject the hypotheses defined in the first phase. Test results are carefully examined and used as feedback to revise the cognitive representation of the problem initially developed. By employing this approach in an iterative way over time, the decision maker develops an accurate understanding of the structure and the distribution of outcomes in the environment, the effectiveness of the identified solution under different contingencies and makes decisions in a probabilistic logic.

2.2 A Scientific Approach to Decision Making: A Stylized Example

We illustrate the use of a scientific approach to decision making in practice with a stylized example. Consider a decision maker with the goal of developing an innovative product or service, or willing to launch a new business. Assume she starts with an intuition coming from observation of real-world phenomena, spotting a problem that would need an innovative solution to be solved. Before deciding to embark in this new project, our decision-maker will evaluate whether her idea is worth the development efforts and this assessment will be made at regular intervals throughout the life of the project. At every round of assessment, her decision can be represented as a choice between three mutually exclusive alternatives (Kirtley and O'Mahony, 2020, Lieberman, Lee, and Folta, 2017, Eisenhardt and Bingham, 2017, Gans, Stern, and Wu, 2019): 1) *terminate* the project, if they believe it won't generate sufficient value; 2) change substantial elements of the idea or project to improve its value (what we refer to as *pivoting*); 3) *continue* the development of the project along its current trajectory.

Along the way, her assessment will be based on considerations regarding the multiple potential scenarios she could face in the environment in which she operates, over which there is uncertainty. This uncertainty could originate, for instance, from the fact that she is not yet familiar with customers' preferences in the environment she targets; or from the fact that these preferences might be subject to change. She will also consider the organizational choices she can make to deal with the multiple scenarios she might be facing. At the very early stage of her process, organizational choices could concern the basic features of the idea. At later stages, choices to be made could be linked to the idea commercialization and could include, for example, the development of different versions of the same product, service, or business model, or the implementation of alternative marketing strategies.

Of course, every choice she envisions might have a different value under different scenarios, and she needs to make these choices in the face of uncertainty. Suppose for instance that our entrepreneur's idea is about developing an innovative service for car-sharing, but there is uncertainty regarding the extent to which cars are going to be relevant in the medium term in the context in which she is operating. If the context in which she operates is going through a massive drop in the use of cars and an increase in the use of bikes, the choice of pursuing such car-sharing project could have a negative value. Instead, if renting cars is a valuable option in the context in which she operates, the choice of pursuing such car-sharing project idea could have a high value. Depending on what her prior on the scenario more likely to manifest itself and what value she envisions her choices to have, she could decide to terminate the project, or to pivot to a new version of the project, or to simply continue the development of the project along its natural trajectory.

If our entrepreneur approached the problem in a scientific way, she would start by developing a theory about the problem that the car-sharing service addresses and the way in which it addresses it, and how the value of each choice would change under different relevant scenarios. She would then decompose the theory into core hypotheses regarding the scenarios she is facing and the value of choices in those scenarios. For instance, relevant hypotheses could be: "Car transportation in large cities is highly valued by families" or "The majority of adults living in a large city believe that owning a car in a large city is not practical due to the high fixed costs and the limited use per person", or "The majority of adults living in a large city consider sharing cars a viable option". She will test such hypotheses by collecting relevant information from a representative group of target customers. She will then evaluate the results obtained from the test against her theory, to ultimately reach a decision about whether to continue with the development of her idea as originally envisioned, make some changes to the original idea (i.e. pivot) or terminate the project, if she believes that no choice that she could implement would lead to achieve value under the scenarios that she expect to be more likely to occur.

2.3 A Scientific Approach to Decision Making: Debiasing and Learning Effects

Camuffo et al. (2020, 2021) show that entrepreneurs who employ a scientific approach terminate their projects more frequently and faster than other entrepreneurs, make more focused changes to their projects and perform better. Whereas these results are interesting, we still lack a precise understanding of the mechanisms underlying the relationship between the use of the approach and its outcomes (more frequent termination and superior performance) as well as its implications.

In this paper we fill in this important gap by focusing on a scientific approach to decision making and elaborating on two main mechanisms that explain the results presented by prior work. We suggest that, first, a scientific approach improves entrepreneurs' ability to develop a more objective and conservative assessment of the value of the business, reducing the impact of decision making biases such as, for instance, overconfidence (e.g., Chen et al., 2018). We call this the *debiasing effect* of the scientific method. The development of a general theory of the problem and its articulation into hypotheses helps scientific entrepreneurs focus on the relevant assumptions behind the business idea that need to hold for the value proposition to generate value, effectively leading to the formulation of more structured prior probabilities. This is complemented with the design of higher quality tests that can provide them with some objective signals about the extent to which their theory and hypotheses, and their priors more in general, are actually supported by data from the environment. Relating signals received back to the broad theory leads to a validation of the theory or to a rejection of it. This results in an update of their priors toward something more objective and to a more conservative expectation on the value of the idea. If this is too low, entrepreneurs will thus choose to terminate the project.

For instance, if the hypothetical entrepreneur we introduced in the previous section collected a negative signal on people's willingness to share cars due to hygiene concerns in a pandemic world, she would be more likely to form a negative value expectation and terminate the project. This effect is likely to lead scientific entrepreneurs to terminate their projects more often than non-scientific entrepreneurs, a result that is in line with qualitative research that suggests that factual grounding speeds decision making, reduces emotional conflict and facilitates deescalation of commitment (Eisenhardt, 1989, McDonald and Eisenhardt, 2020, Raffaelli, Glynn, and Tushman, 2019, Sleesman et al., 2018).

The second effect of a scientific approach that we suggest is that it improves the ability of scientific entrepreneurs to identify the changes to the business proposition would lead it to develop more or less value more easily or more rapidly. We call this the *learning effect* of the scientific method. The development of a theory and its articulation into hypotheses leads to a clear identification of the core elements or the problem and the relationships between them. This facilitates a quicker and more efficient search of the solution space, as it leads actors to identify ex-ante the characteristics of the solution (e.g., Camuffo et al., 2020, Felin, Kauffman, and Zenger, 2020). The subsequent test of hypotheses validates the theory, providing decision makers with feedback that are useful to improve the quality of their projects (e.g. Gross, 2017).

For example, if our entrepreneur obtained a positive signal on her hypotheses that car transportation in large cities is highly valued by households with young children who cannot use other types of transportation such as bikes easily, she would quickly understand that this will also make the service appealing to households that include the elderly and could pivot in this direction. This effect is likely to lead scientific entrepreneurs to perform better, conditional on the fact that they do not terminate their project.

To disentangle and identify both the *debiasing effect* and the *learning effect* of a scientific approach, in the next section we develop a structural decision-making framework and estimate it with a multi-equation simultaneous maximum-likelihood model.

3 Structural Framework

We start with a *value equation*. We consider the realized value (v) of a business idea and model it as:

$$v = a + \theta T + \sigma\epsilon \tag{1}$$

Where $a = \gamma X$, with X being a set of controls recorded at the baseline period. We assume that the realized value of the business idea is a function of a set of controls X and of whether the entrepreneur employs a scientific approach to decision making. Purposefully, the dummy T separates entrepreneurs employing a scientific approach from other entrepreneurs, hence θ identifies what we have labeled to be the *learning effect*.

Our model postulates that entrepreneurs explore their business ideas and form expectations of their potential value and probability of success over time. Let us denote these expectations by $\hat{v}$. We assume that entrepreneurs decide to keep developing their business if such expectations are higher than their outside option w. Therefore, in our framework, such estimations are crucial as the decision between continuing with the development of the business or terminating the project is based on the evaluation of that expectation with respect to an individual outside option.

In line with the RCT design that will be used to estimate the model, we assume that decision makers can be trained to use a scientific approach and our model compares those who will be trained with those who will not be trained (the control group in our RCT). Accordingly, we structure the entrepreneur's decision making process as characterized by four crucial points in time: (i) the baseline, before the training (0 - the *Baseline Evaluation*), (ii) after the training (E - the *Early Evaluation*), (iii) later in time after the training (L - the *Late Evaluation*), and (iv) at the time of the decision whether to remain active or terminate the business (F - the *Final Evaluation*). To clarify what we mean with the *Late Evaluation* period, it is worth explaining briefly the structure of our RCT, which monitors entrepreneurs over time, focusing in particular on whether entrepreneurs' project are still active at any point in time until an entrepreneur decides to terminate his/her project. We consider as *Late Evaluation* the last available data point before such decision. If an entrepreneur never terminates the project within the observation window, we consider as *Late Evaluation* the end of our observation window. Hence, we develop four equations:

$$\hat{v_0} = c + c_0 + \sigma_0 \epsilon \tag{2}$$

$$\hat{v_E} = c + c'_0 + c_E + (c_{ET} + \theta)T + \sigma_E \epsilon \tag{3}$$

$$\hat{v_L} = c + c'_0 + c_L + (c_{LT} + \theta)T + \sigma_L \epsilon \tag{4}$$

$$\hat{v_F} = c + c'_0 + c_F + (c_{FT} + \theta)T + \sigma_F \epsilon \tag{5}$$

The baseline evaluation (Eq. 2) happens before the training and therefore it depends on a series of factors independent of the training, such as education levels, age or previous startup experience, which we include in the vector c_0, whereas c represents a constant term.

Once the intervention starts, we assume it to have an effect on the evaluation. Equations relating the early (Eq. 3), late (Eq. 4) and final (Eq. 5) evaluations include c, which is the constant term, c'_0 that identifies constant idiosyncratic factors as above but that we assume could vary in terms of magnitude (as represented by the apostrophe), and c_j (where $j = E, L, F$), which identify contemporaneous factors affecting the value estimation.

In addition to this, we assume that the intervention has two different effects on the evaluation made by entrepreneurs, i.e.the *learning effect* θ and the *debiasing effect* c_{jT} (which is not restricted to be constant over time). They cannot directly be empirically identified without adding additional structure to our model. This is because we assume that the scientific approach helps entrepreneurs understand the opportunity for positive performance since the beginning of its application, but that its *debiasing effect* might vary overtime. Therefore, to achieve the goal of this paper of identifying these two effects, additional structure in our model is needed.

We first build on the previous steps, and generalize the decision-making process as:

$$\hat{v_j} = c + c_0' + c_j + (c_{jT} + \theta)T + \sigma_j \epsilon \geq w_j \tag{6}$$

With j representing the different time periods, and w_j representing the entrepreneur's outside option (which we assume vary over time and that we represent as w_j=w_{j-1}+b_j). This condition is verified if and only if:

$$\epsilon \geq \frac{w_j - c - c_0' - c_j - (c_j T + \theta)T}{\sigma_j} = z_j \tag{7}$$

We relabel the right hand side of equation (Eq. 7) z_j. When the decision of staying in the market is made (which we labeled with F), for entrepreneurs who choose to terminate their project, we cannot observe the values above. Rather, we only observe the final outcome. Hence, for F, we consider the following equation based on a latent model for the probability of remaining active in the market:

$$Pr(Stay) = \Phi(\frac{-w_F + c + c_0' + c_L + (c_{FT} + \theta)T}{\sigma_F}) \tag{8}$$

We can now re-arrange some equations to retrieve the structural parameters of interest. Let us rewrite Eq. 7 for the first three data points (0, E and L).

$$z_0 = \frac{w_0 - c - c_0}{\sigma_0} \tag{9}$$

$$z_E = \frac{w_E - c - c_0' - c_E - (c_{ET} + \theta)T}{\sigma_E} \tag{10}$$

$$z_L = \frac{w_L - c - c_0' - c_L - (c_{LT} + \theta)T}{\sigma_L} \tag{11}$$

Plugging Eq. 11 into Eq. 8, Eq. 10 into Eq. 11 and Eq. 9 into Eq. 10, we obtain:

$$z_E = \frac{\sigma_0 z_0 + b_E - (c_0' - c_0) - c_E - (c_{ET} + \theta)T}{\sigma_E} \tag{12}$$

$$z_L = \frac{\sigma_E z_E + b_L - (c_L - c_E) - (c_{LT} - c_{ET})T}{\sigma_L} \tag{13}$$

$$Pr(Stay) = \Phi(\frac{-\sigma_L z_L - b_F + (c_F - c_L) + (c_{FT} - c_{LT})T)}{\sigma_F}) \tag{14}$$

In the next section we describe the data and methodology used to estimate the model and the structural coefficients of interest.

4 Methodology and Data

4.1 Experimental Design

To estimate the structural framework we leverage on data from two field experiments, delivered in the context of a business support program that was offered to entrepreneurs in Milan and Turin (Italy). Both RCTs shared the same structure, type of intervention, and data collection process. The two RCTs were held asynchronously.

Both programs were advertised nationally over multiple offline and online channels. The advertisement campaign lasted for several weeks to ensure recruitment of at least 100 entrepreneurs per batch. The campaign promoted the program as a cutting edge business support program, offered free of charge to early stage entrepreneurs operating in any industry. The focus on early stage startups ensured that participants into the programs were highly involved in the decision making process. To apply, entrepreneurs were required to fill in an online survey and complete a telephone interview. In total, the data from the first RCT (Milan) includes 250 entrepreneurs, and the second (Turin) 132. [1]

Entrepreneurs were assigned to either a treatment or a control group through simple randomization. Considering both RCTs, we allocated 192 entrepreneurs in the control group and 190 in the treated group. We checked that the randomization was successful with a set of balance checks across groups (Tables A1 and A2 in the Appendix). Then, each group was broken down into smaller groups and assigned to an instructor, thus creating different classes of entrepreneurs. To avoid potential biases due to instructors' teaching style, each instructor was in charge of teaching to both one treated and one control classroom.

Entrepreneurs in both groups attended the same number of sessions. All the sessions were highly experiential and the division in small classes ensured that instructors provided feedback to each participant. Both groups of entrepreneurs were exposed to (1) general managerial frameworks (such as the balance scorecard or the Business Model Canvas), which would support a cognition-based approach to decision making, and (2) to data gathering techniques (such as interview techniques, surveys and A/B testing), which would support an action-based approach to decision making. The treatment group was taught to apply this content using a scientific approach, i.e. combining cognition and action and make decisions in a probabilistic way. Specifically, treated entrepreneurs learnt to develop

[1] Both experiments have been pre-registered on the AEA RCT Registry. Codes: AEARCTR-0002205 (Milan); AEARCTR-0006579 (Turin). Our registration indicated, respectively, 265 firms and 132 firms as the target. We eventually retained in our sample only 250 for RCT2 as some of the recruited firms eventually decided not to take part in the program. We observed firms for 18 observation points as planned.

a theory of the problem faced, to develop hypotheses that flow logically from it, and to use the evidence gathering techniques to test those hypotheses and relate the results back to the theory. For instance, in one of the first sessions, both groups were exposed to the Business Model Canvas (BMC), a widely used tool in entrepreneurship, which helps entrepreneurs graphically schematize a company's business model. Entrepreneurs in the control group were exposed to this method and simply invited to apply it to their business, as it typically happens in any class in which such a method is taught. Instead, treated entrepreneurs were taught to use the BMC as a starting point for the development of their theoretical conceptualization of the business idea. Each component of the BMC was translated in an hypotheses to be tested. Later on in the module, entrepreneurs were exposed to different testing designs. Entrepreneurs in the control group were generally encouraged to apply these techniques to the problems they were facing in their business. Treated entrepreneurs were explicitly encouraged, instead, to use those techniques to test the hypotheses developed in the previous sessions, to closely assess the results and compare the results with the theory originally envisioned.

Contamination between treatment groups was prevented by scheduling classes in different days or times of the week, according to the offered training. Moreover, all the communication was separated by treatment group and the research team checked whether applicants to the program knew other applicants, allocating them to the same experimental group.

4.2 Data Collection Process

We asked entrepreneurs to provide data on their decision-making processes and business performance throughout the training program for up to 66 weeks after the beginning of the training programs to a team of research assistants (RAs) via a set of phone interviews. RAs were purposefully trained by the research team and were responsible of conducting monthly telephone interviews with entrepreneurs. Overall, for each entrepreneur we collected the baseline and up to 18 data points.

Each phone interview was based on a standardized semi-structured interview script, including both open and closed-ended questions. Inquired topics included business performance, decision-making practices and any change introduced to the business idea. Each interview was recorded and stored in an encrypted storage, while RAs were also instructed to encode qualitative answers into quantitative information.

Each entrepreneur was interviewed up until the end of the project or up until the time they declared to have terminated the development of their business idea; thus, for firms that exited the market we have information only up to such exit decision (what we consider to be the *Late* data point). For firms that did not terminate before the end of our observation window, we have information up to 66 weeks after the beginning of the study.

4.3 Data and Estimation technique

We now describe the data employed in the empirical estimation of the structural model. To allow for a consistent estimation, we collapse our panel dataset into a cross-sectional form (N = 382) creating distinct variables corresponding to the three mentioned data points: the *Baseline* period before the training, the *Early* period, that is 8-weeks after the beginning of the training, and the *Late* period. The latter means that, for entrepreneurs that remained active in the market, we have the full set of information. Instead, for entrepreneurs that terminated their projects before to the end of the full observation window, we have information up to the data point prior to which they declared having terminated, which we treat as our "last available" data point.

To measure selection, i.e. entrepreneurs whose projects are still active at the end of our investigation period, we create a dummy variable that takes value 1 for entrepreneurs whose projects are still active and 0 for those that instead terminate their project at any point in time. For the former, we measure overall performance (or value) by computing the revenue growth between the first (baseline) and last available data point. We also check the robustness of results by computing the average of the revenue growths between each collected data point.

To measure entrepreneurs' perceived value or estimation of future value, we rely on survey and interview data recording two main components. First, we asked entrepreneurs to provide a predicted probability of termination at each of the three data points on a scale from 0 to 100. Second, we asked entrepreneurs to directly estimate the minimum and maximum potential future value of their business ideas, on a scale from 0 to 100. To compute the estimated value, we consider the average of the two (in logarithms).

Finally, as to capture idiosyncratic factors that could affect both the project value and entrepreneurs' estimations, we employ baseline data on team size (number of people in the founding team), team average age, weekly hours worked, years of experience with startups and the team-average education levels.

Table 1 includes some descriptive statistics about these variables by treatment group.

Table 1: Descriptive Statistics

	Scientific		Control		Total	
	Mean	SD	Mean	SD	Mean	SD
Revenue Growth (Stay = 1)	2.00	3.76	1.05	2.69	1.47	3.236
Average Revenue Growth (Stay = 1)	0.11	0.21	0.06	0.15	0.08	0.180
Stay (Dummy)	0.55	0.50	0.68	0.47	0.62	0.487
Probability of Termination (Baseline)	0.17	0.20	0.21	0.21	0.19	0.203
Probability of Termination (Week 8)	0.17	0.23	0.16	0.19	0.16	0.211
Probability of Termination (Last)	0.25	0.29	0.29	0.29	0.27	0.287
Estimated Value (Baseline - log)	4.16	0.29	4.12	0.30	4.14	0.293
Estimated Value (Week 8 - log)	4.06	0.40	4.11	0.30	4.09	0.355
Estimated Value (Last - log)	4.00	0.48	4.00	0.40	4.00	0.439
Startup Experience (Years)	1.34	3.12	1.20	2.32	1.27	2.746
Team Size (Baseline)	2.33	1.49	2.24	1.38	2.29	1.433
Education	1.91	0.79	1.99	0.91	1.95	0.851
Age	31.17	8.48	31.08	7.62	31.13	8.048
Hours Worked (Baseline)	13.24	19.27	12.91	19.28	13.07	19.25

Descriptive statistics on both baseline characteristics and outcomes, by treatment group. $N = 382$. For balance checks and related tests, please refer to Tables A1 and A2 in the Appendix.

By assuming a cumulative normal distribution, we can estimate the value of z_j in our model by simply calculating the inverse of the latter given the predicted probabilities of termination (p_j). Mathematically, since $p_j = \Phi(\frac{w_j - c - c'_0 - (c_{jT} + \theta)T}{\sigma_j}) = \Phi(z_j)$, we can retrieve z_j as:

$$z_j = \Phi^{-1}(p_j) \tag{15}$$

We cannot know the z estimate at the exact time in which the decision has been taken (what we labelled with F). We, therefore, employ a selection model, where we include as our selection variable the estimate z_L, and we rely on a latent estimation for such probability.

If we were to only estimate the first two equations of the structural model described above, this could be done with a standard Heckman model where the exclusion restriction would be satisfied by the inclusion in the selection equation of the estimate z_L from Eq. 15 evaluated in the *late* period. This would allow us to estimate the *learning effect* θ conditional on the decision to stay in the market. However, relying solely on such two equations does not allow us to estimate the *debiasing effect*.

The opportunity to leverage data on the entrepreneurs' estimation of the potential value of their ideas enables us to retrieve all the structural parameters of interest and be able to separate the *debiasing* effect from the *learning effect*. Particularly, we leverage on the first two post-training data points (E and L) and consider such predicted values for two additional equations, that we label with *. It is the availability of the own estimations by entrepreneurs that allow to estimate empirically both Eq. 3* and Eq. 4* and ultimately retrieve the two variances σ_E and σ_L that allow us to estimate the variance σ_F from Eq. 8. This additional step is what allows us to identify the *debiasing effect* in the three different data points we are considering. By estimating the three variances, we are able to subtract θ from the estimated coefficients on T in Eq. 14 and Eq. 12 and finally compute the debiasing effect for Eq. 13.

We thus end up with the following structural model to be estimated, made up of six equations:

$$v = a + \theta T + \sigma\epsilon \tag{1}$$

$$Pr(Stay) = \Phi(\frac{-\sigma_L z_L - b_F + (c_F - c_L) + (c_{FT} - c_{LT})T}{\sigma_F}) \tag{14}$$

$$z_L = \frac{\sigma_E z_E + b_L - (c_L - c_E) - (c_{LT} - c_{ET})T}{\sigma_L} \tag{13}$$

$$z_E = \frac{\sigma_0 z_0 + b_E - (c_0' - c_0) - c_E - (c_{ET} + \theta)T}{\sigma_E} \tag{12}$$

$$\hat{v_L^*} = c + c_0' + c_L + (c_{LT} + \theta)T + \sigma_L\epsilon \tag{4*}$$

$$\hat{v_E^*} = c + c_0' + c_E + (c_{ET} + \theta)T + \sigma_E\epsilon \tag{3*}$$

We estimate these equations through a multi-equation conditional mixed-process estimator using the `cmp` user-written command in STATA 16 (Roodman, 2011). The fitted algorithm is a modified version of a seemingly unrelated regressions estimator. In other words, it employs a maximum likelihood (ML) estimator with the assumption that the errors from the different, independent, equations are distributed according to a joint normal distribution. The `cmp` estimator allows us to model a simultaneous equation framework where endogenous variables in a multi-staged process appear both on the right and left end sides of six empirical equations representing the structural model described in the previous subsection. We estimate the following set of empirical equations, linked to the structural equations above:

Eq. 1 : $v^* = \alpha_v + \beta_v X + \theta T + D + \epsilon_v$

Eq. 14 : $\Phi(\alpha_F + \gamma_F z_L + \beta_F T + D)$

Eq. 13 : $z_L = \alpha_L + \gamma_L z_E + \beta_L T + +\epsilon_L$

Eq. 12 : $z_E = \alpha_E + \gamma_E z_0 + \beta_E T + \lambda_E X + D + \epsilon_E$

Eq. 4* : $v_L^* = \alpha_{v_L} + \beta_{v_L} T + \lambda_{v_L} X + D + \epsilon_{v_L}$

Eq. 3* : $v_E^* = \alpha_{v_E} + \beta_{v_E} T + \lambda_{v_E} X + D + \epsilon_{v_E}$

Where D is a set of dummies for RCT and class instructors [2], X is a set of controls recorded at the baseline period as described above and the α represent constant terms of each equation. All the equations are linearly estimated, but the selection one (Eq. 14) which follows a probit model. Again, equations are estimated simultaneously assuming a joint normal distribution of the error terms. Since the intervention is administered at the classroom level, we cluster standard errors by classroom.

From the estimated coefficients of the above regressions, we can thus retrieve all the parameters of interest that belong to our theoretical structural model. Specifically, the *learning effect* is straightforwardly estimated from the first equation, and it is the coefficient θ on the treatment dummy computed from the first model. All the other structural coefficient have instead to be computed leveraging on the estimated variances and coefficients from the econometric models. Particularly, the computation entails a non-linear combination of different estimated parameters. We conduct such computation using the `nlcom` routine on STATA.

Retrieving the OLS variances from Eq. 3* and Eq. 4*, we can estimate the variance of the model related to the decision (selection equation) from Eq. 14 and we compute all the structural coefficient related to the *debiasing effect* at different points in time from the other equations. Recall that in all equations but the value one, the estimated coefficient on the treatment dummy captures both the hypothesized effects. Thanks to the estimation of variances, we can subtract the estimated *learning effect* (θ) from such coefficients and finally retrieve the correct estimation for the *debiasing effect*. Table 2 details the calculations.

[2]A total of 7 instructors taught in RCT 2, while 5 instructors taught in RCT 3. Four of them taught in both RCTs, thus allowing us to control for both RCT and instructors' effects in regressions.

Table 2: Structural Parameter Computation

Parameter	Computation	Equations
θ	θ	1
σ_E	*OLS variance*	3*
σ_L	*OLS variance*	4*
σ_F	$-\frac{\sigma_L}{\lambda F}$	14, 4*
c_{ET}	$\beta_{v_E} - \theta$	3*, 1
c_{LT}	$\beta_{v_L} - \theta$	4*, 1
c_{FT}	$\beta_F \sigma_F + c_{LT}$	14, 4*

This computation strategy leverages on the straightforward calculations from Eq, 3* and Eq. 4*. An alternative computation strategy is shown in the Appendix.

Before computing the full-fledged structural estimation, we also estimate a three-step extended selection model where we only leverage on the entrepreneurs' prediction of idea value rather than on both the latter and the predicted probability of termination.

First, we account for selection by running a simple Heckman selection model using the predictions at the last available data point to identify the selection equation, always controlling for baseline characteristics. This step is a different way of modelling the first two equations in the full structural model. Differently from that, we directly employ the prediction on the idea value rather than relying on the perceived probability of termination. However, since our theory explains how entrepreneurs' value estimations should be a function of the treatment, such identification might be subject to endogeneity. We thus add a third equation, instrumenting the late predictions with the baseline, exogenous, prediction of idea value. In the final step, to fully disentangle the *debiasing effect* of the treatment, we add a fourth equation introducing entrepreneurs' value estimations in the *early* period, thus setting up a recursive instrumentation structure. To run this stepwise estimation, we also rely on the `cmp` command in STATA.

4.4 Attrition

As it is common in field experiments and in experimental designs with multiple post-treatment periods, experimental units drop out from the study before its natural end, leading to attrition biases (Ghanem, Davis, and Hirshleifer, n.d.). To prevent such problem, we designed a series of monthly events focused on entrepreneurial challenges and issues. These events were offered to both treatment groups and did not include any additional manipulation. We offered these events free of charge to ensure the highest participation rates. The only requirement for attending the events was to participate in the data collection phases.

Across the two RCTs, 10% of the entrepreneurs decided to drop out from the program without having terminated the development of their business ideas. This attrition rate is similar between treatment groups: 12.5% for the control group and 7.4% for the treated group, the difference being significant only at the 10% level (chi-2 test, $p = 0.09$).

In the main analyses that we performed, we input missing values of the "attritors" using their last available data point, considering them as entrepreneurs still active on the market. This relies on the the conservative assumption that both the performance and the perception of these entrepreneurs did not change after they left the program.
Nevertheless, in the Appendix we show tests for selective attrition, comparing baseline characteristics of "attritors" with ventures staying in the market, by treatment group. Moreover, as a robustness check, we re-estimate all the models only considering compliant units, reducing the sample size to 344 observations. Results are available in the Appendix and are fully consistent with the main estimations.

5 Results

5.1 Extended Selection Model

Table 3 reports the results of the four-equations extended selection model. In the Appendix we also report the results of the first two steps described in Section 4.3.

Table 3: Extended Selection Model

	(1) Value Equation	(2) Selection Equation	(3) $\hat{v}_L$	(4) $\hat{v}_E$
Intervention	1.030**	-0.280**	0.043	-0.059**
	(0.440)	(0.120)	(0.043)	(0.02)
$\hat{v}_L$		1.875***		
		(0.320)		
$\hat{v}_E$			1.102***	
			(0.360)	
$\hat{v}_0$				0.256***
				(0.069)
Startup Experience	0.214***	-0.029	0.0045	0.005
	(0.073)	(0.025)	(0.008)	(0.005)
Team Size (Baseline)	0.310*	-0.059	0.011	0.002
	(0.169)	(0.039)	(0.013)	(0.014)
Education	0.269	-0.008	-0.007	-0.035*
	(0.237)	(0.074)	(0.022)	(0.0204)
Age	-0.085***	0.015*	0.003	0.003
	(0.024)	(0.009)	(0.003)	(0.002)
Hours Worked (Baseline)	0.009	0.006	-0.002	0.002**
	(0.013)	(0.005)	(0.002)	(0.001)
Constant	1.745**	-7.462***	-0.652	3.103***
	(0.687)	(1.411)	(1.503)	(0.317)
Correlation	-0.198***			
	(0.075)			
RCT Dummies	Yes	Yes	Yes	Yes
Mentor Dummies	Yes	Yes	Yes	Yes
N	382			

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

The value equation is estimated through an OLS conditioned on the selection equation, estimated through a *probit* model. The last two equations are estimated through OLS. Standard errors clustered at the classroom level ($K = 24$).

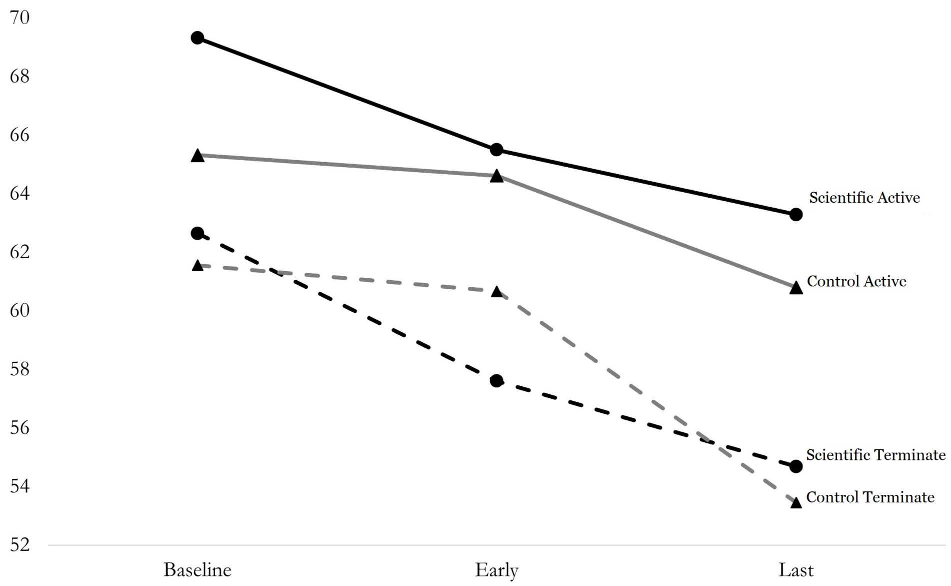
The coefficient on the intervention dummy from the value equation signals that entrepreneurs following a scientific approach have a revenue growth with respect to the baseline of 103 pp higher than those of entrepreneurs in the control group (the *learning effect*). The large magnitude of the coefficient could be driven by few firms experience a sizeable revenue growth over time. Thus, in the Appendix, we show results of the same model estimated on revenue growth trimmed at the 99th percentile, leading to a drop of three potential outliers. Results are consistent but more conservative: the *learning effect* is estimated to be around 0.75 (i.e. firms in the treated group have a revenue growth that is 75 pp higher).

We now turn to the results regarding selection. In the model presented in Column (2),the dependent variable is a dummy taking value 1 if the project is still active at the end of the observation window. The negative coefficient of the intervention dummy signals that treated entrepreneurs are more likely to terminate their projects, a result consistent with Camuffo et al., 2020. The significant correlation ($\rho = -0.198$) between these first two equations signals the necessity of controlling for selection when analyzing performance. The interpretation of the negative correlation aligns with our theoretical expectations that entrepreneurs tend to overestimate the value of their ideas, as it implies that entrepreneurs with higher perceived evaluations are associated with lower realized value.

Results reported in Columns 3 and 4 in Table 3 allow us to observe the *debiasing* effects at the *Late* and *Early* evaluation data points. The negative and significant coefficient of the intervention dummy in the *Early* equation ($\beta = -0.059, SE = 0.027$), indicates that scientific entrepreneurs decrease their value estimates already at the early stage of evaluation. Instead, the non-significant coefficient of the intervention dummy in the *Late* equation ($\beta = 0.043, SE = 0.043$) signals that the *debiasing effect* is mostly found at earlier stages of evaluations rather than at later stages. These results are not surprising if we think that in the *Late* period, all entrepreneurs regardless of their treatment status should make better estimations of the value of their ideas. At the baseline, the value of the average idea is the same between the two groups, thus in principle we should see no differences in the average self-evaluation if the treatment had no effect. Nevertheless, the positive sign of the coefficient might also signal that treated entrepreneurs are taking into account the positive *learning effect* in their predictions, given the early *debiasing effect*. However, this intuition can only be confirmed with the structural model, that fully disentangles the two effects.

To provide additional evidence in support of this early-stage *debiasing effect*, we directly look at the estimates made by entrepreneurs on the potential value of their ideas at different points in time (0-100 scale). We compare the averages of entrepreneurs estimation across four groups defined by two dimensions: whether the entrepreneur belonged to the treatment versus control group and whether they terminated the project within the observation window. We report these metrics in Figure 1 below.

Figure 1: Entrepreneurs' Evaluations



Entrepreneurs' estimation of their idea value over time (scale: 0-100). We show the data for the three main datapoints: baseline (pre-intervention), early (8 weeks from the first lecture of the training), last (last available observation in our dataset). We consider four groups, according to the treatment group and the final decision (terminate or stay) made by the entrepreneur.

Figure 1 shows a set of interesting patterns. First, projects that were not terminated show higher estimation value than those that were terminated, in line with the idea that entrepreneurs, on average, terminate the projects that they assess to have a lower value. Second, for all groups, value estimates are progressively lower over time. This is in line with the idea that, as entrepreneurs collect more information they revise their estimate of the value of an idea accordingly, correcting an initial overestimation that seems to affect all entrepreneurs. Third, the figure also shows that - interestingly - the path of reduction is different for treated and control entrepreneurs. In line with the econometric results, treated entrepreneurs reduced their own estimation already at the very early stages, both in the case in which their final decision was to stay in or exit from the market; instead, the estimates of control entrepreneurs tend to be constant between the baseline and the early period, and are reduced only later in the process.

We now turn to the full-fledged structural estimation, to clearly disentangle the *debiasing* and *learning* effects.

5.2 Structural Estimation Results

Table 4 reports the results of the structural estimation.

Table 4: Estimated Structural Parameters

	Parameter	Std. Err	z-score
θ	1.10	0.455	2.41
σ_E	0.34	0.032	10.72
σ_L	0.43	0.039	10.83
σ_F	1.57	1.088	1.44
c_{ET}	-1.15	0.454	-2.51
c_{LT}	-1.11	0.441	-2.51
c_{FT}	-1.82	0.412	-4.41

Structural parameters retrieved after ML estimation of the six equations described in Section 4.3. Parameters and their standard errors are retrieved using the `nlcom` routine in Stata 16. All estimated equations include dummies for RCT and instructor, with standard errors clustered at the classroom level ($K = 24$). Full estimation results of the six equations are reported in the Appendix.

Our estimation results show a positive and significant θ coefficient (1.10), that represents what we called the *learning effect.* On average, "scientific entrepreneurs" experience an increase in revenues of around 110 percentage points compared to traditional ones, conditional on the decision to stay on the market. This result is in line with the one found by the extended selection model described in section 5.1. As before, we also run a robustness check with the trimmed version of the dependent variable. Results, available in the Appendix, show a more conservative - but still positive - estimate of the effect (0.82).

The three variance estimates, σ_E, σ_L, σ_F grow in magnitude as the final decision is taken. The variance related to the decision equation is around five times the variance experienced in earlier stages (1.57 vs 0.34), coherently with the idea that at the decision moment there is a larger variance of project's perceived value, leading to the termination of less valuable projects.

Parameters c_{ET}, c_{LT} and c_{FT} are those identifying the *debiasing effect* in the three decision periods that we consider in our structural framework. Differently from the extended selection model estimation, we leverage on the structural model to clearly separate them from the *learning effect.* Results show that treated entrepreneurs are more likely to reduce their estimate of the value of the business ideas at each of the three stages that we consider. Particularly, and coherently with the results of the extended selection model, the *debiasing effect* materializes already at an early stage, i.e. eight weeks after the beginning of the training. This effect is persistent over time, signalling a more conservative approach of treated entrepreneurs when estimating the future value of their ideas.

In the Appendix, we also report the results of the six equations estimated via `cmp` and an alternative computation of structural parameters. To test the robustness of the following results, we also run a number of checks other than the mentioned one leveraging on a trimmed version of the revenue growth variable and the other removing "attritors" from the sample. In our first exercise, we include additional controls in all empirical equations to take into account of some imbalances between groups prior to the training. Then, we employ an alternative measurement for z, where - instead of considering the last available data point - we considered the previous one. Finally, we also conduct the same analyses using the average revenue growth over time rather than the simpler revenue growth with respect to the baseline revenues. Results, reported in the Appendix, are all consistent with our main analyses.

Overall, these results suggest that treated entrepreneurs following a scientific approach perform better, on average, compared to traditional entrepreneurs, even when taking into account the effect of selection (the decision to not terminate the project). They also show that they tend to make a downward adjustment to their estimation of their business' ideas values. This downward adjustment on the potential future value of their business ideas explains the higher rate of projects' termination by "scientific entrepreneurs" shown in Camuffo et al. (2020). Interestingly, this effect shows that - despite the positive *learning effect* that would allow treated entrepreneurs to perform better on average- scientific entrepreneurs choose to terminate as a result of a reduction in the value of their expectations.

The fact that treated entrepreneurs tend to be more cautious can have positive implications, since we can expect that many terminated ideas would not have been successful if still active. A positive selection can leads to resource savings, both in terms of time and money. But how does this beneficial effect compares with the termination of potentially good projects? The following section discusses this topic in depth.

6 The Trade-Off between Retaining and Discarding Ideas

In the previous section, we have shown that treated entrepreneurs are more conservative when evaluating their ideas' value and that they reach a more cautious evaluation more quickly than control entrepreneurs. This result is in line with the idea that entrepreneurs following a scientific approach reach a more realistic evaluation of the idea more quickly, and more quickly identify "falsely positive ideas".

However, we cannot exclude the possibility that the treatment rather reduces the confidence of entrepreneurs, leading them to discard truly good ideas and increasing - in other words - the number of falsely negative ideas that they terminate. This is a very important question, but one that is, nevertheless, not trivial to answer. Answering this question would require knowing what could have been the "true" realized value of terminated projects, were they not terminated; this is clearly not possible.

We provide four different suggestive pieces of evidence that can at least partially address this question. First, we leverage additional data from our RCT and we conduct a follow-up data collection regarding the performance of the entrepreneurs after a longer time horizon. Second, we gather an external evaluation of entrepreneurs' ideas from leading professionals, leveraging on idea pitches collected at the baseline period. Third, we further elaborate the data of the structural estimation to derive additional insights. Fourth, we conduct a follow-up experiment using a business simulation game played by MSc students.

6.1 Additional RCT Evidence

An alternative way to obtain a relatively objective assessment of the value of ideas is to examine whether firms have received financing from external investors over time. In our data collection efforts, we asked entrepreneurs whether they received external financing at any data point (for instance, from venture capitalists or business angels). We create a dummy taking value 1 if the firm has received financing within the observation period, and 0 otherwise. In Table 5 we report, for each cell, the share of firms having received external financing, separated by intervention (treatment vs. control) and final decision (termination vs. being active in the market).

Table 5: Share of Firms Having Received External Financing

	Terminate	Active	Difference
Control	2%	10%	-8%
Scientific	1%	20%	-19%
Difference	1%	-10%	11%

Share of firms having received any type of external financing during the RCTs observational window.

Looking at results for the treatment group, only 1% (1 firm out of 86) of firms that terminated the project obtained external finance before their decision to terminate. This share corresponds, instead, to 2% (1 firm out of 61) for those in the control group. This goes in the direction of suggesting that ideas terminated by treated entrepreneurs are not better than those terminated by control entrepreneurs. Conversely, looking at entrepreneurs that decided to keep their projects active, we see that 20% (21 firms out of 104) of treated entrepreneurs received external financing, which corresponds to two times the 10% (13 firms out of 131) recorded in the "control" group[3].

These numbers are also consistent with the intuition that projects retained by scientific entrepreneurs tend to be of higher quality, with a selection of false positive projects taking place. Projects that were kept active were on average more appreciated by external

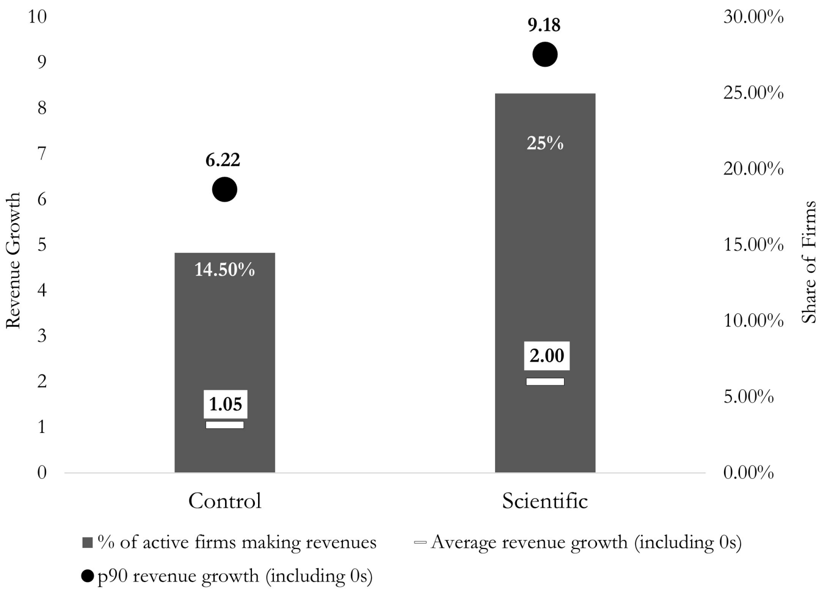
[3]We also run a simple linear probability model, regressing the financing dummy on the interaction between the dummy for staying in the market and the intervention. We add as controls RCT and mentor dummies, clustering the standard errors by classroom. The coefficient on the interaction (i.e. the Difference-in-Differences coefficient) is significant at the 10% level.

investors, who were of course blind with respect to the treatment. This is a strong signal towards our theory that the "scientific approach" helps selecting the best projects ex-ante.

Second, we provide some additional evidence on the distribution of revenues across treatment groups and termination decision, since it is also at the backbone of our empirical estimation.
It is worth noting that in both RCTs all firms started with no revenues, thus explaining the sizeable magnitude of the *learning effect*. Again, we compute the growth in revenues as the difference in logged revenues between the last available data point and the baseline, adding 1 to the latter value as to compute logs for the 0s.

The distribution of revenue growth at the end of the observation window is very skewed, with few firms having positive values. To further explore this phenomenon, we created a dummy variable for firms still active in the market at the end of the observation period, taking value 1 whether a firm shows a positive revenue growth. Figure 2 summarizes the share of firms making revenues together with two key moments of the revenue growth's distribution.

Figure 2: Additional Evidence on Revenue Growth



The columns indicate the share of firms with positive revenue growth, conditional on their decision to stay operational (right axis). The white bar and the black dot indicate, respectively, the 90th percentile and the average of the distribution of revenue growth (including 0s), conditional on the decision to stay operational (left axis).

More precisely, only 14.5% of firms still active on the market in the control group (namely, 19/131) made revenues, versus the 25% (26/104) of those in the scientific group[4]. Looking

[4]We also run a Probit with a Heckman selection model (`heckprobit`) using as a dependent variable the dummy recording positive revenues. The fitted model mimics the one run in the last two steps of the full structural model, using z_L as the selection variable. Results are in line with the intuition that the probability of making revenues conditional on the decision to stay in the market is significantly higher

at the average revenue growth of all operational firms, including those with no revenue growth, Figure 2 shows an higher average revenue growth for scientific entrepreneurs. Whereas the medians for both groups are set to 0, the Figure shows that the value of 90th percentile is higher for scientific entrepreneurs.

This evidence brings further support to the results of our econometric and structural estimations, reinforcing the idea that scientific entrepreneurs make less false positives. Ideas that have been selected by scientific entrepreneurs, despite being fewer, have not only average higher revenues but also a higher chance of reaching the revenue stage

However, since we are also interested in what happens on the false-negative side, we follow the logic from Elfenbein, Knott, and Croson, 2017 [5] and look at the revenue pattern of the firms divided ex-post by their final decision on whether to terminate or stay active on the market.

First, we leverage on the panel structure of our database and look at data on the revenue growth over time. Specifically, for each firm in the sample, we computed its cumulative revenue growth from the baseline to each observation in our panel. For firms that decide to remain active in the market, we expect a growing trend. For firms that terminate, we expect a more noisy pattern, as their revenue growth naturally goes to zero after their decision to terminate the project (and we conservatively set them as missing values in our database). We then compute the average by treatment group and the final termination decision. Figure 3 shows the results of these computations. The figure shows that, looking at firms that remain active, those in the scientific group perform better, in line with previous findings.

Instead, what is more interesting is that firms in the scientific group that terminated their projects did made some revenues, although these revenues were lower than the ones of firms that stayed at the very same point in time. This is a first signal that, on average, ideas that were discarded performed less well than those that were not discarded, at least up to their termination decision.

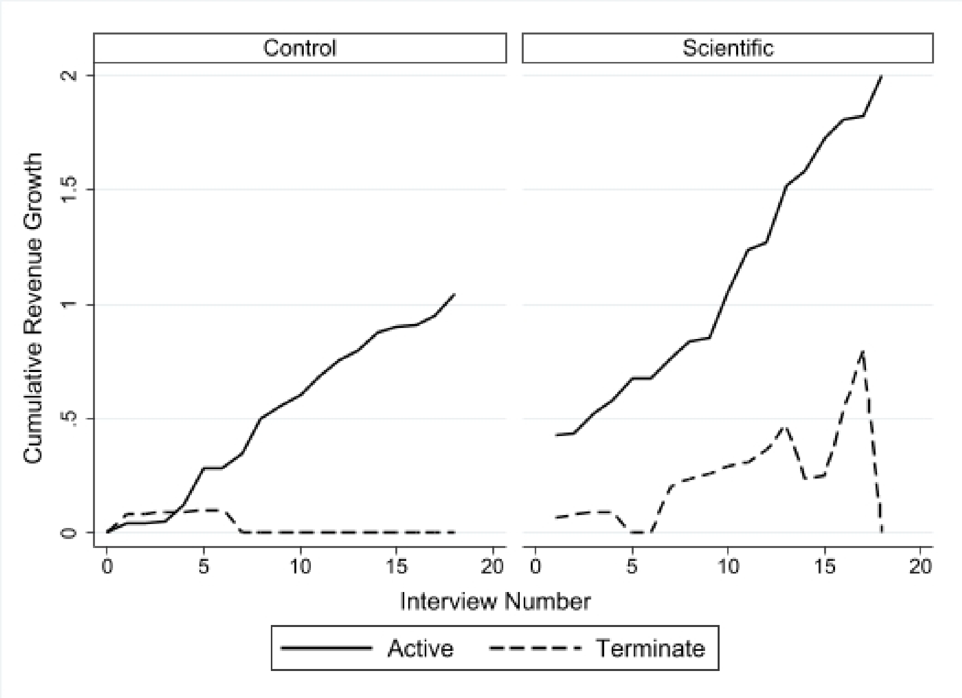
When looking at the control group, our model suggests that the share of firms remaining active in the market is likely to include false positives, but it could also include projects that the treatment group would have discarded as false negatives. However, the facts that 1) the revenue growth of scientific firms that are still active in the market is higher than that of control firms, and 2) that the revenue growth for scientific firms that terminated

for scientific entrepreneurs. We also run simpler tests (probit, t-test and chi2 test) on the subsample of active firms, thus not accounting for selection. All tests lead to a 5% significant difference between groups.

[5] Elfenbein, Knott, and Croson, 2017 run an experiment where they ask participants to deduce their firm type and optimal exit time. Firms can be of two types: high-profit and low-profit. Participants infer the firm type by looking profit streams from period to period, and decide according to the information disclosed at each round of the game. By analogy, in our paper we consider firms that terminated as "low-profit" ones, and firms that are still active as "high-profit" ones and look at the revenue pattern at each of the 18 data points we collected.

is on average always lower than the one of control firms that stayed, suggests that overall the reduction in false positives compensates the potential increase in false negatives experienced by the treatment group.

Figure 3: Panel Data on Revenue Growth



The graph show the pattern of average revenue growth by treatment and final decision of staying or not in the market. For firms that exited, the value of revenue growth is set as missing after their decision to terminate, explaining the noisier pattern.).

A limitation of these findings is that they only enable us to assess the performance of firms during the limited time window of observation covered by our experiment. To address this limitation, we conducted an additional data collection effort in the months of February and March 2022, which is respectively five and four years from the beginning of the first and second RCTs to determine the survival of firms in our sample. The survival rate of startups is notoriously low[6], hence survival can be considered an alternative measure of performance. We recruited a research assistant (RA) and provided him with a list comprising the company name and the founder name for the 382 firms in our sample. The RA was blind to the type of intervention the entrepreneur went through during the training program. We asked the RA to search online for information about the founder and the company, to understand whether the firm was still active. The search was performed on standard research engines (e.g. Google), on LinkedIn and on the official Italian chamber of commerce firm registry. We considered a company as active if clear references to its activities were found online. These references include, among others, the existence of a website, clear references to the company in the founders' LinkedIn profiles and activities,

[6]"Startup, dopo cinque anni ne sopravvive solo una". *Il Sole 24 Ore*, 12 October 2017. Link: https://www.ilsole24ore.com/art/startup-5-anni-ne-sopravvive-su-due–AEB7RAmC

registration to the Chamber of Commerce or recent press coverage. This data collection led to the identification of 67 active companies out of 382 firms (17.5%) as active in March 2022. [7].

When splitting the sample between the two treatment conditions, we find a termination trend consistent with the expectations from our previous analyses. Of these 67 operational firms, 41 belong to the treated group, while 26 belong to the control group. This implies that around 22% of "scientific" firms are still operational versus a 14% of "control" ones[8].

This pattern supports the idea of both a *debiasing* and *learning* effects of the scientific approach taking place. Entrepreneurs following such methodology were able to understand earlier the nature of their projects, terminating potential bad ones before control entrepreneurs whose termination rate kept increasing over time. Looking at this long-term survival rates, chances of being still operational are indeed higher for treated companies. This could mean that treated firms, when compared to control ones, were better able to develop their businesses given the initial stricter selection, bringing further support to the presence of these two intertwined effects.

6.2 External Evaluation of Entrepreneurial Ideas

We partnered with a leading Italian company operating in the innovation and entrepreneurship landscape to obtain an external evaluation on the value of the entrepreneurs' ideas by experienced professionals. The logic behind this choice is that of obtaining an objective evaluation of the ideas that were discarded as well as retained in both groups. We asked the company to evaluate the pitches submitted by entrepreneurs at the baseline, that is, before the start of the training [9]. We assume that the baseline idea is a good proxy of the potential success and future value of the business idea.

The evaluation has been made on three main elements, on a scale from 0 to 100 for each of them: 1) *Profitability Potential*; whether the idea can turn out to be a huge commercial success; 2) *Innovativeness*: whether the idea contains significant innovations;

[7]The number of companies identified as active in this data collection is lower than the expected number based on official statistics, which indicate a 50% likelihood of survival after 5 years Link: https://www.ilsole24ore.com/art/startup-5-anni-ne-sopravvive-su-due–AEB7RAmC). However, it must be considered that these statistics refer to established and registered companies rather than ideas at the very early stage of development as those participating to the two RCTs. In addition, we are only able to classify as active those firms that have an online presence. This implies that we cannot know whether a project is still in a pre-launch phase or whether it has developed under a different name than the one recorded in our database. However, we believe that this analysis still provides some relevant insight on the medium term performance of firms in the two groups

[8]We tested this difference using a two-tailed t-test and also LPM/Probit models with additional controls. In all specifications, the difference is statistically significant at a 5% significance level.

[9]We used 220 pitches for the RCT conducted in Milan, and 110 pitches for the RCT conducted in Turin. The missing pitches were not available due to corrupted data in our storage space. We checked whether the firms for which we do not have the pitch were systematically different from the others, finding no significant differences at the baseline on the variables used in the main analyses. Our final sample included 330 pitches, of which 167 in the control group and 163 in the treatment group. Balance checks still hold for this subsample of firms, meaning that the absence of the pitch is likely a random occurrence.

3) *Feasibility*: whether the idea is realistic and possible to be realized. We then average these three scores to create a *"composite" expert evaluation score* ranging from 0 to 100. In the following analyses, we also look at the *profitability* score alone, as it is the one which is more directly comparable to the potential monetary value of the idea.

The empirical results discussed in the previous sections show that even if a smaller number of treated entrepreneurs remains active in the market, these firms perform better. This is in line with the idea that, by exerting a higher selection, the *scientific approach* reduces the rate of false positives. To ensure that this beneficial reduction in the false positive rate is not more than compensated by a stronger increase in the false negative rate, we would like to see that the *expert evaluation* for treated ideas that terminated is not different from those that terminated in the control group. Table 6 reports the averages for both the *expert evaluation* and *profitability* scores, by treatment group and termination decision.

Table 6: Expert Evaluation Scores

	Composite Expert Score		Profitability Score	
	Terminate	Active	Terminate	Active
Control	40.43 (*22.59*)	36.87 (*21.53*)	41.11 (*25.67*)	35.49 (*25.11*)
Scientific	37.21 (*20.73*)	36.61 (*20.22*)	36.71 (*24.67*)	38.22 (*24.52*)

The Table reports the group averages of the expert evaluations scores. Experts were asked to evaluate each idea on three dimensions, on a scale from 1 to 100: 1) *Profitability Potential*; whether the idea can turn out to be a huge commercial success; 2) *Innovativeness*: whether the idea contains significant innovations; 3) *Feasibility*: whether the idea is realistic and possible to be realized. The "Composite Expert Score" refers to the average between the three items. The "Profitability Score" instead refers to the first item alone.
Standard deviations in parentheses. *N = 330; Control = 167; Treated = 163.* There are no significant differences at conventional levels between groups.

Notably, expert evaluations are significantly lower than the self-evaluations made by entrepreneurs on the same 0-100 scale displayed in Figure 1. This reinforce our initial intuition that entrepreneurs tend to overestimate their ideas' potential at the baseline and that correction in the entrepreneurs self-assessment that we observe overtime is due to a *debiasing effect* induced by the *scientific approach*. Coming back to the trade-off between the *false positive* and *false negative* rates, this data shows how treated firms that decided to terminate do not have an expert evaluation that is significantly higher than those in the control group. This evidence, therefore, does not seem to suggest that scientific entrepreneurs discard a higher rate of false negative projects compared to the control group. In fact, from a qualitative point of view, it even seems that ideas in the control group that terminated have the highest scores. Combined with the evidence coming from the regressions, these results suggest that the selection induced by the *scientific approach* has a net "positive" effect. [10]

6.3 Structural Estimation: Additional Insights

To further support the insights drawn from descriptive data, we go back to our structural model and focus on the first two equations, estimated with linear regression for the *value equation* (Eq. 1) and with a probit model for the selection equation (Eq. 14). We retrieve the correlation coefficient ρ between the two equations and the Mills' ratios from the selection equation for firms that stayed and terminate their projects.

Using the previously computed variances of the *value equation* (σ) and of the selection equation (σ_F), we can thus compute, for each entrepreneur in our sample, the expected value of the correction in the value equation by treatment condition and by decision to terminate the project or not. Mathematically, for firms that stayed in the market, this corresponds to:

$$correction = \rho \times \frac{\sigma}{\sigma_F} \times \frac{\phi(x\beta)}{\Phi(x\beta)} \tag{17}$$

Instead, for entrepreneurs that terminated, this corresponds to:

$$correction = -\rho \times \frac{\sigma}{\sigma_F} \times \frac{\phi(x\beta)}{(1 - \Phi(x\beta))} \tag{18}$$

The intuition behind this analysis is that the correction provides us with a measure of the extent to which the value of ideas needs to be adjusted due to the selection. A positive value of the correction suggests that entrepreneurs using a scientific approach underestimated the value of the project; a negative sign suggests that scientific entrepreneurs overestimate the value of the project. A positive difference in the correction between those who terminate and those who stay suggests that the underestimation of those who terminate is higher than the overestimation of those who stay. We are interested in the difference between control and treated entrepreneurs.

We compute these differences in Table 7, were we make the conservative assumption that the value model for entrepreneurs who remained active and those who terminate their projects is identical for entrepreneurs who terminate and those who stay. We call this the *lower bound* condition.

[10]We checked the robustness of the expert evaluations by comparing the averaged evaluation score between two groups of firms: those receiving external financing during the RCTs observational window (N = 32) and those not receiving any external financing ($N = 298$), as in Section 6.1. We find that the average (median) score of the former group is of 40.4 (38), compared to an average (median) for the latter group of 37.1 (33.3). Despite not being a significant difference at conventional levels, the qualitative evidence points towards a reliable evaluation made by the experts, who evaluated with higher scores project that were indeed funded during the RCTs periods. Evaluators were blind to any outcome/characteristics/treatment of the evaluated projects, including their financing status.

Table 7: Same Value Model for Terminate and Stay

	Terminate	Stay	Difference
Control	0.55 *(0.13)*	-0.26 *(0.11)*	0.805
Scientific	0.43 *(0.14)*	-0.35 *(0.12)*	0.780
Difference	0.125	0.099	0.025

Standard Deviation in Parentheses

The negative ρ coefficient estimated through the structural model leads to a negative correction for firms that stayed in the market. While it can be challenging to interpret such coefficient in the light of the Heckman selection model, the negative direction signals that entrepreneurs on average tend to overestimate the value of their ideas resulting in a negative correlation when looking at realized performance. Such effect moreover could be mostly driven by the weakest bias reduction provided by the control group, given the results from our structural estimation for the treated entrepreneurs. Importantly, the difference-in-difference calculation leads to a number close to zero and not statistically significant (0.015). This suggests that there is not a significant difference between treated and control entrepreneurs when it comes to the balance between overestimated and underestimated projects.

Our theory and empirics also suggest that "scientific" entrepreneurs perform better on average due to what we called "the *learning effect*". But under the stated assumption that the value model for entrepreneurs that terminated vs. did not terminate their project remains the same, the difference-in-difference estimation does not change.

We next relax the assumption that the value model does not change depending on whether projects were discarded or not and rather assume that the value model is different according to the decision taken. This assumption will lead us to what we call our *upper-bound* condition. Under this assumption, we subtract the *learning effect* $\theta = 1.10$ to value of the correction for the projects of scientific entrepreneurs who terminate their projects, which now becomes -0.67. We subtract the estimated learning effect since the value model we estimated already considers the treatment effect for scientific entrepreneurs. The negative correction signals the existence of a bias reduction also for entrepreneurs that terminated their projects. We report these results in Table 8. The difference-in-difference estimation becomes 1.122, suggesting that the selection results in a lower reduction of value for treated (vs. control) entrepreneurs. Bad ideas are effectively ruled out, without a substantial increase in the *false negative* rate.

Table 8: Different Value for Terminate and Stay

	Terminate	Stay	Difference
Control	0.55 (*0.13*)	-0.26 (*0.11*)	0.805
Scientific	-0.67 (*0.14*)	-0.35 (*0.12*)	-0.317
Difference	1.221	0.099	1.122

Standard Deviation in Parentheses

These cases represent two extremes, one where the selection induced by the "scientific approach" is particularly positive (the *upper-bound*) and one where the approach leads to some adverse selection processes (the *lower-bound*), but close to zero. Despite these results should be interpreted with caution as they are based on assumptions, we believe they provide encouraging suggestive evidence of a well-balanced trade-off between the extent to which scientific entrepreneurs discard bad projects at the expense of good projects: in the worst case scenario (*lower bound*) these two effects essentially compensate each other; in the best case scenario (*upper bound*) the positive effect dominates the negative one.

6.4 Business Simulation Game

Our evidence so far suggests that the strong decrease in the false-positive rate caused by the application of the "scientific approach" more than compensates the increase in the false-negative rate. To further corroborate this interpretation, we run an additional experiment using a business simulation game with Master of Science (MSc) students. Business games are widely used in entrepreneurship and strategy education, and are claimed to provide a high value in the whole education process (e.g. Fox, Pittaway, and Uzuegbunam, 2018). We leverage on the potential of a real-life computer simulation for our research purposes, trying to replicate the results found in a real-life setting.

The game simulates the activities of an early-stage startup in its idea validation phase. The player, being in the co-founder role, has to infer the potential value of such startup and ultimately decide whether to launch it on the market or to terminate the project. To understand the idea type with which she is playing, the player can conduct several activities that mimic the real-life experience of an early-stage entrepreneur in the phase of idea validation. For instance, she can brainstorm with the virtual co-founders and create a business model canvas of her idea, or can validate her assumptions by running virtual interviews or questionnaires. The game includes a time dimension, with the market changing conditions over the game days: the player also receives market information in the form of short virtual newspaper articles. Once the player makes her decision, the game ends. The performance are evaluated on a different set of metrics, including market performance and scientific performance (in line with our definition of the scientific approach for decision-making).

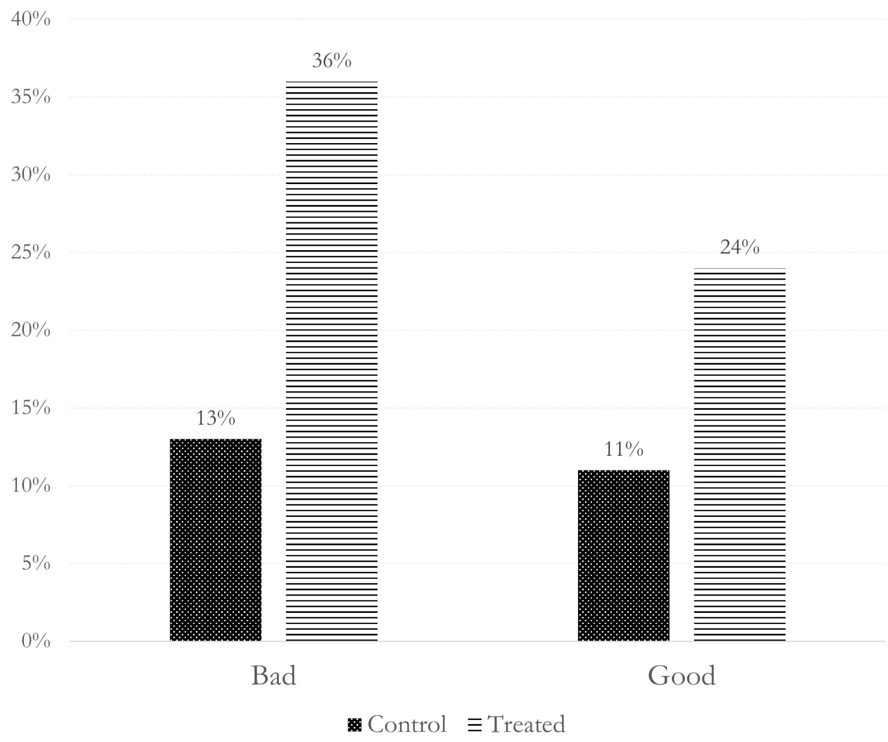
Following again the experiment run in Elfenbein, Knott, and Croson (2017), we force the simulation game to only have two type of scenarios: a *good* and a *bad* one. In the *good scenario*, the underlying idea to be evaluated by the player would be profitable. In this case, the best decision that the player could make would be to *launch* the project. In the *bad scenario* instead, the underlying idea to be evaluated can never be profitable. In this case, the best decision by the player would be to *terminate* the project. Given the nature of the game, the final profitability of the idea also depend on the specific choices that the player makes during the game. However, what is important for our testing is that such choices cannot change the fundamental value of the idea, that can only be low in the *bad scenario* and high in the *good scenario*.

In the context of a common lecture involving different Master of Science (MSc) programs at our home university, we asked students to play this game on a voluntary basis. Students were all enrolled in their first year of studies and came from three different programs. In one program only, students attended a course on scientific decision-making and thus we consider them as our treated group ($N = 28$). Students in the other two MSc programs did not attend such course and thus become our control group ($N = 50$). For game-related technical reasons, we randomized the in-game scenario (*good* vs. *bad*) before students came to class, stratifying by MSc program. The distribution of conditions turned out to be quite balanced, with 54% and 60% of students in the control and treated group respectively being assigned to the *good scenario*.

Students played for a maximum of 60 minutes, meaning that before the end of the lecturing hour they had to make their final decision. Again participation to the common lecture was voluntary. To incentivize players and also reproduce the monetary incentives of an entrepreneurial activity, we offered three 20€ gift cards (one per each MSc program) to the top players according to their in-game performance. To avoid biases, we introduced the game to students without any explicit reference to the "scientific approach". We mentioned that the main goal was to make the best decision given the players' own evaluations of the in-game idea, not suggesting any path or methodology to follow when running such evaulation.

Our outcome variable is a dummy taking value 1 if the player decides to not launch the idea. We have one main hypotheses based on the evidence from our RCTs: treated students following a scientific approach will terminate more often regardless of the underlying idea. Second, we would like to have further confirmation about the positive trade-off between the rate of false-positive and false-negatives. By comparing the two groups within each scenario, we expect that the treated students terminate projects relatively more often in the *bad scenario* than in the *good scenario*. Figure 4 shows the results by treatment group and in-game scenario.

Figure 4: Share of Terminated Projects



The graph show share of players deciding to terminate the idea development in the business simulation game. *Bad* and *Good* refer to the in-game scenarios, identifying the true potential of the underlying business idea to be evaluated. In the *good scenario* the best decision would be not to terminate, while the converse is true for the *bad scenario*.

Our results suggest how treated players are more likely to decide to terminate their ideas. Despite the small sample at our disposal a two-tailed t-test on the termination dummy between the two groups shows a significant difference (two-tailed, $t = -1.84; M1 = 0.12; M2 = 0.29$). When looking within scenarios, we also find an interesting pattern. In the *bad scenario*, where the best decision would be to terminate, treated players are three-times more likely to decide to do so. In the *good condition*, where the best decision would be to launch the startup, treated players are only two-times more likely to choose termination as their final decision. Thus, qualitatively, it seems that the reduction in the false-positive rate more than compensates the increase in the false-negative one. Moreover, the better ability of treated players in discriminating between ideas when compared to the control group is also signalled by the constant share of players belonging to the latter that decide to terminate.

Given the small sample size, these differences are not statistically significant[11]. However, putting together all the evidence coming from expert evaluators, the RCTs, the structural modeling and the simulation game provides a signal favorable towards the idea that

[11] Alongside acknowledging the fact that this experiment is underpowered, we also acknowledge the limitation of using a business simulation game as a testing tool. Indeed, a simulation game cannot reproduce the affection mechanisms and emotional dynamics that real-life entrepreneurs might display when it comes to the development of their own business idea.

applying a scientific approach to decision-making in entrepreneurial contexts could lead to positive selection outcomes, being the reduction in false-positives more than compensating the increase in the false-negative rate.

7 Discussion and Conclusion

In this paper we have explored the implications of encouraging entrepreneurs to employ a "scientific-approach to decision making". This approach, based on developing a theory of the problem faced, a set of hypotheses logically flowing from it, a series of tests to validate those hypotheses and a disciplined evaluation of results, is expected to induce entrepreneurs to reason in more probabilistic terms. Our empirical estimations and structural model predict that entrepreneurs following this approach are more likely to terminate their projects, as a result of a *debiasing effect* that leads entrepreneurs to develop a more conservative estimation of the value of their ideas. They also predict that treated entrepreneurs perform better because the "scientific approach" leads them to a better understanding of the problem and the solution space, an effect that we have called *learning effect*.

We estimated our models using data from two randomized control trial that involved 382 entrepreneurs and their business ideas. The results validate the models and support the intuition that the method leads entrepreneurs to a being more conservative in selecting project, reducing the rate of "false positive", but also to enhance the value of any project they focus on.

To better understand the potential value of this finding for scholars and practitioners, we reflect upon the extent to which the conservative attitude of scientific entrepreneurs might actually lead them to increase their rate of "false negative", that is, of good projects that they discard. We provide suggestive evidence coming from expert evaluations on the entrepreneurs' ideas, the RCTs, the structural estimation and a business simulation game. All these pieces of evidence support the idea that this possible effect is more than compensated by the beneficial effect of the reduction in false positives.

Our paper is not free of limitations. First, while we estimate our models across two distinct RCTs, our results could not be generalizable to other countries or entrepreneurial ecosystems other than the Italian one. However, other research (Novelli and Spina, 2021) show that the main results of Camuffo et al. (2020) generalize to another setting with, so we might expect the same mechanisms we found to be valid also externally. Relatedly, our RCTs encompass entrepreneurs at very early stages of idea development and with different levels of project novelty. Other research (Rindova and Courtney, 2020, von Hippel and von Krogh, 2015) suggest how entrepreneurs with highly novel ideas might follow different procedures to gather additional knowledge, not necessarily linked to hypothesis testing and ex-ante problem formulation. While our results on the *false-positive* and *false-negative* trade-off suggest an overall positive effect of the application of the scientific methodology, we cannot rule out the possibility that experimentation might have hindered the development of this specific subset of entrepreneurs. This is an interesting question that we believe is worth exploring in future research.

Overall, we believe that these findings might inform existing research on organizational decision making, particularly in innovative and entrepreneurial contexts. Our model and unique data on entrepreneurs' own assessment of their ideas enabled us to provide clarity of the effects of approaches to decision making that combine cognition with action (Levinthal, 2017, McDonald and Eisenhardt, 2020). Notably, one of these effects corresponds to an improvement in entrepreneurs' ability to correct their own estimations, reaching a more objective evaluation of the value of their ideas. These results also contribute to research on decision making by introducing the scientific method as a useful debiasing tool. A follow up issue that we investigated was the possibility that the scientific approach might lead entrepreneurs to be excessively conservative. Our complementary exploration of the trade-off associated with the stricter selection process induced by the scientific approach did identify evidence of an increase in the rate of false negatives. We believe that a promising path for future research would be that of exploring the contingencies under which the use of this method is mostly desirable.

We also believe that these results can be relevant for policy and practice. Educating entrepreneurs to follow a scientific approach to decision-making can indeed lead to a better selection process, effectively discarding projects that ultimately would perform poorly. This can lead to large tangible and intangible resource savings. Moreover, teaching entrepreneurs and students to think in "scientific" terms helps them in devising better strategies and development trajectories, resulting in higher performance.

References

Ashraf, Nava, Abhijit Banerjee, and Vesall Nourani (2021). "Learning to Teach by Learning to Learn". en. In: *Working paper*, p. 115.

Bennett, Victor M. and Aaron K. Chatterji (2019). "The Entrepreneurial Process: Evidence from a Nationally Representative Survey". en. In: *Strategic Management Journal*. ISSN: 1097-0266. DOI: 10.1002/smj.3077.

Bingham, Christopher B. and Jason P. Davis (June 2012). "Learning Sequences: Their Existence, Effect, and Evolution". In: *Academy of Management Journal* 55.3, pp. 611–641. ISSN: 0001-4273. DOI: 10.5465/amj.2009.0331.

Bingham, Christopher B. and Kathleen M. Eisenhardt (2011). "Rational Heuristics: The 'Simple Rules' That Strategists Learn from Process Experience". en. In: *Strategic Management Journal* 32.13, pp. 1437–1464. ISSN: 1097-0266. DOI: 10.1002/smj.965.

Bloom, Nicholas and John Van Reenen (Nov. 2007). "Measuring and Explaining Management Practices Across Firms and Countries*". In: *The Quarterly Journal of Economics* 122.4, pp. 1351–1408. ISSN: 0033-5533. DOI: 10.1162/qjec.2007.122.4.1351.

Camuffo, A. et al. (2021). "A Scientific Approach to Innovation Management: Theory and Evidence from Four Field Experiments". In: *CEPR Discussion Paper* No. DP15972.

Camuffo, Arnaldo et al. (Aug. 2020). "A Scientific Approach to Entrepreneurial Decision Making: Evidence from a Randomized Control Trial". In: *Management Science* 66.2, pp. 564–586. ISSN: 0025-1909. DOI: 10.1287/mnsc.2018.3249.

Chen, J.S et al. (2018). "The impact of learning and overconfidence on entrepreneurial entry and exit." In: *Organization Science* 29.6, pp. 989–1009.

Csaszar, Felipe A. and Daniella Laureiro-Martínez (Sept. 2018). "Individual and Organizational Antecedents of Strategic Foresight: A Representational Approach". In: *Strategy Science* 3.3, pp. 513–532. ISSN: 2333-2050. DOI: 10.1287/stsc.2018.0063.

Eisenhardt, Kathleen M. (Sept. 1989). "Making Fast Strategic Decisions In High-Velocity Environments". In: *Academy of Management Journal* 32.3, pp. 543–576. ISSN: 0001-4273. DOI: 10.5465/256434.

Eisenhardt, Kathleen M. and Christopher B. Bingham (Dec. 2017). "Superior Strategy in Entrepreneurial Settings: Thinking, Doing, and the Logic of Opportunity". en. In: *Strategy Science* 2.4, pp. 246–257. ISSN: 2333-2050, 2333-2077. DOI: 10.1287/stsc.2017.0045.

Eisenhardt, Kathleen M. and Timothy E. Ott (2017). "Rigor in Theory Building from Multiple Cases". In: *The Routledge Companion to Qualitative Research in Organization Studies*. Routledge. ISBN: 978-1-315-68610-3.

Elfenbein, Daniel, Anne Marie Knott, and Rachel Croson (2017). "Equity stakes and exit: An experimental approach to decomposing exit delay". In: *Strategic Management Journal* 38 (2), pp. 278–299. ISSN: 1097-0266. DOI: 10.1002/smj.2493.

Felin, Teppo, Stuart Kauffman, and Todd Zenger (Feb. 2020). *Microfoundations of Resources: A Theory*. en. SSRN Scholarly Paper ID 3549865. Rochester, NY: Social Science Research Network. DOI: 10.2139/ssrn.3549865.

Felin, Teppo and Todd R. Zenger (2009). "Entrepreneurs as Theorists: On the Origins of Collective Beliefs and Novel Strategies". In: *Strategic Entrepreneurship Journal* 3.2, pp. 127–146. ISSN: 1932-443X. DOI: 10.1002/sej.67.

Folta, Timothy B. (1998). "Governance and Uncertainty: The Trade-off between Administrative Control and Commitment". In: *Strategic Management Journal* 19.11, pp. 1007–1028. ISSN: 0143-2095.

Fox, Joe, Luke Pittaway, and Ikenna Uzuegbunam (2018). "Simulations in Entrepreneurship Education: Serious Games and Learning Through Play". In: *Entrepreneurship Education and Pedagogy* 1.1.

Gans, Joshua S., Scott Stern, and Jane Wu (2019). "Foundations of Entrepreneurial Strategy". en. In: *Strategic Management Journal* 40.5, pp. 736–756. ISSN: 1097-0266. DOI: 10.1002/smj.3010.

Gary, Michael Shayne and Robert E. Wood (2011). "Mental Models, Decision Rules, and Performance Heterogeneity". In: *Strategic Management Journal* 32.6, pp. 569–594. ISSN: 1097-0266. DOI: 10.1002/smj.899.

Gavetti, Giovanni and Jan W. Rivkin (June 2007). "On the Origin of Strategy: Action and Cognition over Time". In: *Organization Science* 18.3, pp. 420–439. ISSN: 1047-7039. DOI: 10.1287/orsc.1070.0282.

Ghanem, Dalia, UC Davis, and Sarojini Hirshleifer (n.d.). "Testing Attrition Bias in Field Experiments". In: *Working Paper 202010, University of California at Riverside, Department of Economics* (), p. 63.

Gross, Daniel (2017). "Performance feedback in competitive product development". In: *The RAND Journal of Economics* 28.2, pp. 438–466.

Kirtley, Jacqueline and Siobhan O'Mahony (2020). "What Is a Pivot? Explaining When and How Entrepreneurial Firms Decide to Make Strategic Change and Pivot". en. In: *Strategic Management Journal* n/a.n/a. ISSN: 1097-0266. DOI: 10.1002/smj.3131.

Levinthal, Daniel A. (Nov. 2017). "Resource Allocation and Firm Boundaries". In: *Journal of Management* 43.8, pp. 2580–2587. ISSN: 0149-2063. DOI: 10.1177/0149206316667458.

Lieberman, Marvin B., Gwendolyn K. Lee, and Timothy B. Folta (2017). "Entry, Exit, and the Potential for Resource Redeployment". en. In: *Strategic Management Journal* 38.3, pp. 526–544. ISSN: 1097-0266. DOI: 10.1002/smj.2501.

McDonald, Rory M. and Kathleen M. Eisenhardt (June 2020). "Parallel Play: Startups, Nascent Markets, and Effective Business-model Design". In: *Administrative Science Quarterly* 65.2, pp. 483–523. ISSN: 0001-8392. DOI: 10.1177/0001839219852349.

McGrath, Rita Gunther (1997). "A Real Options Logic for Initiating Technology Positioning Investments". In: *The Academy of Management Review* 22.4, pp. 974–996. ISSN: 0363-7425. DOI: 10.2307/259251.

NationalBusinessCapital (Jan. 2020). *2019 Small Business Failure Rate: Startup Statistics by Industry.*

Novelli, E. and C. Spina (2021). "When do Entrepreneurs Benefit from Acting Like Scientists? A Field Experiment in the UK". In: *INSEAD Working Paper* No. 2021/40/EFE.

Ott, Timothy E., Kathleen M. Eisenhardt, and Christopher B. Bingham (2017). "Strategy Formation in Entrepreneurial Settings: Past Insights and Future Directions". In: *Strategic Entrepreneurship Journal* 11.3, pp. 306–325. ISSN: 1932-443X. DOI: 10.1002/sej.1257.

Raffaelli, Ryan, Mary Ann Glynn, and Michael Tushman (2019). "Frame Flexibility: The Role of Cognitive and Emotional Framing in Innovation Adoption by Incumbent Firms". In: *Strategic Management Journal* 40.7, pp. 1013–1039. ISSN: 1097-0266. DOI: 10.1002/smj.3011.

Ries, Eric (Sept. 2011). *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses.* en. Crown. ISBN: 978-0-307-88791-7.

Rindova, Violina and Hugh Courtney (July 2020). "To Shape or Adapt: Knowledge Problems, Epistemologies and Strategic Postures Under Knightian Uncertainty". In: *Academy of Management Review*, amr.2018.0291. ISSN: 0363-7425, 1930-3807. DOI: 10.5465/amr.2018.0291.

Roodman, D (2011). "Fitting fully observed recursive mixed-process models with cmp". In: *Stata Journal* 1, pp. 159–206.

Siggelkow, Nicolaj (2002). "Evolution toward Fit". In: *Administrative Science Quarterly* 47 (1), pp. 125–159. ISSN: 0001-8392.

Sleesman, Dustin J. et al. (Jan. 2018). "Putting Escalation of Commitment in Context: A Multilevel Review and Analysis". In: *Academy of Management Annals* 12.1, pp. 178–207. ISSN: 1941-6520. DOI: 10.5465/annals.2016.0046.

Thomke, Stefan H. (June 1998). "Managing Experimentation in the Design of New Products". In: *Management Science* 44.6, pp. 743–762. ISSN: 0025-1909. DOI: 10.1287/mnsc.44.6.743.

Tversky, Amos and Daniel Kahneman (Sept. 1974). "Judgment under Uncertainty: Heuristics and Biases". In: *Science* 185.4157, pp. 1124–1131. DOI: 10.1126/science.185.4157.1124.

von Hippel, Eric and Marcie J. Tyre (Jan. 1995). "How Learning by Doing Is Done: Problem Identification in Novel Process Equipment". en. In: *Research Policy* 24.1, pp. 1–12. ISSN: 0048-7333. DOI: 10.1016/0048-7333(93)00747-H.

von Hippel, Eric and Georg von Krogh (Dec. 2015). "Identifying Viable "Need–Solution Pairs": Problem Solving Without Problem Formulation". In: *Organization Science*, orsc.2015.1023. ISSN: 1047-7039, 1526-5455. DOI: 10.1287/orsc.2015.1023.

Walsh, James P. (June 1995). “Managerial and Organizational Cognition: Notes from a Trip Down Memory Lane”. In: *Organization Science* 6.3, pp. 280–321. ISSN: 1047-7039. DOI: 10.1287/orsc.6.3.280.

Yang, Mu-Jeung et al. (2020). “How Do CEOs Make Strategy?” In: *NBER Working Paper* 36.

Appendix

Balance Checks

Table A1: Balance Checks Milan RCT

Variable Name	Description	Treatment		Control		Difference	
		Mean	SD	Mean	SD	b	p
Age	Age (Team Average)	31.47	8.18	31.41	7.90	-0.06	(0.950)
Analytic Thinking	Agreement on a 1-10 scale with the following statements (Team Average): "Analyzing the situation and looking at the evidence is critical to our company's decision-making", "We carefully assess all the possible alternatives before making a choice for our company", "We prefer to gather all the relevant information before making a decision for our company", "Multiple elements are taken into account when making a decision for our company, pros and cons are carefully evaluated in every situation"	8.38	3.68	8.07	3.28	-0.32	(0.475)
Background: Economics	Team members with an economics background (%)	0.41	0.42	0.31	0.37	-0.10**	(0.046)
Background: Other	Team members with no economics backgrounds (%)	0.22	0.36	0.20	0.33	-0.02	(0.696)
Background: STEM	Team members with a STEM (Science Technology Engineering Mathematics) backgrounds (%)	0.38	0.40	0.49	0.41	0.11**	(0.032)
Certainty	Agreement on a 1-10 scale with the following statements (Team Average): "We are sure about our business model", "We are sure about our strategy"	5.93	1.94	5.61	1.91	-0.32	(0.191)
Consensus	Answer on a 1-10 scale to the following questions (Team Average): "To what extent do you and your team members have consensus on the long term objectives of the firm?", "To what extent do you and your team members have consensus on the short term objectives of the firm?", "To what extent do you and your team members have consensus on the survival strategy of the firm?"	8.85	1.67	8.86	1.66	0.00	(0.990)
Education	Highest educational level attained by team members (5=PhD, 4=MBA, 3=MSc, 2=BA, 1=high school, 0=otherwise; Team Average)	1.94	0.74	1.95	0.80	0.00	(0.969)
Experience: Entrepreneurial	Number of years of entrepreneurial experience (Team Average)	1.09	2.19	0.93	1.44	-0.17	(0.480)
Experience: Industry	Number of years of experience in industry (Team Average)	2.84	3.82	2.33	3.62	-0.51	(0.280)
Experience: Managerial	Number of years of managerial experience (Team Average)	2.29	3.69	2.27	4.18	-0.02	(0.971)
Experience: Work	Number of years of work experience (Team Average)	8.73	7.75	9.02	8.85	0.28	(0.788)
Full Time	Percentage of team members working full-time	0.57	0.43	0.62	0.42	0.05	(0.390)
Gender (Female)	Proportion of women in the team	0.27	0.37	0.25	0.36	-0.03	(0.541)
Hours: Total Weekly	Weekly hours dedicated to the company (Team Average)	10.17	9.65	10.96	11.45	0.78	(0.560)
Idea Potential	Independent assessment of the value of the idea	47.22	21.22	47.31	23.25	0.09	(0.975)
Idea Value: Max	Maximum estimated value of the project (0 to 100)	85.08	16.29	85.67	16.16	0.59	(0.773)
Idea Value: Mean	Estimated value of the project (mean, 0 to 100)	65.40	15.53	64.52	16.69	-0.88	(0.668)
Idea Value: Min	Minimum estimated value of the project (0 to 100)	45.71	19.86	43.21	22.93	-2.50	(0.357)
Idea Value: Range	Estimated value of the project (range, 0 to 100)	39.37	18.85	42.46	20.99	3.10	(0.221)
Intuitive Thinking	Agreement on a 1-10 scale with the following statements (Team Average): "We are prone to following our intuitions when making company-related decisions", "We consider feelings and intuitions rather than analysis in our startup decisions", "First impressions are important when making decisions", "It is important to rely on gut feelings and intuition when making decisions"	4.09	1.70	3.83	1.74	-0.25	(0.244)
Lombardy	Dummy variable taking value of 1 when the majority of team members comes from the Italian region of Lombardy, 0 otherwise	0.56	0.47	0.57	0.46	0.01	(0.883)
Months to Revenue	Number of months to revenue	11.52	5.80	11.51	5.85	-0.01	(0.987)
Part Time	Percentage of team members working part-time	0.08	0.18	0.08	0.17	0.00	(0.941)
Probability Termination	Probability of terminating the project	31.64	32.53	32.35	31.60	0.70	(0.863)
Team Size	Number of team members	2.25	1.46	2.28	1.37	0.03	(0.858)
Observations		125		125		250	

Table A2: Balance Checks Turin RCT

Variable Name	Description	Treatment Mean	Treatment SD	Control Mean	Control SD	Difference b	Difference p
Age	Age (Team Average)	30.58	9.07	30.48	7.09	-0.10	(0.943)
Analytic Thinking	Agreement on a 1-5 scale with the following statements (Team Average): "Analyzing the situation and looking at the evidence is critical to our company's decision-making", "We carefully assess all the possible alternatives before making a choice for our company", "We prefer to gather all the relevant information before making a decision for our company" and "Multiple elements are taken into account when making a decision for our company, pros and cons are carefully evaluated in every situation"	4.27	0.65	4.39	0.56	0.13	(0.234)
Background: Economics	Team members with Economics backgrounds (%)	0.19	0.32	0.21	0.36	0.02	(0.789)
Background: Other	Team members with no Economics/STEM backgrounds (%)	0.56	0.43	0.44	0.46	-0.12	(0.130)
Background: STEM	Team members with a STEM (Science Technology Engineering Mathematics) backgrounds (%)	0.25	0.37	0.35	0.45	0.10	(0.161)
Confidence	Agreement on a 1-5 scale with the following statements (Team Average): "We are confident in our entrepreneurial skills", "We are sure we are deploying the best strategy for our business", "We are confident in our ability to manage our business", "We master the competences necessary for our venture" and "We are sure there is no better business model for our idea"	3.42	0.53	3.32	0.64	-0.10	(0.335)
Currently Studying	Number of team members enrolled in an education program at the time of training	0.26	0.30	0.22	0.30	-0.04	(0.429)
Education	Highest educational level attained by team members (5=PhD, 4=MBA, 3=MSc, 2=BA, 1=high school, 0=otherwise; Team Average)	1.86	0.88	2.07	1.08	0.21	(0.221)
Experience: Business Plan	Dummy taking value of 1 if the team had years of experience in business plan design, 0 otherwise	0.27	0.37	0.35	0.42	0.07	(0.282)
Experience: Entrepreneurial	Number of years of entrepreneurial experience (Team Average)	1.81	4.38	1.71	3.35	0.11	(0.873)
Experience: Industry	Number of years of experience in industry (Team Average)	2.88	5.65	2.99	5.01	0.10	(0.911)
Experience: Managerial	Number of years of managerial experience (Team Average)	1.56	2.71	1.73	3.74	0.18	(0.756)
Gender (Female)	Proportion of women in the team	0.33	0.39	0.25	0.35	-0.07	(0.256)
Hours: Total Weekly	Weekly hours dedicated to the company (Team Average)	11.26	9.85	11.62	12.32	0.35	(0.856)
Idea Maturity	Maturity of the idea (in months)	9.95	9.54	12.16	11.63	2.21	(0.234)
Idea Potential	Independent assessment of the value of the idea (two evaluators, average) based on five criteria: innovation, feasibility, sustainability, team competence, market size	48.85	12.05	49.17	12.77	0.33	(0.880)
Idea Value: Mean	Estimated value of the project (mean)	66.24.82	18.89	63.54	16.06	-2.69	(0.380)
Intuitive Thinking	Agreement on a 1-5 scale with the following statements (Team Average): "We are prone to following our intuitions when making company-related decisions" and "We consider feelings and intuitions rather than analysis in our startup decisions"	2.79	0.86	2.71	0.98	-0.08	(0.604)
Later Stage	Dummy variable taking value of 1 if the firm is at a more advanced stage than others, 0 otherwise	0.14	0.35	0.10	0.31	-0.03	(0.554)
Locus of Control	Agreement on a 1-7 scale with the following statements (Team Average): "In most jobs you need a lot of luck to excel", "One typically earns what they are worth", "To make money you just need to know the right people", "To get a good position you need luck", "Income is mainly the result of hard work", "There is a direct relationship between a personÕs abilities and the position he/she holds", "Many of the difficulties encountered at work concerns senior colleagues", "Generally, people who work well get rewarded", "Promotions are awarded to people who work well", "To find a good job, having a good network is more important than actual skills", "A well-trained person always finds a satisfycing job" and "To get a really good job you have to have high-level acquaintances"	3.85	0.67	3.78	0.70	-0.07	(0.556)
Months to Revenue	Number of months to revenue	12.42	11.20	14.63	10.51	2.21	(0.245)
Piedmont	Dummy variable taking value of 1 when the majority of team members comes from the Italian region of Piedmont and 0 otherwise	0.52	0.46	0.52	0.48	-0.00	(0.994)
Probability Pivot Idea	Probability of changing the business idea	30.69	22.96	32.42	26.56	1.73	(0.690)
Probability Pivot Other	Probability of changing other components of the business model	51.60	22.46	52.58	26.13	0.98	(0.817)
Probability Pivot Problem	Probability of changing the problem and customer segment	33.75	22.68	34.42	25.20	0.66	(0.873)
Probability Termination	Probability of terminating the project	12.95	16.27	17.31	21.52	4.36	(0.191)
Risk-averse	Agreement on a 1-7 scale with the following statements (Team Average): "In important matters I never take unnecessary risks, which can be avoided", "In important situations I never deliberately chose to take risks I could have avoided", "I always try to avoid situations that put me at risk of getting into trouble with other people", "I am always very careful and I put safety first" and "I prefer to avoid doing things that expose me to criticism and liability"	4.21	1.00	3.95	1.04	-0.26	(0.150)
Risk-taker	Agreement on a 1-7 scale with the following statements (Team Average): "I can be pretty reckless and take some big risks", "I think I often act boldly and courageously", "I am a brave and daring person and I like to tempt fate in various situations", "There is a direct relationship between a personÕs abilities and the position he/she holds" and "I think I am often less cautious than other people"	4.04	1.10	3.98	0.90	-0.06	(0.716)
Scientific intensity: 1 Theory	Theory development score	2.87	1.34	3.02	1.21	0.15	(0.514)
Scientific intensity: 2 Hypothesis	Hypothesis development score	2.12	1.64	1.97	1.50	-0.15	(0.587)
Scientific intensity: 3 Test	Test score	1.32	1.71	1.28	1.67	-0.03	(0.906)
Scientific intensity: 4 Valuation	Valuation score	0.85	1.50	0.94	1.62	0.09	(0.750)
Self-efficacy	Agreement on a 1-7 scale with the following statements (Team Average): "I think I will always be able to achieve a goal even if I have to perform a difficult task", "Faced with new tasks and challenges, I am always confident that I will be able to complete them", "I am sure I will succeed", "When I have a goal, I almost always get better results than others", "When I take a test or an exam I am sure I can pass it successfully", "I am confident that my results will be recognized and appreciated by others", "I am not worried about difficult situations, because so far I have always managed to get by with my skills", "I never had any problem understanding and facing even the most complicated situations" and "I think I get the crux of the matter first"	5.43	1.08	5.56	0.95	0.13	(0.461)
Self-regulation	Agreement on a 1-7 scale with the following statements (Team Average): "People can count on me to meet the set and planned deadlines", "I can hardly say no", "I change my mind quite often", "Others would describe me as an impulsive person", "I wish I had more self-discipline", "I get carried away by my feelings", "I am not easily discouraged", "Sometimes I canÕt stop but do something, even though I know it is wrong", "I often act without thinking about all the alternatives", "I often do things that seem right in the present, even at the expense of future goals" and "When I pursue a goal I follow the original plan, even when I realize that it is not the best"	4.97	0.83	5.23	0.86	0.26*	(0.074)
Startcup	Dummy variable taking value of 1 if the firm takes part to a local competition, 0 otherwise	0.11	0.31	0.18	0.39	0.07	(0.244)
Team Size	Number of team members	2.54	1.61	2.18	1.39	-0.36	(0.173)
Observations		65		67		132	

Extended Selection Model: Full Estimation

We report here the results of the first two steps of the four-equations extended selection model, estimated separately. Table A3 shows the results of the Heckman selection model with entrepreneurs' own predictions of idea value at the Late stage ($\hat{v}_L$) used to identify the selection equation. Table A4 adds the intermediate equation that instruments $\hat{v}_L$ with the pre-training (baseline) evaluations $\hat{v}_0$.

Table A3: Heckman Selection Model

	Value Equation	Selection Equation
Intervention	1.080***	-0.382***
	(0.412)	(0.108)
$\hat{v}_L$		0.585***
		(0.148)
Startup Experience	0.211***	-0.0148
	(0.0734)	(0.0242)
Team Size (Baseline)	0.311*	-0.0400
	(0.171)	(0.0506)
Education	0.269	-0.0878
	(0.233)	(0.0868)
Age	-0.087***	0.028***
	(0.025)	(0.008)
Hours Worked (Baseline)	0.008	0.009*
	(0.013)	(0.005)
Constant	1.850**	-2.441***
	(0.742)	(0.755)
Correlation	-0.223***	
	(0.0700)	
RCT Dummies	Yes	Yes
Mentor Dummies	Yes	Yes
N	382	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Standard errors clustered at the classroom level

Table A4: Extended Heckman Selection Model

	Value Equation	Selection Equation	$\hat{v}_L$
Intervention	1.019**	-0.280**	-0.0213
	(0.435)	(0.120)	(0.0496)
$\hat{v}_L$		1.875***	
		(0.320)	
$\hat{v}_0$			0.282***
			(0.0689)
Startup Experience	0.213***	-0.029	0.009
	(0.072)	(0.025)	(0.007)
Team Size (Baseline)	0.308*	-0.0591	0.0127
	(0.169)	(0.039)	(0.0140)
Education	0.265	-0.007	-0.045**
	(0.239)	(0.073)	(0.0230)
Age	-0.083***	0.015*	0.006**
	(0.024)	(0.009)	(0.003)
Hours Worked (Baseline)	0.009	0.006	0.001
	(0.013)	(0.005)	(0.002)
Constant	1.710**	-7.460***	2.768***
	(0.686)	(1.413)	(0.321)
Correlation	-0.197***		
	(0.0730)		
RCT Dummies	Yes	Yes	Yes
Mentor Dummies	Yes	Yes	Yes
N	382		

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Standard errors clustered at the classroom level

Full Estimation Results

We report in Table A5 the results of the full `cmp` estimation, used to retrieve the structural coefficients shown in Table 4 the main text.

Table A5: Structural Model: Full Estimation

Model	Value (Eq 1) *OLS*	Selection (Eq 14) *Probit*	z_L (Eq 13) *OLS*	z_E (Eq 12) *OLS*	v_L^* (Eq 4*) *OLS*	v_E^* (Eq 3*) *OLS*
Intervention	1.097**	-0.452***	-0.175**	0.049	-0.012	-0.049*
	(0.455)	(0.122)	(0.079)	(0.086)	(0.049)	(0.027)
z_L		-0.272				
		(0.181)				
z_E			1.018***			
			(0.183)			
z_0				0.312***		
				(0.041)		
Startup Experience	0.187***			0.005	0.011	0.008
	(0.072)			(0.019)	(0.007)	(0.005)
Team Size (Baseline)	0.301**			-0.137***	0.023	0.009
	(0.145)			(0.044)	(0.015)	(0.013)
Education	0.231			0.123*	-0.044*	-0.031
	(0.231)			(0.064)	(0.024)	(0.021)
Age	-0.079***			-0.025***	0.005	0.002
	(0.025)			(0.007)	(0.003)	(0.002)
Hours Worked (Baseline)	0.011			-0.001	0.001	0.002***
	(0.013)			(0.003)	(0.001)	(0.001)
Constant	1.695**	0.256	0.396	-0.318	3.931***	4.153***
	(0.760)	(0.296)	(0.357)	(0.286)	(0.141)	(0.059)
Correlation	-0.278*					
	(0.145)					
RCT Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Mentor Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Equation $ln(\sigma)$ (OLS only)	1.120***		0.176**	-0.00585	-0.849***	-1.066***
	(0.079)		(0.073)	(0.036)	(0.092)	(0.093)
N				382		

All equations contain dummies for RCT and instructor, with standard errors clustered at the classroom level (in parentheses). *** p<0.01, ** p<0.05, * p<0.1

Alternative Computation of Structural Coefficients

Table A6: Structural Parameter Computation

Alternative computation from Z equations

Parameter	Computation	Equations
θ	θ	1
σ_E	*OLS variance*	3*
σ_L	*OLS variance*	4*
σ_F	$-\frac{\sigma_L}{\lambda F}$	14, 4*
c_{ET}	$beta_E \sigma_E - \theta$	3*, 12
c_{LT}	$-\beta_L \sigma_L + c_{ET}$	4*, 13, 12
c_{FT}	$\beta_F \sigma_F + c_{LT}$	14, 4*, 13, 12

Table A7: Estimated Structural Parameters

Alternative estimation from Z equations

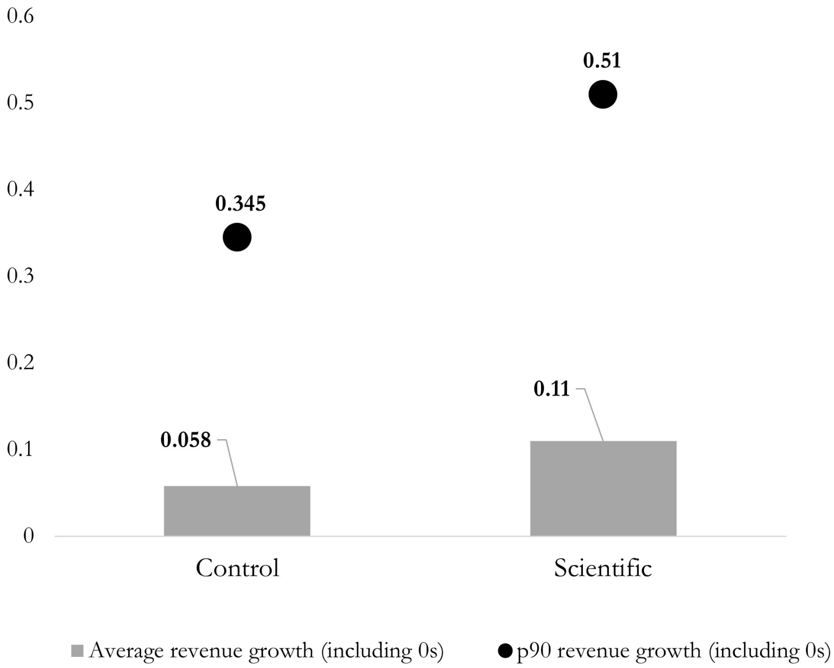
	Parameter	Std. Err	z-score
θ	1.10	0.455	2.41
σ_F	0.34	0.032	10.72
σ_L	0.43	0.039	10.83
σ_F	1.57	1.088	1.44
c_{ET}	-1.11	0.435	-2.56
c_{LT}	-1.04	0.443	-2.34
c_{FT}	-1.75	0.411	-4.26

Structural parameters retrieved after ML estimation of the six equations described in Section 4.3. This alternative computation retrieves the parameters c_{ET} and c_{LT} from Eq. 12 and 13 rather than from Eq. 3* and Eq. 4*. Parameters and their standard errors are retrieved using the `nlcom` routine in Stata 16. All estimated equations include dummies for RCT and instructor, with standard errors clustered at the classroom level.

Robustness Checks: Average Revenue Growth results

In this subsection we report the results for the extended selection model and structural models when using the average revenue growth over time as dependent variable. First, Figure 5 shows the statistics pertaining to average revenue growth. The latter has been computed as the average of the growth of revenues between each data point in the sample, only for firms remaining active for the whole observation window.

Figure 5: Average Revenue Growth Over Time



The graph show the mean and 90th percentile values of the average revenue growth over time for firms active in the market, by treatment condition.

Consistently with the figures on revenue growth with respect to the baseline discussed in the main text, treated firms have a higher growth rate when compared to control firms, both on average and at the 90th percentile.

Table A8 reports the results of the four-equations extended selection model, using as dependent variable the average revenue growth over time.

Table A8: Extended Selection Model

	Value Equation	Selection Equation	$\hat{v}_L$	$\hat{v}_E$
Intervention	0.057**	-0.280**	0.043	-0.058**
	(0.024)	(0.120)	(0.043)	(0.027)
$\hat{v}_L$		1.875***		
		(0.320)		
$\hat{v}_E$			1.102***	
			(0.360)	
$\hat{v}_0$				0.256***
				(0.069)
Startup Experience	0.012***	-0.029	0.005	0.005
	(0.004)	(0.025)	(0.008)	(0.005)
Team Size (Baseline)	0.017*	-0.059	0.011	0.002
	(0.009)	(0.040)	(0.013)	(0.014)
Education	0.015	-0.008	-0.007	-0.035*
	(0.013)	(0.074)	(0.022)	(0.020)
Age	-0.005***	0.015*	0.003	0.003
	(0.001)	(0.009)	(0.003)	(0.002)
Hours Worked (Baseline)	0.000	0.006	-0.002	0.002**
	(0.001)	(0.005)	(0.002)	(0.001)
Constant	0.097**	-7.462***	-0.652	3.103***
	(0.038)	(1.411)	(1.503)	(0.317)
Correlation	-0.198***			
	(0.075)			
RCT Dummies	Yes	Yes	Yes	Yes
Mentor Dummies	Yes	Yes	Yes	Yes
N	382			

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

The value equation is estimated through an OLS conditioned on the selection equation, estimated through a *probit* model. The last two equations are estimated through OLS.

Standard errors clustered at the classroom level.

The *learning* effect is still significant and estimated to lead to an additional average growth of 5.7 pp for treated entrepreneurs. The effect is quite sizeable given the averages shown in Figure 5.

Table A9 reports the results of the structural estimation using the average revenue growth over time as dependent variable. The *learning effect* is consistent with previous estimations and the same is valid for the different *debiasing effects*, despite the substantial change in the nature of the dependent variable.

Table A9: Estimated Structural Parameters

	Parameter	Std. Err	z-score
θ	0.06	0.025	2.41
σ_E	0.34	0.032	10.70
σ_L	0.43	0.039	10.83
σ_F	1.57	1.088	1.44
c_{ET}	-0.11	0.035	-3.15
c_{LT}	-0.07	0.035	-1.53
c_{FT}	-0.78	0.457	-1.71

Structural parameters retrieved after ML estimation of the six equations described in Section 4.3, but using the average revenue growth over time as dependent variable. Parameters and their standard errors are retrieved using the `nlcom` routine in Stata 16. All estimated equations include dummies for RCT and instructor, with standard errors clustered at the classroom level.

Robustness Checks: Additional Controls

In this robustness check, we include two additional controls to the main estimation that resulted to be statistically different between the two treatment groups at the baseline. The dependent variable for the *value equation* is the revenue growth with respect to the baseline, as in the analyses reported in the main text.

We add the variable *Self-regulation*, which accounts for the team-level discipline in organization and decision-making activities measured through a 11-item Likert scale. This variable is only available for the RCT conducted in Turin, with a statistical difference between treatment groups significant at 10% (see Table A2).

Second, we add the variable *Background: Economics*, which records the percentage of team members with a degree in economics. This variable is only available for the RCT conducted in Milan, with a statistical difference between treatment groups significant at 5% (see Table A1). We set at 0 the value of these two variables for the RCTs where they were not recorded.

Table A10 reports the results of the four-equation extended selection model with these additional controls. Estimates show a consistent *learning effect*, with a slightly lower magnitude estimated to be around an additional 89 pp of revenue growth for the treated entrepreneurs. The likelihood of being active on the market at the end of the observation window is still lower for treated entrepreneurs. Consistently with the main results, treated entrepreneurs tend to estimate lower values when asked to estimate the value of their ideas in the early data point.

Table A11 shows the structural parameters from the full fledged estimation. All the parameters estimated are consistent with what we find in the main analyses, with a smaller estimation of the *learning effect*.

Table A10: Extended Selection Model

	Value Equation	Selection Equation	$\hat{v}_L$	$\hat{v}_E$
Intervention	0.894**	-0.279**	0.060	-0.056**
	(0.403)	(0.113)	(0.046)	(0.027)
$\hat{v}_L$		1.861***		
		(0.327)		
$\hat{v}_E$			1.150***	
			(0.376)	
$\hat{v}_0$				0.245***
				(0.070)
Startup Experience	0.208***	-0.028	0.004	0.005
	(0.076)	(0.026)	(0.008)	(0.005)
Team Size (Baseline)	0.306*	-0.060	0.007	0.003
	(0.160)	(0.041)	(0.014)	(0.014)
Education	0.216	-0.014	-0.002	-0.037*
	(0.247)	(0.079)	(0.024)	(0.021)
Age	-0.074***	0.014*	0.002	0.003
	(0.023)	(0.008)	(0.003)	(0.002)
Hours Worked (Baseline)	0.010	0.005	-0.002	0.002**
	(0.012)	(0.005)	(0.002)	(0.001)
Background: Economics	1.109*	-0.023	-0.141**	0.022
(RCT Milan only)	(0.580)	(0.211)	(0.071)	(0.035)
Self-regulation	-0.576	0.101	0.039	0.039
(RCT Turin only)	(0.397)	(0.162)	(0.046)	(0.026)
Constant	1.164*	-7.376***	-0.784	3.149***
	(0.693)	(1.483)	(1.561)	(0.319)
Correlation	-0.179**			
	(0.091)			
RCT Dummies	Yes	Yes	Yes	Yes
Mentor Dummies	Yes	Yes	Yes	Yes
N	382			

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

The value equation is estimated through an OLS conditioned on the selection equation, estimated through a *probit* model. The last two equations are estimated through OLS.

Standard errors clustered at the classroom level.

Table A11: Estimated Structural Parameters

	Parameter	Std. Err	z-score
θ	0.95	0.428	2.22
σ_E	0.34	0.032	10.62
σ_L	0.42	0.039	10.77
σ_F	2.43	3.014	0.81
c_{ET}	-1.00	0.426	-2.34
c_{LT}	-0.95	0.413	-2.29
c_{FT}	-1.90	0.914	-2.08

Structural parameters retrieved after ML estimation of the six equations described in Section 4.3, adding additional controls to the equations. Parameters and their standard errors are retrieved using the `nlcom` routine in Stata 16. All estimated equations include dummies for RCT and instructor, with standard errors clustered at the classroom level.

Robustness Checks: Trimmed Revenue Growth

In this robustness check, we re-run the main specifications using as a trimmed version of the dependent variable (i.e. revenue growth with respect to the baseline). We trim the variable at the 99th percentile, resulting in the drop of three outlier firms. The average of the revenue growth variable drops from 1.47 to 1.34. The sample size employed in this exercise is thus of 379 observations.

Table A13 reports the results for the four-equation extended selection model. We find results consistent with the main estimation, with a smaller coefficient on the intervention variable (the *learning effect*). This signals that the larger magnitude of the coefficient found in the main specification might be highly influenced by the presence of three well-performing companies. However, the estimated effect in this robustness test is still high in magnitude and significant, signalling that the detected effect is robust. Results on the *debiasing effect* are consistent, also looking at Table A13 that reports the structural results from the full-fledged model. Consistently with the main estimation, we indeed find negative and significant parameters for the three *debiasing effects* over time.

Table A12: Extended Selection Model

	Value Equation	Selection Equation	$\hat{v}_L$	$\hat{v}_E$
Intervention	0.748*	-0.293**	0.042	-0.060**
	(0.425)	(0.122)	(0.044)	(0.027)
$\hat{v}_L$		1.844***		
		(0.325)		
$\hat{v}_E$			1.103***	
			(0.362)	
$\hat{v}_0$				0.254***
				(0.069)
Startup Experience	0.229***	-0.028	0.005	0.005
	(0.068)	(0.026)	(0.008)	(0.005)
Team Size (Baseline)	0.277*	-0.059	0.011	0.002
	(0.165)	(0.040)	(0.013)	(0.014)
Education	0.284	-0.010	-0.007	-0.034*
	(0.239)	(0.074)	(0.022)	(0.020)
Age	-0.088***	0.015*	0.003	0.003
	(0.025)	(0.009)	(0.003)	(0.002)
Hours Worked (Baseline)	0.013	0.006	-0.002	0.002**
	(0.012)	(0.005)	(0.002)	(0.001)
Constant	1.164*	-7.376***	-0.784	3.149***
	(0.693)	(1.483)	(1.561)	(0.319)
Correlation	-0.188**			
	(0.081)			
RCT Dummies	Yes	Yes	Yes	Yes
Mentor Dummies	Yes	Yes	Yes	Yes
N	379			

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

The value equation is estimated through an OLS conditioned on the selection equation, estimated through a *probit* model. The last two equations are estimated through OLS. Standard errors clustered at the classroom level.

Table A13: Estimated Structural Parameters

	Parameter	Std. Err	z-score
θ	0.82	0.450	1.83
σ_E	0.35	0.032	10.66
σ_L	0.43	0.039	10.86
σ_F	1.59	1.089	1.46
c_{ET}	-0.87	0.448	-1.96
c_{LT}	-0.84	0.434	-1.93
c_{FT}	-1.58	0.411	-3.84

Structural parameters retrieved after ML estimation of the six equations described in Section 4.3, using a trimmed version of the dependent variable ($N = 379$). Parameters and their standard errors are retrieved using the `nlcom` routine in Stata 16. All estimated equations include dummies for RCT and instructor, with standard errors clustered at the classroom level.

Attrition

In Table A14, we report tests on a subset of covariates to compare "attritors" with respondents by treatment conditions. We report the results of F-test for joint orthogonality, comparing all four groups, and t-tests comparing "attritors" and respondents within treatment.

Table A14: Selective Attrition Tests

	Control Attritors	Control Respondents	Treated Attritors	Treated Respondents	F-Test
Startup Experience	0.842	1.251	0.286	1.425	0.982
	[0.258]	[0.188]	[0.163]	[0.243]	
Team Size	2.125	2.262	1.500	2.392**	1.856
	[0.284]	[0.106]	[0.251]	[0.114]	
Education	1.642	2.040**	1.964	1.911	1.807
	[0.151]	[0.071]	[0.259]	[0.059]	
Age	29.441	31.319	31.780	31.120	0.412
	[1.323]	[0.599]	[1.848]	[0.649]	
Industry Experience	2.028	2.637	3.810	2.780	0.527
	[0.894]	[0.319]	[1.243]	[0.340]	
Managerial Experience	1.083	2.226	1.571	2.076	0.741
	[0.371]	[0.327]	[0.650]	[0.261]	
Hours Worked	13.243	12.864	13.393	13.225	0.012
	[3.475]	[1.515]	[5.432]	[1.451]	
Predicted Value	65.896	63.875	67.036	66.244	0.681
	[3.170]	[1.288]	[5.285]	[1.191]	
Probability of Termination	0.168	0.215	0.220	0.166	1.865
	[0.039]	[0.016]	[0.062]	[0.015]	
RA Evaluation	44.229	48.493	40.750	48.339	1.004
	[4.270]	[1.551]	[5.604]	[1.382]	
Observations	24	168	14	176	

Selective attrition tests, by treatment condition.

T-tests are conducted comparing attritors and respondents within treatment condition.

All variables are recorded at the baseline. *RA Evaluation* refers to an evaluation made by Program Assistants on the quality of the applicants' ideas. For the definition of the different variables, please refer to the Balance Check tables.

***$p < 0.01$, **$p < 0.05$, * $p < 0.10$

The within-treatment comparison signals that there are differences between respondents and attritors in terms of education and team size. However, the comparison between respondents across conditions, does not reveal any significant difference, but for the *Predicted Value*, which is sligthly higher for respondents in the treated group. Nevertheless, any of the F-tests for joint orthogonality is significant at conventional levels.

To ensure our main results are robust to attrition, we estimate our main models excluding those observations. Thus, we run the estimation on a sample of 344 ventures (Control

= 168; Treated = 176). Table A15 reports the result for the four-step simplified model, while Table A16 reports the results from the structural estimation.

Table A15: Extended Selection Model

	Value Equation	Selection Equation	$\hat{v}_L$	$\hat{v}_E$
Intervention	0.927**	-0.248**	0.035	-0.049*
	(0.443)	(0.120)	(0.043)	(0.026)
$\hat{v}_L$		1.954***		
		(0.318)		
$\hat{v}_E$			1.001***	
			(0.383)	
$\hat{v}_0$				0.238***
				(0.077)
Startup Experience	0.216***	-0.020	0.006	0.003
	(0.079)	(0.024)	(0.007)	(0.005)
Team Size (Baseline)	0.315*	-0.046	0.017	-0.004
	(0.177)	(0.039)	(0.014)	(0.015)
Education	0.298	-0.008	-0.000	-0.032
	(0.259)	(0.074)	(0.022)	(0.022)
Age	-0.098***	0.011	0.003	0.003
	(0.026)	(0.010)	(0.003)	(0.003)
Hours Worked (Baseline)	0.010	0.006	-0.002	0.002**
	(0.016)	(0.006)	(0.002)	(0.001)
Constant	2.528***	-7.868***	-0.286	3.187***
	(0.841)	(1.355)	(1.597)	(0.352)
Correlation	-0.237***			
	(0.088)			
RCT Dummies	Yes	Yes	Yes	Yes
Mentor Dummies	Yes	Yes	Yes	Yes
N	344			

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

The value equation is estimated through an OLS conditioned on the selection equation, estimated through a *probit* model. The last two equations are estimated through OLS.

Standard errors clustered at the classroom level. Attritors are excluded by this estimation.

Table A16: Estimated Structural Parameters

	Parameter	Std. Err	z-score
θ	1.03	0.458	2.26
σ_E	0.34	0.036	9.59
σ_L	0.43	0.043	10.00
σ_F	1.28	0.692	1.84
c_{ET}	-1.07	0.455	-2.36
c_{LT}	-1.04	0.445	-4.33
c_{FT}	-1.60	0.450	-3.55

Structural parameters retrieved after ML estimation of the six equations described in Section 4.3, excluding "attritors" ventures ($N = 344$). Parameters and their standard errors are retrieved using the `nlcom` routine in Stata 16. All estimated equations include dummies for RCT and instructor, with standard errors clustered at the classroom level.