# THE LEAN STARTUP METHOD:
# EARLY-STAGE TEAMS AND HYPOTHESIS-BASED PROBING OF BUSINESS IDEAS

**Michael Leatherbee**
Pontificia Universidad Católica de Chile
Departamento de Ingeniería Industrial y de Sistemas
ORCID iD: https://orcid.org/0000-0002-9125-8271
Vicuna Mackenna 4860, Macul
Santiago, Chile
mleatherbee@ing.puc.cl

**Riitta Katila**
Stanford University
Department of Management Science & Engineering
ORCID iD: https://orcid.org/0000-0001-8125-515X
475 Via Ortega
Stanford, CA 94305
rkatila@stanford.edu

*Strategic Entrepreneurship Journal, 2020*

We are grateful to Steve Blank, Tom Byers, Kathy Eisenhardt, Todd Morrill, Alexander Osterwalder, Yves Pigneur, and Michelle Rogan for discussing the research design and interpretation of the findings with us. Participants at the Strategic Management Society Special Conference in Santiago, the Frontiers of Entrepreneurship Conference at UNC, STVP Research Seminar at Stanford, and the Academy of Management Conference in Vancouver provided thoughtful feedback. The authors gratefully acknowledge the financial support from the National Science Foundation (Grant # 1305078), the Stanford Technology Ventures Program, Stanford Institute for Innovation in Developing Economies, Chile's National Commission for Scientific and Technological Research (CONICYT) Núcleo Milenio Research Center in Entrepreneurial Strategy Under Uncertainty (NS130028), ANID (The Chilean Agency for Research and Development) for its support through FONDECYT Iniciación Project number 11170537, CORFO, and the Aubrey Chernick Foundation. We thank Delara Fadavi, Macarena Gonzalez, Cassi Janakos, Nicholas Morrill, and the Bay Area I-Corps Node personnel for their logistical and research assistance. Much appreciation goes to Co-Editor Christoph Zott and two anonymous reviewers for their constructive feedback.

Keywords: Lean startup method, hypothesis-based probing, business idea, young-firm teams, MBA education, learning-by-thinking, learning-by-doing

**ABSTRACT**

**Research summary**

We examine a learning-by-doing methodology for iteration of early-stage business ideas known as the "lean startup." The purpose of this paper is to lay out and test the key assumptions of the method, examining one particularly relevant boundary condition: the composition of the startup team. Using unique and detailed longitudinal data on 152 NSF-supported lean-startup (I-Corps) teams, we find that the key components of the method—hypothesis formulation, probing, and business idea convergence—link up as expected. We also find that team composition is an important boundary condition: business-educated (MBA) members resist the use of the method, but appreciate its value ex-post. Formal training in learning-by-thinking methods thus appears to limit the spread of learning-by-doing methods. In this way, business theory constrains business practice.

**Managerial summary**

Lean startup methodology has rapidly become one of the most common and trusted innovation and entrepreneurship methods by corporations, startup accelerators, and policymakers. Unfortunately, it has largely been portrayed as a one-size-fits-all solution—its key assumptions subject to little rigorous empirical testing, and the possibility of critical boundary conditions ignored. Our empirical testing supports the key assumptions of the method, but points to business education of team members as a critical boundary condition. Specifically, MBAs resist the use of the method despite being in a strong position to leverage it. Results from a post-hoc analysis we conducted also suggest that more engagement with the method relates to higher performance of the firm in the 18-month period following the lean startup intervention.

*We decided our company was a go! We have enough data to support our hypotheses and some good feedback from potential partners.*

–NSF I-Corps team, 2016.

Bounded rationality—finite information, finite minds, and finite time—makes young firms imperfect decision-makers (March and Simon, 1958; McGrath and Macmillan, 2009; Hallen and Pahnke, 2016). Early-stage teams, especially under high levels of environmental uncertainty and in technology-based industries, are unlikely to decide on an optimal business idea[1] from the very beginning of their ventures (Contigiani and Levinthal, 2018). Too much relevant information is missing, and capturing it requires time and effort. To address this problem and reduce uncertainty about their businesses' viability, entrepreneurs are encouraged to iterate their business ideas using both "learning-by-thinking" and "learning-by-doing" methods that are outlined in research (Gans, Stern, and Wu, 2019; Ott, Eisenhardt, and Bingham, 2017). Understanding these methods is particularly important because getting the business idea "right" early on can have a high impact in the long run (McDonald and Gao, 2019; Zott and Amit, 2007).

In the past decade or so, a new learning-by-doing methodology known as the "lean startup" has emerged. The goal is simple: to help early-stage teams iterate business ideas until they are able to make a sound decision about them (Blank, 2003; Ries, 2011). The method's significant features are (1) formulation of hypotheses in nine pre-identified areas of the business idea (using a shorthand visualization called canvas) and (2) "getting out of the building" to probe each hypothesis by interviewing customers and other stakeholders. The expected outcome of this hypothesis-probing process is to converge on a business idea by "confirming" or "disconfirming"

[1] We define "business idea" as startup's business concept including all the elements that would typically be included in a business plan (Delmar and Shane, 2003).

the hypotheses.[2]

Lean startup is a blend of previously identified learning-by-doing methods. It draws inspiration from previous blended approaches such as discovery-driven planning that similarly urge teams to articulate their underlying assumptions and to get data to iterate them (McGrath and MacMillan, 1995, 2009), and it builds on some existing principles of experimentation (i.e., hypotheses and their testing). It is novel, however, in its strong emphasis on interviewing customers and in its shorthand visualization of the core components of the business idea (Contigiani and Levinthal, 2018).

Lean startup is currently one of the most widely embraced entrepreneurship methods, and it is particularly valuable under environmental uncertainty (Contigiani and Levinthal, 2018; Kerr, Nanda, and Rhodes-Kropf, 2014). Nevertheless, the method's key assumptions and boundary conditions have not been subject to rigorous empirical analysis. Addressing this gap will thus be key to improving the method and to improving business idea processes in general.

The boundary conditions in particular require examination. The belief that early-stage teams would promptly revise their ideas in response to market information (e.g., customer interviews) is in sharp contrast with the description in recent research of some founders as reluctant to change their original ideas, even when external information prescribes it (Parker, 2006; Crilly, 2018; Grimes, 2017). In particular, teams deeply steeped in educational expertise often resist implementing changes in response to external feedback (Katila and Shane, 2005),

---

[2] The lean startup method uses the terms "hypothesis testing," "validation" and "experimentation" in reference to the process of checking the assumptions underlying the business idea. Throughout this paper we use the term "probing", as it more accurately reflects what the teams are actually doing, i.e., interviewing customers, but not necessarily testing or running experiments in a scientifically rigorous way. We also use the term "convergence" rather than validation to clearly differentiate the method from scientific experiments that are not the method's focus. Convergence is defined as a team's belief that a particular assumption in the business-model canvas now has reduced level of uncertainty and is not considered an ongoing hypothesis to test.

suggesting educational background as a potentially relevant boundary condition.[3] A better understanding of team composition is also important because this factor can be adjusted by organizations directly, providing a clear point of intervention.

We examine 152 early-stage teams engaged in a standardized implementation of the lean startup method over an 8-week period. Our particular research setting is I-Corps—a National Science Foundation (NSF) program developed by Steve Blank using the lean startup method, which has been offered in a consistent manner since 2011. During the program, all teams iterate a business idea for a core technology originally created by a scientist (one of the team members) and financially supported by a grant from the National Science Foundation.

There are several findings. First, the results from our baseline hypotheses shed light on the *underlying assumptions* of the method. Our results confirm the central role of probing (i.e., customer interviews). Probing not only motivates convergence on a business idea, as the method predicts. We discover that it also motivates the formulation of new ideas, with their corresponding hypotheses-to-be-tested. Probing can thus help teams to dislodge an entrenched vision of the business idea, as needed, and find a new plan. We also find that if a team converges on an idea, it is likely to formulate fewer new hypotheses, which suggests that the method involves a natural stopping mechanism.[4] In contrast, we do not find confirmation for the central role of hypothesis formulation: formulating a large number of hypotheses does not drive or help teams' probing. Altogether, we find support for the main assumptions of the method, but we note that one criticism—that endless formulation of new hypotheses may tire out entrepreneurs and

[3] Parker (2006), for example, found that older entrepreneurs were less willing to change business ideas even when external feedback suggests otherwise, but there is very little evidence about team expertise and "learning-by-doing" methods in general.

[4] These findings are particularly interesting given prior literature's critique of the method that probing leads entrepreneurs to change ideas too frequently, become disheartened and give up (Ladd, 2016); or converge towards business ideas that are bounded by customers' explicit feedback, and thus obvious in nature (Felin *et al.*, 2019).

prevent scaling (Ladd, 2016)—is indeed correct, highlighting the dangers of hypothesis proliferation as an end in itself.

Second, we identify *boundary conditions* for the method. While it is beyond the scope of this study to examine all possible boundary conditions, our findings clearly show their importance for the method and its results. In particular, we find that certain teams embrace the method, while others tend to resist it. In particular, teams with formal *business education* (i.e., MBAs on the team) are the least likely to formulate hypotheses and to converge on a business idea. We suggest that this is because learning-by-thinking, which is the staple of MBA programs, can limit the appreciation of learning-by-doing methods like lean startup. However, compared to teams with no business education, MBA teams who *do* engage in hypothesis probing also formulate more new business ideas and their corresponding hypotheses *and* achieve idea convergence faster. MBAs thus may be particularly apt at some components (such as interpreting data from probing) if they engage with the method in the first place.

## RESEARCH BACKGROUND

Prior research outlines two main approaches—learning-by-thinking and learning-by-doing—that early-stage teams can use to iterate their business ideas (Gans *et al.*, 2019; Ott *et al.*, 2017). Iteration by learning-by-thinking, that is, thinking through potential consequences for alternative firm actions before choosing, received much early attention (e.g., Delmar and Shane, 2003; Gavetti, Levinthal, and Rivkin, 2005). Recently, scholarly attention has shifted to learning-by-doing, in other words, trying potential alternatives before choosing, most notably categorized as trial-and-error, bricolage, and experimentation approaches. Table 1 summarizes the differences between the two learning approaches and their relation to the lean startup method.

Learning-by-thinking methods use mental representations of the environment to plan how a firm's solutions address problems. A "cognitive structures" approach, for example, suggests that early teams form better strategies when they have a holistic understanding of opportunities, markets, and their own firms, and think through the potential consequences for alternative firm actions before choosing (Felin and Zenger, 2009; Ott *et al.*, 2017). An exemplar is Delmar and Shane's (2003) study of a random sample of 223 Swedish ventures, which finds that ventures with a holistic business plan are more likely to introduce products and less likely to fail. "Analogies," in turn, are mental representations of past solutions used to solve current problems. The case teaching approach common in business schools is an example. Although the lean startup features a blueprint of the environment (i.e., nine areas visualized in the canvas), in general it shares very few similarities with learning-by-thinking methods (see Table 1).

Learning-by-doing methods, in contrast, are based on taking action and learning from experience. There are three primary approaches: trial-and-error, bricolage, and experiment. In trial-and-error, many ideas are tried (typically in the neighborhood of existing good solutions) in the hopes of stumbling upon a promising one (Bingham and Davis, 2012; Katila and Ahuja, 2002). An analogy may be made to evolutionary theory, where rounds of variation and selection produce adaptive results over time—although trial-and-error is less "blind" than natural selection is. Callander (2011) says, "Trial and error search begins where theory insight ends… but is not blind as agents are using experience—accumulated information from successes and failures—to guide future choice." Lean startup resembles trial-and-error as it too accumulates information (i.e., interviewees' opinions) to gauge business ideas' potential success or failure. If feedback from the initial interviews is positive, focus is expanded in that direction; if not, focus is moved elsewhere and new hypotheses are formulated and probed.

Bricolage is another commonly studied learning-by-doing approach. In bricolage, solutions that already exist at hand are combined in new ways (Baker and Nelson, 2005). Effectuation is a close variant. Effectuation processes, like those in bricolage, imply that ventures draw on their knowledge and networks to select between possible effects that can be created using this known set of solutions (Sarasvathy, 2001). In lean startup, weekly business-idea iterations have some resemblance to bricolage (i.e., trying out different combinations until a solution is found).

Experiment, the third major learning-by-doing approach, is based on testing hypotheses. Experimentation approaches include theory development; causal propositions; and controlled variation of activities with somewhat controlled circumstances for testing them (Aghion *et al.*, 1991; Camuffo *et al.*, 2019)—or, alternatively, more light-weight experiments lacking controlled circumstances or clear theory (Miner, Bassoff, and Moorman, 2001). Again, the lean startup method shares several similarities with experiments, but also differs: the business-model canvas acts as a rough outline of areas where learning is needed (but no explicit theory is stated), and hypotheses are formulated and probed with customers (but not under controlled circumstances, nor systematically). Like many experiments (e.g., Andries *et al.*, 2013), it also embraces parallel probing of hypotheses.

Lean startup, as we have suggested and will now discuss in detail, is a blend of these three "learning-by-doing" approaches, particularly drawing from experimentation.

---Table 1 and Figures 1a and 1b about here---

**The Lean Startup Method**

Popularized by practitioners Eric Ries and Steve Blank (Blank, 2003, 2013; Ries, 2011), the lean startup method aims to iterate business ideas, helping entrepreneurs make an early decision about their feasibility. To that end, it encourages entrepreneurs to make explicit their assumptions about a business idea (i.e., formulate hypotheses) and then probe them (i.e., interview customers).

One of the main tools of the lean startup is the business-model canvas introduced by Osterwalder and Pigneur (2010). Canvas provides structure for the teams to articulate their assumptions about the business idea in nine pre-specified areas.[5] In particular, the teams use it to visually depict the goods and the services the firm plans to offer to customers, the parties participating, and the ways in which the ownership of goods and services is to be exchanged. These articulated assumptions then serve as hypotheses for the team to probe by "getting out of the building"—another significant feature of the method. Probing by interviewing customers and other stakeholders is done to confirm or disconfirm hypotheses. Through hypothesis-based probing, the lean startup method promises to help early-stage teams reduce uncertainty about the viability of ideas. It may also encourage them to "pivot," that is, to find a different idea when the initial idea is not confirmed. See Figure 1a for visualization of the method.

Despite some skepticism (Gans *et al.*, 2019; Ladd, 2016; Felin *et al.*, 2019), the lean startup method has been broadly embraced by entrepreneurs and corporate intrapreneurs, is taught widely in startup accelerators, and is embraced by policymakers in different countries. For example, Start-Up Chile, a renowned ecosystem startup accelerator created by the Chilean government (Clarysse, Wright, and Hove, 2015) with proven impact on the performance of participating startups (Gonzalez-Uribe and Leatherbee, 2018), encourages the use of the method by its members. Our online appendix summarizes the research on these initiatives.

Notwithstanding its popularity, the method's key assumptions have not been subject to rigorous empirical analysis. This is a surprising gap, as examining the key assumptions and components of the method is important to successfully implementing and improving it. Another

---

[5] The components of the canvas include value proposition, customer segments, exploitation activities, strategic partners, resources, customer relationships, distribution channels, revenue streams, and cost structure.

key gap in our understanding is the boundary conditions. Because composition of the team influences its learning approach (Katila and Shane, 2005; Jung and Shin, 2019), the team's background is particularly relevant for identifying the boundaries for the use of the method. Prior research has tied educational backgrounds to different styles of learning, for example, including some teams deciding *not* to engage in learning, even when provided with evidence that supports change (Hallen and Pahnke, 2016). Analyzing how the lean startup method can be leveraged or be limited by team composition is important for identifying the boundaries of the method.

## HYPOTHESES

To better understand the steps of the lean startup method, we test the underlying assumptions of the method (H1-H4) followed by potential boundary conditions, particularly regarding the composition of the teams that use the method (H5-H8).

### Lean Startup Method and the Underlying Assumptions

**Formulating hypotheses.** In H1 we test the first underlying assumption of the lean startup method regarding hypothesis formulation. The assumption is that formulating several different hypotheses about the business idea serves as a nudge for the team to gather feedback on its idea from different perspectives. In particular, teams that formulate more hypotheses are proposed to probe the hypotheses more (i.e., interview more).

The practice of formulating several hypotheses—i.e., breaking up problems and the proposed solutions into the nine areas of the business-model canvas—is believed to activate cognitive processing in teams. By explicitly stating a hypothesis, teams are expected to "turn on" a mindful, conscious activity rather than rely on habitual patterns of thought (Louis and Sutton, 1991). Sowing a seed of doubt in the team's mind about an assumption that was previously taken for granted (and thus ignored) is believed to help the team reconsider its assumptions. In particular, such a switch to active thinking is thought to "seed the search landscape" more broadly, highlighting areas in the team's opportunity landscape that are not fully understood. This process is then assumed to push the team to note the many elements in the environment which now warrant greater probing.

> ***H1.*** *Number of hypotheses formulated in a given period will relate positively to hypothesis probing in the following period.*

**Probing hypotheses.** In H2 we test another underlying (although rarely discussed) assumption of the lean startup method: namely, that probing is positively related to inspiration to formulate new hypotheses. Investigating one hypothesis helps the team think of new ones. This relationship is important to test as it contradicts much of the entrepreneurship literature that portrays listening to current customers and related stakeholders as counterproductive to forming new ideas (e.g., Felin *et al*., 2019). One of the main tenets of lean startup is that interviews help the team to learn about the feasibility of the hypothesized ideas and, if necessary, pivot to new ones. H2 tests this less discussed path to pivoting: that new information from interviews triggers new hypotheses that teams would not naturally think of in the absence of the method (from ii to iv in Figure 1b).

First, because probes are tied to hypotheses, the new information they generate often makes the team revisit the original cause-and-effect relationship. Without an explicit causal hypothesis, new information from an interview can seem obvious because it is easy to rationalize a novel phenomenon if no prior assumptions exist (Davis, 1971). In contrast, revisiting a stated hypothesis is likely to result in richer analysis of new information, questioning of prior beliefs, and formulation of new hypotheses.

Second, by nudging teams to probe hypotheses with external stakeholders they would not normally communicate with, probing is believed to encourage idea-networking (Dyer, Gregersen, and Christensen, 2008), which allows teams to access novel information embedded in social networks different from their own. Thus, teams may learn about areas in the opportunity landscape that they were not originally aware of, prompting new hypotheses.

Third, probing of hypotheses is expected to go hand-in-hand with formulating new hypotheses because it is through the concrete process of probing that members of the team become more acutely aware of the views held by their colleagues. Not all members of the early-stage team may fully comprehend other team members' assumptions about the business idea. Probing thus has the potential to increase communication and idea-exchange within the team as interview data are unpacked on a weekly basis (cf. Edmonson *et al.*, 2001 on extensive communication that enables executives to understand surprising results more effectively). Increases in idea exchanges between team members are then likely to inspire team members to think about new ideas not originally considered. And, as new ideas emerge that need to be tested, teams will likely formulate new hypotheses-to-test.

***H2.*** *Hypothesis probing in a given period will relate positively to new hypothesis formulation in the following period.*

**Business idea convergence.** Another assumption is that probing and convergence on a business idea—a team's belief that a hypothesis no longer requires testing—go hand in hand. When the team probes, gaining more understanding of the hypothesis, it is consequently more likely to converge on treating it as a quasi-fact. There are two reasons for the proposed relationship (from ii to iii in Figure 1b). First, because probing is conducted with external stakeholders (such as customers), the lean startup method assumes that information gathered is "salient" and "vivid." As such, the information is assumed to attract the team's attention and mobilize the team to converge. As Li *et al*. (2012) note, because salient and vivid information departs from expectations and norms of the team, it is more likely to enter team members' consciousness and affect subsequent action. Thus, this information will push the team to discuss alternatives, focus on possible tensions, and converge on a decision (cf. Li *et al*., 2012; Edmonson *et al.*, 2001). In contrast, information that is perceived as similar to what the team already knows—such as information gathered from team members themselves—does not similarly call the team to action (see also Fiske & Taylor, 2008; Sullivan, 2010).

Second, because it involves outsiders, probing with external stakeholders is more likely to introduce novel data and information to the team. These novel data can then be used to resolve contradictory beliefs within the team and help dislodge barriers to convergence (Edmonson *et al.*, 2001). Taken together, when the team is contrasting the cause-effect relationship outlined in a hypothesis against the information gained through probing (i.e., vivid and new information), the assumption of the lean startup method is that the process can lead to a reduction in the team's perceptions about the uncertainty of a given idea, encouraging convergence (from ii to iii in Figure 1b). Moreover, as the team becomes ever more confident that they are reducing the uncertainty about their business idea, the assumption of the lean startup method, as proposed in H4, is that the need for new hypotheses should also diminish (from iii to iv in Figure 1b).

> ***H3.*** *Hypothesis probing in a given period will relate positively to business idea convergence in the next period.*
>
> ***H4.*** *Business idea convergence in a given period will relate negatively to new hypothesis formulation in the next period.*

Although we have so far discussed the lean startup process as a "one-size-fits-all" solution, the arguments above point to much of the method's functioning being related to teams' level of engagement with the learning-by-doing aspects of the method. In H5-H8 we focus on the composition of the team as a potentially critical boundary condition.

**Boundary Conditions: Lean Startup Method and the Composition of Early-Stage Teams**

Research finds that the background and prior experiences of early-stage team members shape the decisions and strategic decision-making processes that ventures follow (Beckman, 2006; Eisenhardt and Schoonhoven, 1990; Lerner and Malmendier, 2013). Early teams' educational background has been found to be particularly influential, at least partly because education is typically an intense, formative experience for individuals' thinking styles and worldview (Jung and Shin, 2019). Research finds, for example, that even long after graduation, MBA students closely follow academic theories taught in business schools (Priem and Rosenstein, 2000) and surgeon-CEOs' decisions about a venture's R&D strategies are strongly shaped by their medical education (Katila *et al.*, 2017).

In H5-H8 we argue that team members' educational background is particularly relevant for the use of the lean startup method. Prior research has tied educational backgrounds to different styles of learning, including some teams deciding *not* to engage in learning, even when provided with evidence that supports change (Hallen and Pahnke, 2016). Understanding how such sources of inertia may be tied to the team's educational background is important for fully understanding the power and limits of the lean startup method. In the language of opportunity landscapes, team members with a particular educational background may have a fixed "x" landscape in mind that they are reluctant to reshape, either because they are averse to using the method to explore the landscape or because they are married to their original starting position within it.

**Business education.** Business education of team members is particularly relevant for lean startup because much of the required curriculum in U.S. business schools emphasizes learning-by-thinking rather than learning-by-doing methods (Navarro, 2017). Thus, in H5 we propose that formal business education, which we operationalize as team members with an MBA degree, is likely to particularly limit probing as a core learning-by-doing step in lean startup.

First, in their education, MBAs are trained to use learning-by-thinking methods, that is, the use of analogies (case studies), frameworks, and synthesized information such as market reports, to quickly and efficiently map out industry structures and strategies (Gary and Wood, 2011; Gavetti *et al.*, 2005; Sarasvathy, 2001).[6] Corporate strategy, for example, a required course in 92% of the MBA programs in the U.S., heavily emphasizes learning-by-thinking tools (Navarro, 2017; Gavetti, Levinthal, Rivkin, 2005). Corporate strategy staples such as the five forces (Porter, 1980), transaction costs (Williamson, 1979), and capabilities (Teece, Pisano, and Shuen, 1997) teach students to cognitively blueprint, synthesize, and think through the dimensions of firms and their environments before they act. This repeated activation and use of learning-by-thinking methods in an early stage of one's career is particularly likely to strengthen the individual's subsequent emphasis on problem-solving using learning-by-thinking methods, possibly making them skeptical of other methods. This point is consistent with a study by Gary and Wood (2011) where the authors showed, using a sample of 63 MBA students, that MBAs are comfortable using "learning-by-thinking" methods, in other words, holistic mental models of the business situation. All in all, it is likely that team members with formal business education would be less likely to embrace learning-by-doing methods. A close parallel is when prior industry experience with legacy technologies limits executives' ability to switch to new technologies,

[6] Other examples include studies that have shown that management training (i.e., finance or operations) improved business exploitation and efficiency (Bloom *et al.*, 2013; Bruhn, Karlan, and Schoar, 2010).

even in the face of discontinuous change (Furr *et al.*, 2012).[7] We similarly propose that teams that are trained to use "legacy" methods may be skeptical of new learning-by-doing methods, and particularly of their distinct features such as probing.

Second, we propose that because MBA training is about business, the team is more likely to treat members with an MBA as domain experts in forming business ideas. Thus, they would likely give outsize emphasis to such members' preferences in setting and iterating the business idea. This tendency may reflect the authority principle (Cialdini, 2001), by which team members become less likely to challenge a member with business education, believing that the expert is more likely to identify the best solution. Either way, with an MBA on the team, we propose that the team will feel less compelled to apply probing, the core step of the lean method.

***H5.*** *Business education (i.e., MBA) on the team will relate negatively to hypothesis probing.*

In H6 we particularly focus on how business education (MBA) of team members interacts with the use of the lean startup method in terms of probing (see fig 1c). Research on learning-by-doing methods has singled out learning from data and experiments (i.e., probing) as particularly significant yet challenging for some executives. In particular, executives that lack relevant expertise to interpret data benefit less from experiments, because experimentation by an inexperienced team member is likely to be costly and error-prone (Thomke, 2003); and conversely, because there is some evidence that teams with more experience derive more benefits through better interpretation (cf. Koning *et al.* (2019) on A/B testing).

Although these prior studies were about experiments, not the lean startup, they indicate that team composition is likely related to how the results of probing are utilized. Formal business

[7] Prior industry experience is typically cited as the major barrier. For instance, Furr *et al.*'s (2012) analysis of photovoltaic companies over a 15-year period showed that if the CEO's prior experience was in the solar industry, the firm was more resistant to switch to new photovoltaic technologies. In contrast, CEOs with no industry experience were quick to embrace technological changes.

training can potentially interact with probing in several ways, depending on how team members use their formal training to process the data that are gained. One possibility is that the learning-by-thinking training makes MBAs skeptical of any interview data and thus could make them less likely to even start processing the data. In this scenario we would see a non-significant or even a negative interaction. Another possibility, which we propose in H6, is that learning-by-thinking training acts as a complement to learning-by-doing. In this scenario, MBAs who embrace the learning-by-doing method could use appropriate analogies and frameworks from their formal training (including how to interpret traditional focus group data) to interpret what they have learned through the interviews and from their interviewees, suggesting a positive interaction. Thus, it may be more difficult for the teams that lack business education to interpret the results from probing or become inspired with new ideas (cf. Katila *et al.*, 2017; Sorenson, 2003). In contrast, team members with MBAs are more likely to have relevant experience, helping them to interpret the results of probing more effectively:

> ***H6.*** *Business education (i.e., MBA) in the team will amplify the positive relation between (a) hypothesis probing and business idea convergence and (b) hypothesis probing and hypothesis formulation.*

**Diversity in education.** In contrast to the previous two hypotheses that focused on MBA-teams, in H7 and H8 we move to consider the education of all team members more broadly. Although we have prior understanding of how diversity (e.g., in industry experience; Furr *et al.*, 2012) makes teams more flexible users of learning-by-thinking methods, we do not understand well how diversity is related to the use of learning-by-doing methods such as lean startup. In H7-H8 we focus on the diversity regarding a team's education (i.e., range of degrees in disciplines including law, medicine, engineering, and business), as each degree is tied to different styles of learning and knowledge bases. In H7 we propose that educational diversity in the team is likely to prompt the use of the lean startup method, and probing in particular.

First, overconfidence in the team's existing approach and over-optimism regarding initial business ideas is more likely when the team is homogeneous. This is because homogeneous teams share a similar domain knowledge and are likely to consolidate their preexisting beliefs when observing that their teammates have the same beliefs. In contrast, more diverse teams are likely to be collectively less confident in their methods and in their starting position, because of the lack of consensus about a given domain of knowledge and how it relates to the opportunity landscape. Consistent with this argument, Almandoz and Tilcsik (2016) found that more diverse teams were more "skeptical," whereas in more homogeneous teams (where non-experts in the domain were a minority), it was "difficult… to challenge experts and consider alternatives." Therefore, we propose that the presence of educational diversity in the team would go hand-in-hand with a predisposition to engage in probing.

Second, educationally diverse teams can more easily envision access to a broader set of potential stakeholders, making it mechanically easier for such teams to engage in hypothesis probing (such as arranging meetings for interviews and absorbing insights). Ability to reach out to—and meet with—a greater number of stakeholders is thus likely to make these teams more receptive to the method and to probing in particular.

Third, we propose that, by having a broader knowledge-base and stakeholder arena to build on, educationally diverse teams are likely to have a broader view of the opportunity landscape. This broader view could help diverse teams see both needs and solutions from different perspectives (Gruber, MacMillan, and Thompson, 2013), making them more prone to probe. A more expansive view leads to more expansive questioning.

***H7.*** *Educational diversity in the team will relate positively to hypothesis probing.*

We further propose that educationally diverse teams that probe more will be able to interpret the data from probing better. This is because they are likely to have broader absorptive capacity—that is, their collective knowledge base is more heterogeneous. As a consequence, they are quicker to converge on an assessment about the soundness of their ideas. At the same time, we expect that probing across a broader and more diverse group of stakeholders will help diverse teams generate new hypotheses. Conversely, it is more difficult for teams that lack educational diversity to decide how to interpret results or become inspired by new hypotheses, as they are likely to have a narrower base to draw on. Thus, we propose:

***H8.*** *Educational diversity in the team will amplify the positive relation between (a) hypothesis probing and business idea convergence and (b) hypothesis probing and hypothesis formulation.*

## METHOD

### Research setting

**Lean Startup at I-Corps.** To test our hypotheses about the lean startup method, we use the U.S. National Science Foundation's Innovation-Corps (I-Corps) program as our setting. This Congress-approved program was launched in 2011 to offer the lean startup training to prior NSF grantees. The program is aimed at helping NSF-funded scientists and engineers to identify business ideas for their technology-based inventions. While implementation of the lean startup method varies widely across the world, NSF has strived to train in a highly standardized manner, strictly adhering to method's founding principles (Arkilic, 2019).[8] "We imposed a standard curriculum," an early program officer at NSF's I-Corps said in an interview. "We wanted uniformity… to get the innovation process to scale."

[8] Steve Blank advised the NSF on the program's design and implementation.

Standardized features of I-Corps include a standard curriculum, a lead instructor who trains all the trainers, standardized length (2-3 day in-person kickoff, 8-week online modules, 2 day in-person wrap-up), and uniform requirements for team formation. All I-Corps teams are also required to use the common LaunchPad Central platform (which incorporates the business-model canvas). Other uniform features include transparent reporting on the platform including details of hypotheses and interviews, enabling rapid feedback from instructors and subsequent iteration by teams; weekly reporting deadlines; parallel probing of multiple hypotheses per interview; and weekly performance goals. These features were imposed from the first training and have not changed (Arkilic, 2019).

The original motivation for NSF to use the lean startup method was the lack of knowledge among scientists (i.e., NSF grantees) about business and the related perennial challenge in academic entrepreneurship, namely, low rates of commercialization of technologies by scientists and engineers (Wright *et al.*, 2009). The learning-by-doing orientation of the I-Corps program was also deemed a particularly good fit for the educational background of NSF scientists.

**Program elements.** During the I-Corps, each team participates in an intensive 8-week program with a cohort of about 12-15 other teams. The program includes both on-site (in a local I-Corps node) and off-site weeks, and is frequently described by participants as "intense" and a "24-7 bootcamp" (post participant surveys).

Before the start of the 8-week intervention, all teams are required to submit the first version of their business idea using the program's online platform, LaunchPad Central. At this point, the teams' business ideas are likely to be ill-defined and imprecise. Nevertheless, they must be explicitly outlined for the team to start the program. In particular, all teams are required to articulate the underlying assumptions of their business idea in nine pre-specified areas of the business-model canvas, that is, to formulate the first hypotheses.

Program delivery is standardized. Once a week, teams attend a “flipped classroom session,” meaning that material on lean startup is mostly studied beforehand, with classroom time devoted to promoting deeper understanding. During the classroom sessions, the teams present weekly progress, and receive critique from other teams and the instructors.

The teams’ core tasks during the program are to formulate and log “hypotheses to probe,” the “number of weekly interviews”, the people interviewed (role and typically the name of the interviewee), the key insights from the interview, and changes in the team’s business-model canvas (e.g., confirmed or disconfirmed hypotheses, new hypotheses). It is mandatory to log each of these changes on the program’s online platform on a weekly basis. Instructors monitor the weekly progress and provide written feedback to teams through the platform.

Our detailed longitudinal data on the week-by-week evolution of each team’s business-model canvas (using LaunchPad Central) records the changes that teams undergo. For example, a team may start with an assumption of specific market segments that would benefit from their technology, only to remove or add market segments in the following weeks as a result of the probing conducted with potential business stakeholders.

**Probing**. Consistent with the lean startup method, the program strongly emphasizes the idea of “getting out of the building” to conduct customer and stakeholder interviews. A weekly target is set for the teams to conduct about 10 interviews with real potential business stakeholders (mostly customers) to probe the hypotheses, including parallel probing so that multiple hypotheses can be probed in a single interview. Teams are required to use more than one interview to “confirm” or “disconfirm” each hypothesis.

To qualitatively validate teams' probing efforts, we asked all 14 teams in one cohort to answer, on a weekly basis, the questions "What were the 2 to 3 most important decisions you made this week related to your project? Why did you make each decision?" We content-analyzed these responses to isolate the instances where teams referred to probing that they had conducted. We also counted the number of statements in each week that made a direct reference to the effort of gathering real information from the market (see tables A1a and A1b in the online appendix). Despite our prompt making no reference to probing, participants described it frequently. This analysis strongly validates probing as the central component in lean startup.

**Sample**

Our dataset consists of 1,061 early-stage team members (divided into 388 teams from 23 cohorts) that participated in the National Science Foundation's I-Corps program across the United States over eight weeks. While our sample spans several cohorts (we include cohort fixed effects), the teaching materials, supervision and feedback methods, length of the program, and evaluation protocols are standardized across all cohorts.

To construct our team-level characteristic variables, we handpicked data from team members' LinkedIn profiles and cross-validated with data from Crunchbase and pre-program surveys (e.g., Gonzalez-Uribe and Leatherbee, 2018). We gathered biographical and demographic information, including each member's gender, age, level of education, and prior work and entrepreneurial background. Teams with incomplete individual-level information were dropped. Our sample consists of 152 early-stage teams (381 team members from 16 different cohorts) and 1,216 team-week observations.[9]

[9] Observations drop to 1,064 when the lagged dependent variable is used in regression models.

Our sample design is particularly appropriate for our research objectives. All teams are early-stage but have a preconceived business idea (NSF-funded technology to be commercialized). Eligibility for I-Corps excludes teams with significant revenues and private financing, consistent with our focus on early-stage teams. NSF further mandates that each team have the following members: entrepreneurial lead, technical lead (PI), and an industry lead (mentor) reducing team heterogeneity. Further, we have complete data on all sample teams' decision-making regarding business ideas with *high levels of granularity* and in *real time*.

**Effort by teams.** Due to the competitive nature of the I-Corps program (participants must go through an application process and commit to an intense program-related workload), teams are highly motivated to spend considerable effort on iterating their business ideas during the eight weeks of the program. Altogether, over the course of the program, teams typically solicit and perform close to 100 interviews (estimated interview workload is 20-30 hours per week), give 10 presentations, and participate in individual meetings with instructors.

In addition to the documented intensity of teams' engagement with the program, both the pre- and follow-on surveys[10] conducted with the teams indicate that they are indeed building a company, and not merely engaging in an executive education course. On average, 55 percent of the teams in our sample had incorporated their company within six months after the I-Corps intervention; and about 80 percent of these firms were still operational 18 months after I-Corps and had sought additional funding.

**Entrepreneurial team background.** The I-Corps team members in our sample closely represent the broader population of *professional entrepreneurs* in science-driven businesses.

[10] We had access to two sets of data: our own surveys and those of NSF. Response rates per team in our follow-on surveys range from 75 to 95 percent for the six-month post survey. The NSF data was anonymized so we could not link respondents to their teams or to individual characteristics.

Based on our data, the average age of team members is 32 years old, and 23 percent are female. Slightly under 12 percent of the members have an MBA, 38 percent hold a Master of Science degree, and 50 percent have a Ph.D. (as is typical in science-intensive sectors; Belz *et al.*, 2018). At the time of the study, 24 percent are university professors, typically in sciences and engineering. On average, participants have held five previous job positions in the past. Twenty-four percent have previously held a corporate C-level position, and 30 percent of the participants had founded a company prior to joining the program.

As noted above, teams are required to have three members, with an occasional team having a co-entrepreneurial or a co-industry lead. Fourteen percent of the I-Corps teams have a member with an MBA, while 80 percent of the teams have at least one member with a Ph.D. Only 7 percent of the teams are entirely female, while 58 percent of the teams are entirely male.

**Measures**

**Core variables of interest.** There are three core variables of interest: hypothesis formulation, hypothesis probing, and business idea convergence. We measured *hypothesis formulation* as the number of new assumptions that each team articulated about their business idea on the program's platform in a given week.

We measured *hypothesis probing* by the number of stakeholder interviews a team conducts each week. Because we focus on how often the teams check the feasibility of their ideas with external stakeholders, the number of interviews by a given team is an appropriate measure. Detailed minutes of each meeting were uploaded on the program's platform and the quality was monitored by the program instructors, further bolstering the measure. In alternate tests we included a squared term of the variable with no changes in our hypothesized findings.

We measured *business idea convergence* as the number of items in a given week that were either removed from a team's business-model canvas (i.e., they were deemed invalid assumptions) or added to it without being associated with a hypothesis (i.e., they were considered a fact, not an assumption). Because teams register changes in their business-model canvas on a weekly basis, we are able to track items that are "removed" or "disconfirmed". By removing an item, teams are explicitly declaring that their perception of the validity of the item was incorrect or undesirable. In contrast, if teams *add* new hypotheses (hypothesis formulation) in a given week, we consider this a sign that they recognize they are still dealing with relatively uncertain elements. These new hypotheses are not included in *business idea convergence*, but rather are directly measured as part of the *hypothesis formulation* variable.

We measured each team's *business education* by a binary variable labeled *MBA on team* that takes a value of one if the team has at least one member with an MBA degree and zero otherwise. An MBA, or master's degree in business administration, focuses on providing expertise in management analysis and is accredited to ensure consistency and quality of education. We do not consider undergraduate business degrees or certificates of executive education programs because their curricula are less standardized than the MBA. During the time period of our study, as noted above, the "learning-by-thinking" methods were the staple of MBA education making *MBA on team* a particularly relevant measure.[11]

Like Harrison and Klein (2007), we measured diversity in expertise using a Blau index of *educational diversity*, based on the classification of each team member's highest degree (Doctor

[11] Note that team members with Ph.D. degrees have overwhelmingly received their doctorate training in science and engineering fields (i.e., typical NSF grantees) so they have *not* been taught the learning-by-thinking skills in business that are the staple of MBA education.

of Law, Doctor of Medicine, MBA, Ph.D., MS, BS). The index takes the value of zero if all members have the same degree and higher values indicate higher diversity.

**Controls.** We included several controls. Because bigger teams may be able to codify more of their assumptions, conduct more interviews, and have a greater capacity to change (Barker and Duhaime, 1997), we controlled for team size, measured as the number of members in each team. Moreover, we controlled for gender diversity, as diverse teams may be more likely to embrace new tools such as lean startup. Gender diversity was measured using a Blau index based on the number of same-gender team members.

We also controlled for the training cohorts using *cohort fixed effects*. Although the goal of the NSF is to provide a standardized training experience across all cohorts, we included cohort fixed effects to control for any unobserved differences not controlled for otherwise. In addition, we included *week fixed effects* to account for systematic program-level differences during the course of each 8-week program, and because teams may have become more skilled at hypothesis-based probing or idea convergence as they matured in the program.

**Statistical analysis**

The data consist of a panel of observations on team-weeks. We used negative binomial models as our primary models and Poisson models for robustness because our dependent variable is a count variable with overdispersion (variance exceeds the mean; Hilbe, 2007). To account for team heterogeneity (repeat observations of the same team across multiple weeks), we used Generalized Estimating Equations (GEE) regressions (Liang and Zeger, 1986).

We repeated the analysis with *fixed-effects Poisson regression* that attempts to account for time-invariant team heterogeneity with team fixed effects; we chose this method as it is generally regarded as a "truer" fixed effects model than fixed-effects negative binomial models. We also included a lagged dependent variable (e.g., *hypothesis formulation*) to control for time-variant unobserved team heterogeneity and to facilitate causal inference (Heckman and Borjas, 1980). We report all results with robust standard errors.

**Causal inference.** While few research designs can completely rule out reverse causality or other possible alternative explanations, we have tried to account for these issues in several ways. In addition to lagging the dependent variable, we also used random-effects analysis, which, unlike fixed effects, allows estimation of variables that do not change over time. Findings were highly consistent (available from the authors). We also used an *instrumental variables* analysis, as described in more detail in the online appendix, although we recognize its limitations. To further facilitate causal inference, we included a rich-control set and lagged our independent variables by one week. Finally, we conducted several post-hoc analyses to better understand the results.

## RESULTS

Table 2 shows the descriptive statistics and correlations for all variables. Each week, teams formulated 3 to 4 hypotheses, conducted around 10 interviews to probe hypotheses, and converged on 3 business ideas, on average. Overall, the independent variables show considerable variance, and the correlation matrix indicates low to moderate correlations among them (team size and MBA on team were entered separately and together, with no changes in the results).

---Tables 2-6 about here---

The results for the GEE negative binomial regressions are shown in tables 3 through 6.[12] All models include week and cohort fixed effects. Model 1 in all tables includes the control variables only. Of the control variables, path-dependencies are interesting: probing in the past is

[12] Because Poisson fixed-effects models drop time-invariant covariates, they allow for testing hypotheses 4 through 7 only. Results are consistent with the reported GEE negative binomial ones (results available from the authors).

not related to probing in the future, and formulating more hypotheses is related to fewer new hypotheses in the future. However, teams that converged in the past are likely to do so again.

Model 2 introduces hypothesis formulation, hypothesis probing, and business idea convergence variables in all tables, corresponding to H1-H4. Model 3 introduces MBA on the team (business education) and its interaction with the main independent variables (H5, H6). Model 4 repeats with educational diversity (H7, H8). We use the full model (Model 5 in all tables) to interpret the results because omitting any key explanatory variables can lead to bias in the estimation of the remaining parameters (Kennedy, 1998: 103). Model 6 replicates the full model but excludes teams with extensive pre-program joint experience (i.e., the roughly 20% of the teams that had formed before the start of the I-Corps program[13]). After excluding teams with prior joint experience, the results (Models 6 in all main tables) were consistent and strongly supported our original findings, as described in detail below.

Table 3 focuses on the relationship between hypothesis formulation and hypothesis probing (see Figure 1b). H1 argued that formulation would go hand in hand with probing. Our results did not support this expectation. The only exception is educationally diverse teams, for whom there is some indication that more hypotheses result in more probes compared to less diverse teams. We return to this unexpected finding in the Discussion.

Table 4 focuses on the relation between hypothesis probing and hypothesis formulation. In H2, we argued that probing would inspire teams to come up with new ideas not previously considered, priming new hypotheses. Our results confirm the hypothesis.

Tables 5 and 6 focus on convergence. We argued that probing hypotheses would help

[13] For this analysis, we collected objective data for our teams about whether a venture had been formed before the I-Corps program. For robustness, we triangulated our data with anonymized I-Corps participant surveys that NSF shared with us. In these self-reported data, 23% of the respondents report that a venture was formed before the start of the program, increasing confidence in our team-level data that reports a corresponding number of 20%.

teams gather information that would allow them to converge (H3), and converging on a business idea would result in fewer new hypotheses (H4). Our results support both H3 and H4.

H5 through H8 focused on team composition as a boundary condition. Models 3 and 5 in tables 3-6 test H5. Teams with MBAs formulate fewer hypotheses and are slower to converge on their business ideas, but the results on probing are not significant. However, the positive coefficient on the interaction between MBAs on the team and probing (tables 4 and 5) indicates that probing results in more new hypotheses formulated and more convergence in teams with an MBA member, confirming H6. These results are consistent with the argument that MBAs as business experts can possibly evaluate the results of probing more effectively.

In H7 we proposed that educational diversity is related with more probing, while H8 focused on the interactions. We do not find the expected effect on probing (H7) in our results. However, post-hoc analyses show that educational diversity is positively related with *major* business idea convergence (results available from the authors). We also find that probing results in more convergence and more new hypotheses in educationally diverse teams, as we expected (H8), although the results are only moderately significant.

Because standard errors can be inaccurately reduced in models with a lagged dependent variable, resulting in overstated significance, we also ran the models by *excluding the lagged dependent variable*. Our original findings stayed. Because standard errors were almost identical across the models that included or excluded the lagged dependent variable, artificially small standard errors are less likely to have influenced the significance of our main findings.

Our post-hoc analyses also provide evidence on long-term performance outcomes of lean startup. As reported in the online appendix, teams that engage more with the method (more hypotheses, more interviews) have more positive outcomes in the 18-month period following the lean startup intervention (venture foundings, employment), further lending credence to our findings. Other sensitivity analyses (reported in the online appendix) use instruments, and examine several alternative explanations for our findings including idea quality and teams' startup experiences, again confirming the original findings.

**DISCUSSION**

Our research started with the observation that, although the lean startup methodology is increasingly popular, its key assumptions and boundary conditions have been subject to little rigorous testing. Ours is one of the first empirical studies—if not the first—to pin down the assumptions underlying the method, as implemented by NSF's I-Corps, which arguably most faithfully represents its principles. Our results also highlight that the method is not a one-size-fits-all solution; rather, critical boundary conditions such as team composition may influence its functioning.

There are several key findings. First, our results challenge some common assumptions about the lean startup method while confirming others. They pinpoint central role of *probing* (i.e., customer interviews), confirming the significance of "getting out of the building" in the method. It is probing that motivates business idea convergence—the key tenet of the method. Probing also motivates new business ideas (with corresponding hypotheses), a channel that is not often discussed by the method's adherents or anticipated by its critics (Felin *et al*., 2019). Intriguingly, probing can be helpful for teams to dislodge their original vision of the business idea, as needed, and find a new plan. Thus, we confirm that probing is a central component in the lean startup method, for reasons that are both more and less obvious.

In contrast, hypothesis formulation is not as central as expected. In fact, teams that formulate more hypotheses subsequently probe *fewer* of them. One possible reason is that one crisp, well-formulated hypothesis may lend itself much better to probing than many vague and poorly formulated ones. If this reasoning is correct, it suggests a need to emphasize quality over quantity in hypothesis formulation. Emphasis on quantity seems particularly problematic given the benefits of probing outlined above. Another possible reason for the unexpected result hinges on differences in how different types of teams approach hypothesis formulation: perhaps MBAs think that if they have thought through a hypothesis, they no longer need to probe it, an interpretation that is consistent with MBAs' affinity for learning-by-thinking methods. In contrast, educationally diverse teams seem to embrace hypotheses and their probing more. Either way, the goal should be to formulate a meaningful, but not too large, number of hypotheses.

Our results also address some of the recent criticisms of the method. Lean startup is often criticized as a self-reinforcing loop from hypothesis formulation to probing and further formulation of new hypotheses, which may tire out and prevent scaling (Ladd, 2016). A closer examination of the results suggests a more nuanced interpretation however. The criticism may indeed be correct to highlight the need to formulate hypotheses only in moderation, but we also find that convergence leads to *fewer* new hypotheses; that is, there is a natural stopping mechanism. At least in the context of the NSF's I-Corps program, the concerns related to endless iteration with no clear end result in sight do not seem as alarming as previously assumed.

Our second key area of examination relates to boundary conditions. Our findings extend prior literature by suggesting that expertise using the "legacy" learning approaches can be a handicap in the team's use of newer approaches. Specifically, we find that MBA teams comfortable with learning-by-thinking methods are resistant to use the lean startup, a learning-by-doing method. In contrast, teams with no business education (no MBAs) are more likely to formulate new hypotheses and converge on new business ideas than business-educated (MBA) teams, suggesting more affinity for learning-by-doing methods. However, it is noteworthy that the lean startup can be an effective tool if business-educated MBAs try it in the first place, because probes by MBA-teams are more likely to lead to more new ideas (with the corresponding hypotheses) and more business idea convergence, suggesting that MBA training helps interpret data from probes. It is also intriguing that, while teams with an MBA are significantly *less* likely to conduct a *major* change in their business idea in general, they are *more* likely to make major changes if they engage in hypotheses probing (additional results available from the authors). This is particularly encouraging for corporate entrepreneurship, as many established corporations are likely to have MBA-led teams. So, to assuage Felin *et al*.'s (2019) fears, the use of the method appears to actually help MBA-teams make changes.

Fourth, our findings indicate that lean startup may be a particularly suitable method for technology-driven academics as a first step to start building their ventures. A core challenge is that academic scientists often "do not leave the building" (Franklin, Wright, and Lockett, 2001; Wright *et al.*, 2009), meaning that they err on the side of too little iteration and too little customer information. Prior work has suggested (Wright *et al.*, 2009) that connecting MBA-degree holding alumni with scientists could be a solution. Our results suggest that a more nuanced interpretation is in order and that future work should consider important boundary conditions for involvement of MBAs, particularly if learning-by-doing methods are used.

Finally, and more broadly, we provide both qualitative and quantitative data about *early-stage teams*. In particular, we help address this gap in quantitative work focused on these teams and their business-idea probing (cf. Demil *et al.*, 2015). Our rich empirical data (hand-collected longitudinal data, week-by-week granularity during the intervention, and a lagged dependent variable and fixed effects to address team heterogeneity) are key to a more fine-grained understanding of how early-stage teams go about deciding what business idea to pursue.

**Limitations.** As in any study, there are limitations. Because we studied real teams pursuing real business ideas, we were not given the freedom to assign a control group or to randomize, as one would have in a laboratory experiment. Thus, the causal interpretation of our results is limited. However, because all teams went through the same standardized intervention, as described above, we were able to isolate the effects of the method particularly well, shielding from potential contamination by the "social context" (e.g., interactions with the rest of the organization that might blur the results in larger established firms that use the method).

Despite the advantages of the setting, future work should examine the use of the lean startup method outside of a standardized intervention such as I-Corps. Future research could also identify the many other possible boundary conditions on which the key relationships in the method might hinge (e.g., team composition, feedback from instructors and stakeholders, team formation process, program length)—that is, precisely the elements that were relatively standardized in I-Corps but are likely to differ in other implementations of the method.

It is also possible that lean startup as a method is particularly suitable for engineers and scientists, for whom technological development is relatively well thought out but market needs are not (Franklin *et al.*, 2001), as is the case in I-Corps. If the conditions were reversed (e.g., teams lack technology expertise but understand the market well), the value of the method could be different. It is also unclear whether we can observe similar results in contexts where the solution is more evident for the team, such as tastier meals in a fast food restaurant, or where uncertainty about the opportunity landscape is low. Again, this is a path for future work.

Another possible limitation is that our sample teams are relatively small (about three). Although this is close to a prototypical team size in new firms (Eesley, Hsu, and Roberts, 2014), it is not clear how our findings extend to larger teams in more established organizations. Perhaps, in these larger teams, diversity plays a more prevailing role than what we found in our setting.

Moreover, we do not measure venture performance beyond the preliminary post-hoc analyses reported in the online appendix that indicate that more intense interviewing is related to more ventures formed. Performance effects of the method are thus ripe for more exploration.

Finally, our measure of business idea convergence is from the perspective of the teams. Therefore, we are not able to identify whether what teams are learning is actually valuable. These questions offer intriguing ideas for future research.

**Conclusion.** Although the scholarly community is still in its early stages of embracing lean startup as a learning-by-doing method for entrepreneurship, our research anchors the concept theoretically, and shows that it has potential to be an engine for innovation and change in organizations. Increasing consensus about theoretically grounded assumptions and boundary conditions of lean startup should help elevate the theory-practice understanding of the method and help improve it and similar methods in the future. Altogether, empirical research on entrepreneurial learning methods holds great promise in helping us understand better how new business ideas are formed and shaped.

## REFERENCES


Aghion P, Bolton P, Harris C, Jullien B. 1991. Optimal Learning by Experimentation. *The Review of Economic Studies* **58**(4): 621.

Almandoz J, Tilcsik A. 2016. When experts become liabilities: Domain experts on boards and organizational failure. *Academy of Management Journal* **59**(4): 1124–1149.

Andries P, Debackere K, van Looy B. 2013. Simultaneous Experimentation as a Learning Strategy: Business Model Development Under Uncertainty. *Strategic Entrepreneurship Journal,* **7**(4): 288–310.

Arkilic, E. 2019. Raising the NSF Innovation Corps. In *Does America need more innovators?* Wisnioski, M. Hintz, E. and Stettler Kleine, M. editors. The MIT Press: 69-82.

Baker T, Nelson RE. 2005. Creating Something from Nothing: Resource Construction through Entrepreneurial Bricolage. *Administrative Science Quarterly* **50**: 329–366.

Barker VL, Duhaime IM. 1997. Strategic Change in the Turnaround Process: Theory and Empirical Evidence. *Strategic Management Journal* **18**(1): 13–38.

Baron RA, Ensley MD. 2006. Opportunity Recognition as the Detection of Meaningful Patterns: Evidence from Comparisons of Novice and Experienced Entrepreneurs. *Management Science* **52**(9): 1331–1344.

Batova, T. Card, D. Clark, D. 2016. Challenges of Lean Customer Discovery as Invention. IEEE International Professional Communication Conference (IPCC), Austin, TX.

Belz AP, Giga A, Zapatero F. 2018. Early-Stage Financing for Academic Entrepreneurs.

Bingham CB, Davis JP. 2012. Learning sequences: Their existence, effect, and evolution. *Academy of Management Journal* **55**(3): 611–641.

Blank S. 2003. *The Four Steps to the Epiphany: Successful strategies for products that win*.

Blank S. 2013. Why the Lean Start-Up Changes Everything. *Harvard Business Review* **91**(5): 63–72.

Bloom N, Eifert B, Mahajan A, Mckenzie D, Roberts J. 2013. Does Management Matter? Evidence from India. *The Quarterly Journal of Economics*. Narnia **128**(1): 1–51.

Bruhn BM, Karlan D, Schoar A. 2010. What Capital is Missing in Developing Countries? *The American Economic Review* **100**(2): 629–633.

Callander S. 2011. Searching and learning by trial and error. *American Economic Review* **101**(6): 2277–2308.

Camuffo A, Cordova A, Gambardella A, Spina C. 2019. A Scientific Approach to Entrepreneurial Decision Making: Evidence from a Randomized Control Trial. *Management Science* (September).

Cialdini RB. 2001. The Science of Persuasion. *Scientific American* **284**(2): 76–81.

Clarysse B, Wright M, Hove J Van. 2015. *A Look Inside Accelerators: Building Businesses*. *Nesta*. London, February.

Contigiani A, Levinthal DA. 2018. Situating the Construct of Lean Startup: Adjacent 'Conversations' and Possible Future Directions. *Industrial and Corporate Change*.

Crilly N. 2018. 'Fixation' and 'the pivot': balancing persistence with flexibility in design and entrepreneurship. *International Journal of Design Creativity and Innovation*. Taylor & Francis **6**(1–2): 52–65.

Davis MS. 1971. That's Interesting: Towards a Phenomenology of Sociology and a Sociology of Phenomenology. *Philosophy and Social Science* **4**: 309–344.

Delmar F, Shane SA. 2003. Does business planning facilitate the development of new ventures? *Strategic Management Journal* **24**(April): 1165–1185.

Demil B, Lecocq X, Ricart JE, Zott C. 2015. Introduction to the SEJ Special Issue: Business Models Within the Domain of Strategic Entrepreneurship. *Strategic Entrepreneurship Journal*.

Edmondson, A., Bohmer, R. Pisano, G. 2001. Disrupted routines: team learning and new technology implementation in hospitals. *Administrative Science Quarterly*, 46(4):685-716.

Eesley CE, Hsu DH, Roberts EB. 2014. The contingent effects of top management teams on venture performance. *Strategic Management Journal* **35**(12): 1798–1871.

Felin T, Zenger TR. 2009. Entrepreneurs as theorists: On the origins of collective beliefs and novel strategies. *Strategic Entrepreneurship Journal* **146**: 127–146.

Felin, T. Gambardella, A. Zenger, T. 2019. Lean startup and the business model: Experimentation revisited. Long Range Planning.

Fiske ST, Taylor SE. 1991. *Social Cognition*, 2nd ed. McGraw-Hill: New York, NY.

Franklin SJ, Wright M, Lockett A. 2001. Academic and Surrogate Entrepreneurs in University Spin-out Companies. *Journal of Technology Transfer* **26**(1–2): 127–141.

Furr NR, Cavarretta F, Garg S. 2012. Who Changes Course? The Role of Domain Knowledge and Novel Framing in Making Technology Changes. *Strategic Entrepreneurship Journal* **6**(3): 236–56.

Gans JS, Stern S, Wu J. 2019. Foundations of entrepreneurial strategy. *Strategic Management Journal* **40**(5): 736–756.

Gary MS, Wood RE. 2011. Mental models, decision rules, and performance heterogeneity. *Strategic Management Journal* **32**(6): 569–594.

Gavetti G, Levinthal DA, Rivkin JW. 2005. Strategy Making in Novel and Complex Worlds: the Power of Analogy. *Strategic Management Journal* **26**(8): 691–712.

Gonzalez-Uribe J, Leatherbee M. 2018. The Effects of Business Accelerators on Venture Performance: Evidence from Start-Up Chile. *Review of Financial Studies*, **31**(4): 1566–1603.

Grimes MG. 2017. The Pivot: How Founders Respond To Feedback Through Idea and Identity Work. *Academy of Management Journal* **61**(5): 1692–1717.

Gruber M, MacMillan I, Thompson J. 2013. Escaping the prior knowledge corridor. *Organization Science* **24***:* 280-300.

Grose, T. 2014. To Market, To Market. *ASEE Prism*, 24(4): 22-27.

Hallen BL, Pahnke EC. 2016. When do entrepreneurs accurately evaluate venture capital firms' track records? *Academy of Management Journal* **59**(5): 1535–1560.

Harrison DA, Klein KJ. 2007. What's the Difference? Diversity Constructs as Separation, Variety, or Disparity in Organizations. *Academy of Management Journal* **32**(4): 1199–1228.

Heckman JJ, Borjas GJ. 1980. Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence. *Economica* **47**(187): 247–83.

Hilbe J. 2007. *Negative binomial regression*. Cambridge University Press: Cambridge, England.

Jung J, Shin T. 2019. Learning Not to Diversify: The Transformation of Graduate Business Education and the Decline of Diversifying Acquisitions. *Administrative Science Quarterly*, 64: 337-369.

Katila R, Ahuja G. 2002. Something old, something new: A longitudinal study of search behavior and new product introduction. *The Academy of Management Journal* **45**(6): 1183–94.

Katila R, Shane SA. 2005. When Does Lack of Resources Make New Firms Innovative? *Academy of Management Journal* **48**(5): 814–829.

Katila R, Thatchenkery S, Christensen MQ, Zenios S. 2017. Is there a doctor in the house? New ventures, expert end-users in organizational roles, and innovation. *Academy of Management Journal*, **60**(6): 2415–2437.

Kearney, M. 2019. Essays on Managing Innovation. Ch 5. PhD Dissertation. MIT.
Kennedy P. 1998. *A Guide to Econometrics*. Blackwell Publishers: Malden, MA.
Kerr WR, Nanda R, Rhodes-Kropf M. 2014. Entrepreneurship as Experimentation. *Journal of Economic Perspectives* **28**(3): 25–48.
Koning R, Hasan S, Chatterji A. 2019. Experimentation and Startup Performance: Evidence from A/B Testing. *SSRN Electronic Journal*.
Ladd T. 2016. The Limits of the Lean Startup Method. *Harvard Business Review*.
Lerner, J, Malmendier, U. 2013. With a little help from my (random) friends: Success and failure in post-business school entrepreneurship. *The Review of Financial Studies* 26(10): 2411–2452
Li, Q., Maggitti, P. G., Smith, K. G., Tesluk, P. E., & Katila, R. 2013. Top management attention to innovation: The role of search selection and intensity in new product teams. *Academy of Management Journal*, 56(3), 893–916.
Liang K-Y, Zeger SL. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* **73**(1): 13–22.
Louis MR, Sutton RI. 1991. Switching Cognitive Gears: From Habits to Mind to Active Thinking.pdf. *Human Relations* **44**(1): 55–76.
March JG, Simon HA. 1958. Cognitive Limits on Rationality. In *Organizations*. McGraw-Hill.
McDonald R, Gao C. 2019. Pivoting Isn't Enough? Managing Strategic Reorientation in New Ventures. *Organization Science*
McGrath RG, MacMillan IC. 1995. Discovery Driven Planning. *Harvard Business Review* **73**(4): 44–54.
Miner AS, Bassoff P, Moorman C. 2001. Organizational and Improvisation Learning: A Field Study. *Administrative Science Quarterly* **46**(2): 304–337.
Navarro 2017. The MBA Core Curricula of Top-ranked U.S. Business Schools: A study of failure? *Academy of Managament Learning and Education*: 7(1).
NSF 2020. https://www.nsf.gov/news/special_reports/i-corps/resources.jsp Accessed May, 2020.
Osterwalder A, Pigneur Y. 2010. *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. John Wiley & Sons, Inc.: Hoboken, New Jersey.
Ott TE, Eisenhardt KM, Bingham CB. 2017. Strategy Formation in Entrepreneurial Settings: Past Insights and Future Directions. *Strategic Entrepreneurship Journal* **11**(3): 306–325.
Parker SC. 2006. Learning about the unknown: How fast do entrepreneurs adjust their beliefs? *Journal of Business Venturing* **21**(1): 1–26.
Ries E. 2011. *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. *Crown Business*. Crown Business, New York.
Sarasvathy SD. 2001. Causation and Effectuation: Toward a Theoretical Shift from Economic Inevitability to Entrepreneurial Contingency. *Academy of Management Review* **26**(2): 243–263.
Shane SA. 2000. Prior Knowledge and the Discovery of Entrepreneurial Opportunities. *Organization Science* **11**(4): 448–469.
Sorenson O. 2003. Strategy as quasi-experimentation. *Strategic Organization*, **1**(3): 337–343.
Sullivan B. N. 2010. Competition and beyond: Problems and attention allocation in the organizational rulemaking process. *Organization Science*, 21: 432–450.
Thomke S. 2003. Experimentation Matters: Unlocking the Potential of New Technologies for Innovation. *Harvard Business Press*.
Wright M, Piva E, Mosey S, Lockett A. 2009. Academic entrepreneurship and business schools. *Journal of Technology Transfer* **34**(6): 560–587.
Zott C, Amit R. 2007. Business Model Design and the Performance of Entrepreneurial Firms.

*Organization Science* **18** (2):181–199.

**Table 1. Learning-by-thinking, learning-by-doing, and the lean startup methods.**

| | Learning-by-Thinking | | | Learning-by-Doing | | |
|---|---|---|---|---|---|---|
| **Key elements** | Cognitive model | Analogies | **Lean Startup** | Trial-and-error | Experimentation | Bricolage |
| **Definition** | Problem-solving by constructing mental representations to plan solutions that will address current problems | Problem-solving by using mental representations of past solutions to solve current problems | Problem-solving by hypothesis-probing using interview data | Problem-solving by trying many solutions in the hopes of stumbling upon a good one | Problem-solving by hypothesis-testing using controlled variation of activities and context | Problem-solving by combining resources that already exist in new ways to solve problems |
| **Key references** | Delmar & Shane, 2003 | Gavetti et al. 2005 | | Callander 2011 | Camuffo et al. 2019 | Baker & Nelson, 2005 |
| **Time Horizon** | Up-front, preplanned | | **Real-time, experiential** | **Real-time, experiential** | | |
| **Method** | Elaborate cognitive planning | | **Rapid iterative experimentation in pre-selected target areas** | Iterative trial-and-error experimentation | **Spotlighting: Targeted, narrowly focused experiments** | Experimentation with re-combinations of existing solutions to create new solutions |
| **Hypotheses** | N/A | N/A | **Formulation of hypotheses about solutions** | N/A | **Formulation of cause-effect relationships to test** | N/A |
| **Seed for problem-solving** | Some theory; prior packaged information about the environment | Past solutions | **Team's assumptions about solutions** | **Accumulated information from past solutions (local hill-climbing)** | Unclear: Theory, although not always | **Existing solutions and resources at hand** |
| **Mental representations of the environment** | **Environment is mapped and a vision for the future is envisioned** | Mapping of similarities between past and current environments | **Scaffolding: visualization of pre-identified components of the environment** | N/A | Unclear: Theory, although not always | N/A |
| **Role of theory** | Holistic framework may draw on theories | N/A | N/A | N/A | Causal theory may underlie hypotheses | N/A |
| **Evaluation criteria** | Intuition and thinking through the consequences of solutions that the firm proposes | | **Parallel probing; Hypothesis confirmation or disconfirmation by interviewees** | Comparison to past performance | **Hypothesis confirmation or disconfirmation** | Unclear: subjective assessment relative to aspiration levels |

Adapted from Ott et al., 2017 and Gans et al., 2019. Bolded areas indicate overlap between the lean startup and the existing methods.

**Table 2. Descriptive Statistics and Correlations**

| | | Mean | S.D. | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Hypothesis Formulation | 3.63 | 6.22 | | | | | | |
| 2 | Hypothesis Probing | 10.45 | 7.4 | .07 | | | | | |
| 3 | Business Idea Convergence | 2.97 | 5.01 | .00 | .28 | | | | |
| 4 | MBA on Team | 0.14 | 0.35 | -.01 | -.07 | -.01 | | | |
| 5 | Educational Diversity | 0.39 | 0.31 | -.01 | -.03 | .07 | .19 | | |
| 6 | Team Size | 2.51 | 1.06 | .01 | .06 | .08 | .44 | .29 | |
| 7 | Gender Diversity | 0.16 | 0.22 | -.01 | -.01 | .05 | .22 | .09 | .30 |

N=152 teams

**Table 3. Negative Binomial GEE Regressions on Hypothesis Probing (t)**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Hypothesis Formulation (t-1) | | -0.000 | -0.000 | -0.001 | -0.001 | -0.000 |
| | | (0.520) | (0.828) | (0.084) | (0.094) | (0.280) |
| MBA on Team | | | 0.005 | | 0.006 | 0.008 |
| | | | (0.518) | | (0.485) | (0.313) |
| MBA on Team x Hypothesis Formulation (t-1) | | | -0.001 | | -0.001 | -0.002 |
| | | | (0.193) | | (0.110) | (0.031) |
| Educational Diversity | | | | -0.002 | -0.003 | 0.000 |
| | | | | (0.736) | (0.646) | (0.940) |
| Educational Div. x Hypothesis Formulation (t-1) | | | | 0.001 | 0.002 | 0.001 |
| | | | | (0.066) | (0.036) | (0.059) |
| Gender Diversity | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.006 |
| | (0.830) | (0.838) | (0.842) | (0.824) | (0.818) | (0.406) |
| Team Size | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.003 |
| | (0.285) | (0.277) | (0.291) | (0.311) | (0.330) | (0.389) |
| Lagged DV | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | (0.630) | (0.596) | (0.531) | (0.610) | (0.525) | (0.979) |
| Constant | -0.176 | -0.175 | -0.175 | -0.173 | -0.173 | -0.140 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Observations | 1,064 | 1,064 | 1,064 | 1,064 | 1,064 | 833 |
| deviance | 806.8 | 806.7 | 806.0 | 805.4 | 804.4 | 677.9 |

Note: P-values in parenthesis, two-tailed tests. All models include week and cohort fixed effects. Robust standard errors. Model 6 excludes teams with prior joint experience.

**Table 4. Negative Binomial GEE Regressions on Hypothesis Formulation (t)**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Hypothesis Probing (t-1) | | 0.004 | 0.004 | 0.002 | 0.007 | 0.015 |
| | | (0.001) | (0.000) | (0.038) | (0.000) | (0.000) |
| MBA on Team | | | -0.206 | | -0.214 | -0.490 |
| | | | (0.059) | | (0.065) | (0.051) |
| MBA on Team x Hypothesis Probing (t-1) | | | 0.011 | | 0.009 | 0.030 |
| | | | (0.026) | | (0.071) | (0.047) |
| Educational Diversity | | | | 0.001 | 0.029 | -0.053 |
| | | | | (0.990) | (0.639) | (0.379) |
| Educational Div. x Hypothesis Probing (t-1) | | | | 0.001 | 0.000 | 0.004 |
| | | | | (0.707) | (0.917) | (0.081) |
| Gender Diversity | -0.073 | -0.007 | 0.084 | 0.071 | -0.005 | 0.003 |
| | (0.124) | (0.881) | (0.003) | (0.002) | (0.896) | (0.944) |
| Team Size | 0.002 | 0.034 | -0.004 | -0.007 | 0.040 | 0.112 |
| | (0.899) | (0.134) | (0.734) | (0.616) | (0.035) | (0.000) |
| Lagged DV | -0.011 | -0.006 | -0.010 | -0.011 | -0.007 | -0.010 |
| | (0.000) | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) |
| Constant | 0.001 | -0.544 | -0.242 | 0.006 | -0.427 | 0.310 |
| | (0.972) | (0.000) | (0.023) | (0.897) | (0.001) | (0.176) |
| Observations | 1,064 | 1,064 | 1,064 | 1,064 | 1,064 | 833 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Deviance | 1441 | 1371 | 1378 | 1431 | 1361 | 1139 |

Note: P-values in parenthesis, two-tailed tests. All models include week and cohort fixed effects. Robust standard errors. Model 6 excludes teams with prior joint experience.

**Table 5. Negative Binomial GEE Regressions on Business Idea Convergence (t)**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Hypothesis Probing (t-1) | | 0.012 | 0.010 | 0.006 | 0.006 | 0.005 |
| | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| MBA on Team | | | -0.203 | | -0.166 | -0.192 |
| | | | (0.000) | | (0.001) | (0.000) |
| MBA on Team x Hypothesis Probing (t-1) | | | 0.019 | | 0.018 | 0.015 |
| | | | (0.000) | | (0.000) | (0.000) |
| Educational Diversity | | | | -0.080 | -0.048 | -0.020 |
| | | | | (0.362) | (0.475) | (0.747) |
| Educational Div. x Hypothesis Probing (t-1) | | | | 0.010 | 0.006 | 0.006 |
| | | | | (0.030) | (0.052) | (0.052) |
| Gender Diversity | -0.006 | 0.036 | -0.029 | 0.070 | -0.033 | 0.007 |
| | (0.925) | (0.012) | (0.535) | (0.187) | (0.414) | (0.883) |
| Team Size | 0.033 | -0.011 | 0.062 | 0.034 | 0.033 | 0.051 |
| | (0.036) | (0.017) | (0.000) | (0.013) | (0.033) | (0.005) |
| Lagged DV | 0.004 | 0.007 | 0.009 | 0.007 | 0.007 | 0.008 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Constant | -0.822 | -0.024 | -0.897 | -0.907 | -0.814 | -0.644 |
| | (0.000) | (0.046) | (0.000) | (0.000) | (0.000) | (0.000) |
| Observations | 1,064 | 1,064 | 1,064 | 1,064 | 1,064 | 833 |
| Deviance | 1495 | 1607 | 1441 | 1454 | 1450 | 1140 |

Note: P-values in parenthesis, two-tailed tests. All models include week and cohort fixed effects. Robust standard errors. Model 6 excludes teams with prior joint experience.

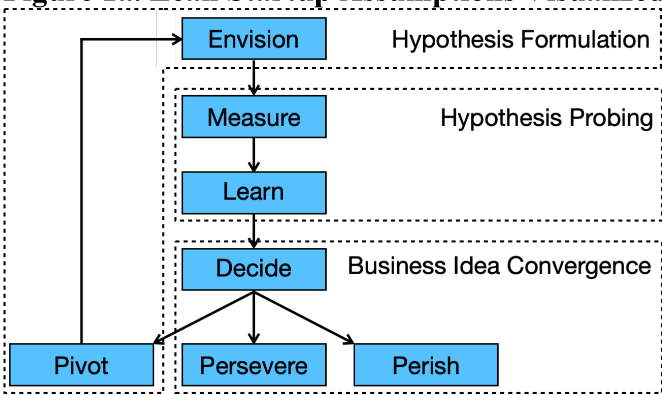
**Table 6. Negative Binomial GEE Regressions on Hypothesis Formulation(t)**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Business Idea Convergence (t-1) | | -0.006 | -0.006 | -0.009 | -0.006 | -0.003 |
| | | (0.012) | (0.013) | (0.000) | (0.011) | (0.166) |
| MBA on Team | | | -0.147 | | -0.135 | -0.329 |
| | | | (0.085) | | (0.131) | (0.050) |
| MBA on Team x Hypothesis Probing (t-1) | | | 0.006 | | 0.004 | 0.020 |
| | | | (0.121) | | (0.258) | (0.002) |
| Educational Diversity | | | | -0.082 | -0.049 | -0.169 |
| | | | | (0.105) | (0.389) | (0.093) |
| Educ. Div. x Hypothesis Probing (t-1) | | | | 0.012 | 0.005 | 0.018 |
| | | | | (0.000) | (0.045) | (0.000) |
| Gender Diversity | -0.073 | 0.017 | 0.023 | 0.018 | 0.011 | 0.009 |
| | (0.124) | (0.635) | (0.520) | (0.588) | (0.800) | (0.848) |
| Team Size | 0.002 | 0.043 | 0.047 | 0.024 | 0.042 | 0.071 |
| | (0.899) | (0.066) | (0.048) | (0.330) | (0.084) | (0.003) |
| Lagged DV | -0.011 | -0.004 | -0.004 | -0.005 | -0.003 | -0.008 |
| | (0.000) | (0.032) | (0.038) | (0.010) | (0.052) | (0.003) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Constant | 0.001 | -0.525 | -0.460 | -0.429 | -0.443 | -0.362 |
| | (0.972) | (0.000) | (0.002) | (0.004) | (0.003) | (0.049) |
| Observations | 1,064 | 1,064 | 1,064 | 1,064 | 1,064 | 833 |
| Deviance | 1441 | 1370 | 1366 | 1373 | 1361 | 1057 |

Note: P-values in parenthesis, two-tailed tests. All models include week and cohort fixed effects. Robust standard errors. Model 6 excludes teams with prior joint experience.

**Figure 1a. Lean Startup Assumptions Visualized.**




Adapted from Eisenmann, Ries and Dillard (2013).

**Figure 1b. Conceptual Model of H1-H4**

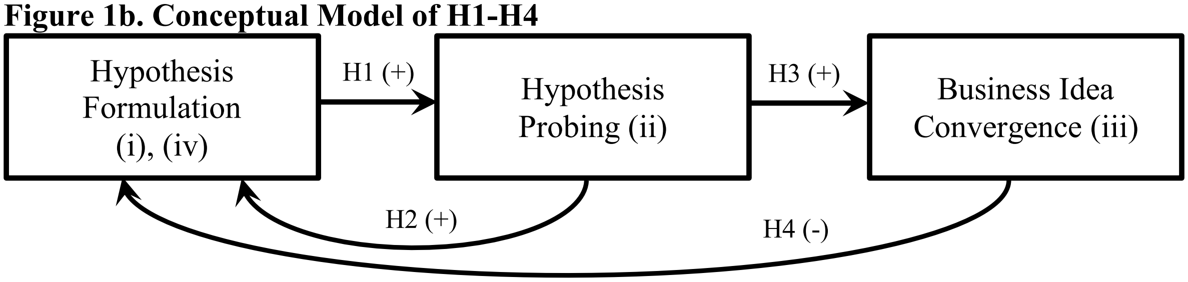
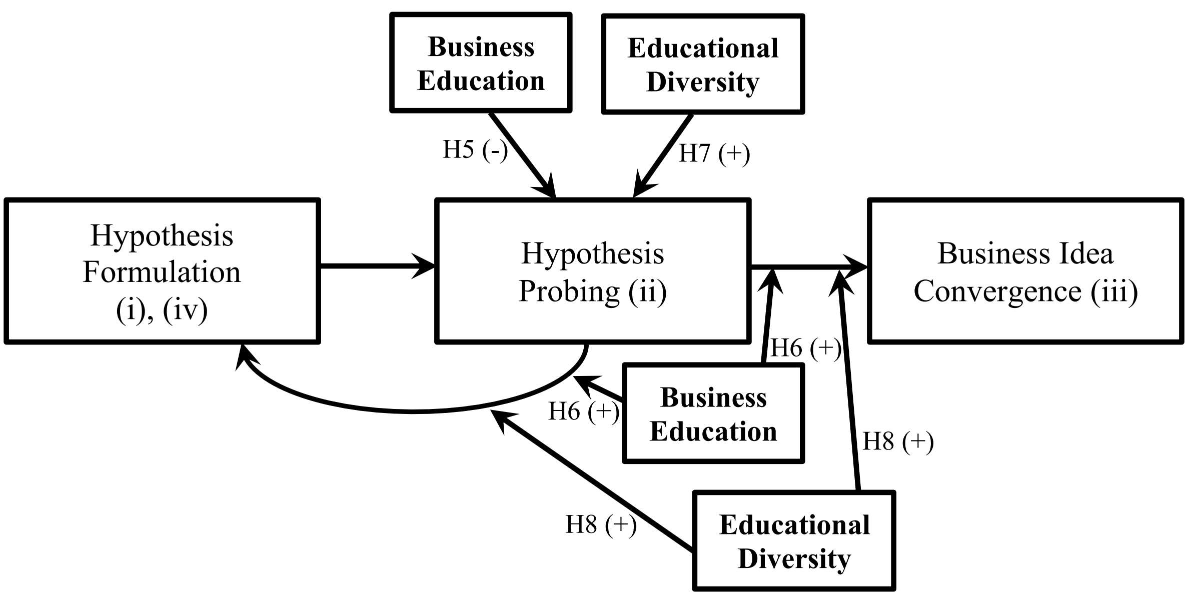



**Figure 1c. Conceptual Model of H5-H8**



## ONLINE APPENDIX

**Research on I-Corps**. Two streams of academic research have emerged around I-Corps. The first empirical stream examines performance outcomes of I-Corps training, and documents positive results. Descriptive data from the first several years of the program reveals that about 50% of the teams that went through the training started a company (NSF, 2020) and that about 10% of these companies received equity financing subsequently—rates that are higher than other venture acceleration programs (Grose, 2014). Belz, Giga and Zapatero (2018) similarly find that a relatively large percentage of the I-Corps teams that they surveyed received subsequent SBIR (27%), angel (20%), and VC (8%) funding. Finally, using a staggered design of the gradual rollout of I-Corps nodes in different states and a difference-in-differences design, Kearney (2019) finds that NSF grantees that went through the program had higher patenting and publication rates compared to similar NSF grantees that did not attend the program. Case studies also document broader, positive impacts on the entrepreneurial culture of a university campus (University of Wisconsin-Milwaukee, 2015). Overall, this stream points to positive performance outcomes.

The second stream is conceptual and raises several concerns about the method. Ladd (2016) argues that the method encourages endless iteration while Felin *et al*. (2019) in turn emphasize—in the context of the lean startup in general—the shortcomings of listening to current customers especially in projects that aspire to develop radical ideas that depart from what is already known. Batova *et al*. (2016) raise concern that academic research on customer discovery is not incorporated in the method. Overall, in two separate streams, prior research points to positive performance outcomes of I-Corps while also raising potential shortcomings.

**Sensitivity analyses.** We did several post-hoc analyses to better understand our findings. First, we included a ratio of team members with *prior entrepreneurial experience* (Shane, 2000). Some research suggests that individuals with more entrepreneurial experience may have a larger reserve of knowledge from which to build and so could be more engaged with hypothesis formulation and probing. Other research, in contrast, suggests that more experienced individuals (such as serial entrepreneurs) develop more rigid cognitive prototypes of what a valid business idea should look like (Baron and Ensley, 2006), and therefore would engage in fewer iterations and fewer interviews. This alternative control did not change our original results.

An alternative explanation for our MBA findings, is, of course, that business education helps teams to elicit *superior business ideas* (i.e., a starting-point strategy that is close to optimal one), thus requiring less iteration and less engagement with the method. However, our original explanation seems more likely, given that those MBA teams that *did* engage in probing in fact ended up converging more (Table 5). The OLS regression results in Table 7a also contradict the idea that MBA teams have superior initial ideas: These results that are based on our surveys within 18 months since the I-Corps intervention ended indicate that teams with an MBA are related with *fewer* ventures formed compared to teams that did not have an MBA. If MBA teams indeed had superior initial ideas, we would expect the opposite. Overall, it is likely that our MBA-results reflect resistance to change by teams steeped in "learning-by-thinking" methods, not better initial business ideas. A promising direction for future studies is to examine this intriguing finding using a regression discontinuity design or perhaps a randomized controlled trial at NSF to more accurately understand these outcomes.

Another interesting finding from our main results was that teams with an MBA (vs. those without) have higher rates of business convergence when they probe (interview) more. We suggested that this is perhaps because MBAs are able to interpret the findings from interviews more skillfully. To enhance causal inference regarding this finding, we examined whether endogeneity was present using Durbin and Wu–Hausman tests which were borderline significant. We then *instrumented* hypothesis probing and its interaction with MBA (by instructor engagements and cohort dummies as instruments), using two-stage instrumental variables limited information maximum likelihood regressions. Through this analysis, we attempted to consider unobserved factors that may simultaneously influence both hypothesis probing, and convergence. We used both ivregress and ivpoisson in Stata with no significant difference. Although the instruments are relatively weak as is typical of organizations research (F-statistic 8.9), instrumental variables results are strongly consistent with our original analyses, providing one further point of confidence in our findings.

**Table 7a. OLS Regressions on Venture Performance**

| | (1)<br>Venture Formed | (2)<br>Number of Employees |
|---|---|---|
| MBA on Team | -0.320 | -2.418 |
| | (0.017) | (0.403) |
| Educational Diversity | -0.139 | -3.542 |
| | (0.294) | (0.042) |
| Gender Diversity | -0.020 | -1.766 |
| | (0.908) | (0.531) |
| Team Size | 0.187 | 0.913 |
| | (0.001) | (0.399) |
| Constant | 0.701 | 10.692 |
| | (0.001) | (0.032) |
| Observations | 104 | 104 |
| R-squared | 0.597 | 0.314 |

P-values in parenthesis, two-tailed tests. All models include cohort fixed effects. Robust standard errors. Venture formed takes the value of 1 if a venture was formed, 0 otherwise. Number of employees is the number of jobs created within 18 months after the end of the program. Models exclude teams for which no information was found about the formation of a venture.

**Table 7b. Venture Performance and Lean Startup Method Intensity**

| | Hypotheses formulated | | | | Hypotheses probed | | | |
|---|---|---|---|---|---|---|---|---|
| | Below | Above-median | t-test | p | Below | Above-median | t-test | p |
| **Venture Formed** | 0.448 | 0.609 | -1.635 | 0.053 | 0.508 | 0.533 | -0.249 | 0.402 |
| | (0.501) | (0.493) | | | (0.504) | (0.505) | | |
| **Number of employees** | 2.534 | 3.978 | -1.17 | 0.123 | 2.424 | 4.156 | -1.489 | 0.07 |
| | (4.301) | (7.443) | | | (4.868) | (6.990) | | |

Teams were divided into above-median vs. equal or below median hypotheses formulated/hypotheses probed, at the cohort level. One-tailed t-test (Camuffo *et al.*, 2019). Standard errors in parentheses. Teams for which no information was found about the formation of a venture were excluded.

**Table 8a. Definitions and examples of instructor feedback.**

- **Positive Reinforcement:** A comment that supports the team's actions without guiding them.
  - Example: "I like the change you made in your Key Activities."
- **Negative Reinforcement:** A comment that opposes the team's actions without guiding them.
  - Example: "I don't like the change you made in your Key Activities."
- **Bossy Comment:** A comment that guides the team without explained reasoning.
  - Example: "You need to rework your Value Proposition."
- **Reflective Advice:** Advice that guides the team to think deeper about an issue or some information.
  - Example: "What have you learned from this interview?"

**Table 8b. OLS Regressions on Progress Feedback from Instructors**

| DVs | (1) Positive Reinforcement | (2) Negative Reinforcement | (3) Bossy Comment | (4) Reflective Advice |
|---|---|---|---|---|
| MBA on Team | 0.109 | 0.059 | 0.168 | 0.025 |
| | (0.266) | (0.210) | (0.078) | (0.876) |
| Educational Diversity | 0.081 | -0.003 | -0.029 | -0.200 |
| | (0.439) | (0.953) | (0.664) | (0.199) |
| Gender Diversity | -0.200 | 0.018 | 0.263 | 0.256 |
| | (0.084) | (0.752) | (0.017) | (0.254) |
| Team Size | 0.017 | -0.007 | -0.030 | 0.073 |
| | (0.627) | (0.585) | (0.251) | (0.173) |
| Constant | -0.238 | -0.079 | -0.124 | -0.340 |
| | (0.067) | (0.077) | (0.160) | (0.105) |
| Observations | 1,064 | 1,064 | 1,064 | 1,064 |
| R-squared | 0.124 | 0.070 | 0.137 | 0.255 |

P-values in parenthesis, two-tailed tests. All models include team and cohort fixed effects. Robust standard errors. Total number of teams is 152, observed through 7 weeks.

**Intensity of effort and post-I-Corps outcomes.** We also wanted to understand whether the intensity of the team's "learning-by-doing" effort in I-Corps made a difference. That is, we wanted to examine whether the intensity of using the lean method (hypotheses formulated, probing conducted) was meaningfully related to different post-intervention outcomes for the teams. To do so, we split our sample into two groups (more vs less than median number of hypotheses or interviews per team; see Camuffo *et al.* (2019) for a similar analysis regarding scientific experimentation). One group comprised all the teams that had less effort than the median team, and the other group had teams with more than median effort. In order to account for heterogeneity across cohorts, we used cohort-level medians to construct the groups. We then compared these two groups on two post-performance outcomes for each team: whether a venture was formed, and the number of jobs created. Simple t-tests in Table 7b suggest that teams that use the lean method more "intensively", that is, formulate more hypotheses, and probe them more, are more likely to start a new venture, and create ventures with more jobs, respectively.

**Idea quality.** Next, we wanted to further explore heterogeneity in teams and the perceived quality of the teams' ideas. We first collected and analyzed additional data on *instructor feedback* for the teams. We hand-coded feedback that each team received from instructors weekly, and then examined how characteristics of the team were related with the type of feedback the team received. We categorized feedback into four: positive, negative, bossy, reflective (see Table 8a for definitions and examples). The results indicate that teams with an MBA are more likely to receive a “bossy comment” from instructors (Table 8b). This finding is consistent with the idea that instructors factor their feedback differently to teams that need a stronger nudge to engage with the method.

We also used instructor feedback to better understand whether an alternative explanation for our empirical findings is that the composition of the teams goes hand in hand with the *quality of the business idea*, which could be reflected in instructor feedback. Although no objective measures of business idea quality were available, teams with better ideas possibly receive more positive, and fewer negative reinforcements from instructors. However, our empirical analyses show that feedback is not a significant predictor (main analyses or performance), and thus suggest that idea quality is probably not the chief explanation. Moreover, in Table 8b, we show that team characteristics (other than the MBA in the team) are not related with the instructor feedback regarding the ideas.

**Magnitude of business idea changes.** Finally, to give a more intuitive, live feel of the business ideas, we analyzed data on different types of business idea convergence of 111 teams in our sample for which the detailed data were available. We classified each change in business idea from previous week to next as minor vs. major change. We gave the two independent coders instructions to define *minor changes* as “a cognitively local, exploratory change compared to the previous week”, and *major changes* as “a cognitively distant change compared to the previous week.” Examples of changes we coded as minor include: (1) The channel in the business-model canvas has two items initially—“Retail Stores” and “Healthcare Providers”—and subsequently “Retail Stores” is removed; and (2) An item “Monthly Subscription Fee” is added to the revenue stream in the business-model canvas which previously only had items related to percentages of total revenue earned. Examples of changes we coded as major include: (1) The value proposition in the business-model canvas is changed from “fundraising for entrepreneurs” to an “online buy-sell gaming platform”; and (2) The customer segment in the business-model canvas is changed from one item “Smartphone users who attend Burning Man” by adding the item “Parents who want to teach their children about their emotions”. The 111 teams engaged in 183 major and 978 minor changes during the program as documented on the business-model canvas. As expected, roughly one of every seven changes (15.7%) were major changes, and 84.2% were minor. This is consistent with the notion that minor changes are easier, more likely, and more frequent than larger changes. Our results (available from the authors) strongly support our main findings. Teams with an MBA are significantly *less* likely to conduct a major change in their business idea (89.8% decrease in major changes relative to teams without an MBA), unless they engage in hypotheses probing, which significantly increases the likelihood of making major changes (by 8.8%).

### Table A1a. Sample statements that refer to interviews in medical-focused teams

| | | |
|---|---|---|
| "*Previously we had problems getting pharma interviews, but now no issues. Just have to be persistent.*" | "*[the team's solution] that can be delivered in the waiting room is desired by nearly all of the clinicians we spoke with*" | "*Decided to let [team member name], our large-animal vet, talk to dairy producers and other veterinarians. They are more receptive to her calls and she speaks their language*" |
| "*we interviewed several providers—self-employed and clinic—both are very interested in increasing their revenue through increased customer acquisition*" | "*we spent time at a community hospital and see that gaining access to those patients is problematic and gated at reception*" | "*we decided to build a minimum viable product because it is clear to us that we are not getting as much information now out of our interviews and cannot proceed further with getting more data out of [target customers]*" |
| "*Based on feedback from interviews regarding spoilage bacteria and also mention of these as an issue at the [market stakeholder name] meeting, considering major pivot and looking more closely at spoilage bacteria in processed foods as market*" | "*[…] went in person and based on the facilities they have available […] and knowledge they have in house especially regarding ideas for future products or use of our platform, they are the right partner to use for not just animal trials but lots of other research too*" | "*patients kept telling us that this is a significant value proposition and differentiated from other offerings*"<br><br>"*decided to add coaches as a customer segment because our interviewees light up when we talked about them*" |

### Table A1b. Teams' statements about interview efforts (2-3 Key Decisions Survey)

| | Week | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Team** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| 1 | 1 | | 2 | 2 | | 3 | 4 | 3 |
| 2 | | 1 | 1 | | | | | 1 |
| 3 | 2 | 2 | 2 | na | na | na | na | na |
| 4 | 2 | 2 | 1 | | 3 | 2 | 1 | 1 |
| 5 | | 1 | | 2 | 1 | 1 | 1 | |
| 6 | 1 | 1 | 2 | 1 | | 2 | 2 | 1 |
| 7 | 1 | 1 | 2 | 3 | 1 | 2 | 2 | |
| 8 | 3 | 3 | 2 | | | | 1 | |
| 9 | 3 | 2 | 2 | 3 | 2 | | 1 | 2 |
| 10 | 2 | 1 | 3 | 1 | | | | |
| 11 | 2 | 1 | 3 | 1 | 2 | 1 | 2 | 2 |
| 12 | 1 | 2 | 1 | 1 | 1 | 1 | | |
| 13 | 3 | 2 | na | na | na | na | na | na |
| 14 | 3 | 1 | 3 | 1 | | | | |
| Sum* | 19 | 16 | 24 | 15 | 10 | 12 | 14 | 10 |

Note: Numbers indicate the count of independent statements reflecting a meaningful effort to gather market information through potential customer or stakeholder interviews. Teams answered the question "What were the 2 to 3 most important decisions you made this week related to your project? Why did you make each decision?" Except for teams 3 and 13, all teams responded to this question every week. Blank cells represent a count of zero references to interviews that week. The sum of all statements excludes teams 3 and 13 to maintain a comparable value across all weeks.