Research-Technology Management

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/urtm20

A Process Model for Early-Stage Experimentation to Accelerate Innovation

Jeanne Liedtka, Stefano Magistretti & Elizabeth Chen

To cite this article: Jeanne Liedtka, Stefano Magistretti & Elizabeth Chen (2024) A Process Model for Early-Stage Experimentation to Accelerate Innovation, Research-Technology Management, 67:4, 36-48, DOI: 10.1080/08956308.2024.2349834

To link to this article: https://doi.org/10.1080/08956308.2024.2349834

Published online: 27 Jun 2024.

A Process Model for Early-Stage Experimentation to Accelerate Innovation

Companies can use the process model and practices presented to make early-stage experimentation tangible, easy to implement, and scalable, and turn it into an embedded practice.

Jeanne Liedtka, Stefano Magistretti, and Elizabeth Chen

OVERVIEW: Innovation has become increasingly urgent for companies to maintain competitive advantage, yet deep uncertainty persists regarding the end results of their innovation efforts. Early-stage experimentation offers a risk management approach that can dramatically accelerate cost-effective innovation. Experimentation relies on the scientific method to craft and test hypotheses. Insufficient guidance exists regarding how to help managers harness experimentation in early-stage innovation projects. Based on our observations of experienced experimenters engaged in early-stage innovation projects in four distinct settings, we developed a process model that offers practitioners a structured and scalable approach to foster early-stage experimentation. Managers in all parts of the organization can apply the process model and operationalize experimentation. The article's contribution is twofold: we articulate specific practices associated with choices around *what is being tested, how it is being tested,* and *what is learned,* and we offer a model designed to help build managerial literacy in early-stage experimentation.

PRACTITIONER TAKEAWAYS:
- Conducting fast, simple experiments at the early stages of the innovation process offers a valuable way to better manage innovation costs and risks.
- Experimentation can be codified into a simplified and scalable process, using a 5-step process model proposed here, that can be practiced across organizational levels and functions.
- An organizational strategic capability for experimentation, developed in a systematic and scalable way, can enhance a firm's ability to adapt to change and exploit opportunities in an uncertain world.

KEYWORDS: Innovation, Experimentation, Management practices, Innovation capabilities

In managing innovation, managers face deep uncertainty, especially at the front end where the future is unknown, technology is not fully developed, and market validation can be difficult due to the limited accessibility of users (Durante et al. 2024). Experimentation is a scientific approach involving hypothesis crafting and testing. It offers a promising approach to address these challenges, accelerate innovation, and improve innovation project outcomes by offering a more

Jeanne Liedtka is the UTC Professor of Business at the University of Virginia's Darden Graduate School of Business. With research interests at the intersection of strategy and design, she has written eight books and multiple articles on the subject of strategy, innovation, and design thinking. She has also consulted with diverse organizations, including IBM, Samsung, NASA, The United Nations, and the government of Singapore. She coauthored her most recent book, *The Experimentation Field Book: A Step-by-Step Project Guide* (Columbia Business School Press 2024), with Elizabeth Chen, Natalie Foley, and David Kester. JML3S@virginia.edu

Stefano Magistretti is an assistant professor in agile innovation and design management at the School of Management, Politecnico di Milano. He also holds a senior research position in the LEADIN' Lab, the Laboratory of Leadership, Design, and Innovation. Within the School of Management, he serves as the codirector for the Observatory Design Thinking for Business and Futures: Sense Making by System Thinking. He has a PhD in management engineering. His research is situated at the intersection of design thinking and digital technologies, with a focus on the expanding role of design and design thinking in entrepreneurship and strategy. stefano.magistretti@polimi.it

Elizabeth Chen is an associate professor in the Department of Health Behavior at the UNC Gillings School of Global Public Health. She is also design thinking lead for Innovate Carolina, the university's unit for innovation, entrepreneurship, and economic development. She has more than a decade of experience as a teacher, public health practitioner, and researcher. She earned her PhD and MPH in the Department of Health Behavior at UNC Gillings. Her research interests are at the intersection of adolescent health, technology, and design thinking (specifically human-centered design). Her current research focuses on how to integrate design thinking mindsets and methodologies into public health training, practice, and research. lizcchen@unc.edu

DOI: 10.1080/08956308.2024.2349834

timely and cost-effective set of processes and practices that optimize resource use amid early-stage uncertainty. According to Narduzzo and Forer (2024, p. 1), "When uncertainty prevails, deliberately risking failure may be the only way to gain new knowledge and innovate." In essence, they argue that experimentation, with its recognition of the core role played by failure, may be *the* critical path to learning under conditions of uncertainty.

Traditionally, experimentation is associated with scientific research. Based on their thematic synthesis of the innovation literature, Baxter, Trott, and Ellwood (2023, p. 5) argue for a broader definition of experimentation and who engages in it: "*Experimentation* could be used to categorize a variety of innovation activities in addition to those technical experiments conducted by scientists and engineers. Thus, we interpret the equivocality of the literature regarding the influence of managers on failure as reflecting the experimental nature of the innovation management task. Experimentation is thus an important organizational learning mechanism and is occurring all the time during the innovation process."

Experimentation is now central to a broadening group of fields, including design management (Micheli et al. 2019; Magistretti, Ardito, and Messeni Petruzzelli 2021), business model innovation (Andries, Debackere, and Van Looy 2013), and entrepreneurship (Patel et al. 2015). Baxter, Trott, and Ellwood (2023, p. 7) observe that "failure on a small scale (at the project level, and ideally early)" is essential to a firm's ability to innovate successfully. Euchner and Ganguly (2014) argue that even though innovation is inherently iterative, "there is a sequence to the analyses and experiments." Thomke (2020) contends that a lack of a structured process and clearly specified best practices prevent experimentation from being widely operationalized and scaled in organizations.

The challenges companies face in using experimentation effectively, especially in the early stages of innovation, make it fertile ground for analysis by academics and practitioners alike. This study offers a granular view of a set of experimentation endeavors in four cases. We offer a detailed analysis of the process of experimentation as practiced by expert experimenters that recognizes three discrete bundles of activities relating to *what is tested*, *how it is being tested*, and *what is learned*. Based on our findings, we propose a process model and tools that practitioners can use to conduct early-stage experimentation with the goal of improving innovation outcomes. Thus, this article contributes to managers' ability to explore new ideas and uncover their value through a rigorous and scalable experimentation practice. It also fills an important gap in the academic literature about experimentation in innovation.

Theory of Experimentation

Researchers over the last decade have tried to frame processes and tools that can support new product development to increase iteration and thus reduce uncertainty in innovation (Cooper and Sommer 2018). Yet scholars are still researching how to support a more agile innovation approach, especially in the early stage of innovation (Magistretti and Trabucchi 2024). When the focus is on the initial assessment of the innovation proposed (von Hippel and Kaulartz 2021), managers are left with little guidance about how to structure their decision-making (Sanasi, Manotti, and Ghezzi 2021).

Experimentation is considered a scientific approach involving hypothesis crafting and testing (Magistretti et al. 2023). For our purposes, we adopt Euchner and Ganguly's (2014, pp. 37–38) definition: "A business experiment is generally short term and limited in scale, but it answers a question that is critical to the success of the business." They state that business experiments are "conducted in the real-world using prototypes, simulated user experiences, and short trials"; they are "designed to provide enough insight to allow us to validate (or invalidate) a crucial assumption"; and they are "generally short-term and limited scale but answering a question that is crucial to the success of the business" (Euchner and Ganguly 2014, p. 37). Ganguly and Euchner (2018, p. 29) outline four attributes of an experiment: "It is focused on a few well-chosen factors, ideally one; it measures outcomes against a concrete prediction (the hypothesis); it is simple and inexpensive to conduct (given its context); [and] it is designed to get answers quickly."

Experimentation is challenging to accomplish: it requires a fundamental shift in both mindset and behavior (Liedtka, Hold, and Eldridge 2021). Learning in action is an important factor in successful experimentation. Cognitive (Tversky and Kahneman 1992) and social psychologists (Dweck 2008; Higgins 1998) have documented that barriers are deeply grounded in the basic human desire to avoid error. Similarly, organizational cultures punish failure, when they need instead to embrace failure to allow for success in experimentation—and by extension, innovation. As Zaharee et al. (2021, p. 31) point out, "Even though failure is an essential part of learning, it carries a huge stigma. 'Fail-fast' has become a common mantra, yet companies struggle to implement processes that can systematically encourage failure and capture lessons learned from failures." Zaharee et al. (2021, p. 32) also argue that "capturing learning from failures requires a psychologically safe environment where project team members feel comfortable sharing their experiences and perspectives about unsuccessful projects." Edmondson (1999) introduced the concept of psychological safety and has also pioneered research on failure in organizational working groups (Cannon and Edmondson 2001), suggesting the need for a clear and compelling process and practices that provide both detailed instruction and psychological safety to overcome these deep-seated challenges.

The Case Study

This case study includes four cases where companies use experimentation in the early stages of their innovation processes to reduce uncertainty. Our research involved tracing the activities of experienced experimenters working on early-stage innovation projects in four organizations. The first case is a shoe subscription service offered by Nike for young athletes (under age 18). The second case involves three concepts by the UK-based Southwestern Railway (SWR) to improve its passenger service. The third case by the Project Management Institute focuses on offering greater value to members outside its core product. The fourth case is MyHealthEd, a technology nonprofit aiming to use technology to improve teens' mental, emotional, and behavioral health.

Methodology

Using a comparative case study method (Eisenhardt 2021), we study four cases from organizations where experienced experimenters worked with managers to use the experimentation process to reduce uncertainty in their innovation activities. Our research study fulfills the criteria set out by Goffin et al. (2019) for doing quality case research in innovation: the case method is appropriate given the exploratory nature of our topic; we selected the cases deliberately based on explicit criteria; we used multiple data sources, and multiple investigators performed the analysis, which was validated externally by interviewees.

We selected cases designed to increase heterogeneity, with differing industries and concepts, spanning non-profit startups to large corporations (Table 1). The cases selected are highly influential and inspiring for the intent of the investigation (Siggelkow 2007). The experiments studied were conducted by different teams, each with extensive experience in experimentation. In fact, all of the experimental work conducted met the criteria outlined by Euchner and Ganguly (2014) to distinguish well-designed business experiments.

We conducted 21 semi-structured interviews, lasting 60–90 minutes, with experimentation consultants and company managers (cofounder, business and service heads, project partners, and team members). Interviews covered the projects undertaken and involved multiple stakeholders at different points in time, increasing the heterogeneity in perspective. The protocol of the interview followed open-ended questions pertaining to project structure (How was the project organized?), project aim (What was the object and motivation

TABLE 1. Organizations, context, concepts, interviewee roles, and secondary data

Organization	Case Context	Interviewees (by title)	Secondary Data
Nike (global athletic apparel company)	Nike wanted to consider new services to encourage brand loyalty among young customers under the age of 18. **Concept:** *Easykicks:* a shoe subscription service for young athletes based on a technology platform	1. Sustainable Business Head 2. Engagement Lead 3. Project Partner 4. Team Member (2)	126 pages* (slidedecks from presentations detailing different phases of the project)
Southwestern Railway (SWR) (UK joint venture, operating extensive train network into London)	SWR wanted to improve the passenger experience. **Concepts:** *Wayfinding Audit:* addressed poor signage with smart customer-led tools. *Concierge* concept: aimed at offering a warm welcome and better infrastructure for ticketing *Cleaner and Cleaner*: a hygiene awareness and nudge campaign	1. SWR Customer Service Head 2. Station Manager 3. Research & Marketing Lead 4. Engagement Lead 5. Team Member (3)	282 pages* (slidedecks from presentations at different phases of the project, written reports, video transcripts of employee interviews)
Project Management Institute (PMI) (premier global project management association)	PMI wanted to add greater value to its members outside of its core product, the PMP certification. **Concepts:** *Snippets:* a microlearning platform *Career Navigator:* a self-assessment tool *Hive:* peer-to-peer connection platform *Spot*: an experiential learning opportunity	1. Product Manager, Global Membership 2. Engagement Lead 3. Project Partner 4. Team Member (2)	216 pages* (slidedecks from presentations detailing different phases of the project)
MyHealthEd (Technology non-profit aimed at using human-centered design to improve the health and well-being of youth)	MyHealthEd wanted to use technology to improve teen mental, emotional, and behavioral health. **Concepts:** *ExploreMe:* an app-based avatar experience with educational missions for youth to learn health content and earn badges *xHealth:* a story sharing app where youth can write and read stories related to health, respond to others' stories, and discuss health topics	1. Cofounders, MyHealthEd (2)	207 pages* (Slidedecks from presentations and feedback sessions, spreadsheets tracking tests and test results as well as story submissions, Miro frames, written reports, published manuscripts)

*Page totals do not include transcripts of interviewees listed by title in the previous column.

for the project?), project execution (How has the project been conducted?), and project delivery (What were the final results achieved?). We organized two follow-up meetings with selected interviewees in each case to discuss the evidence gathered. We recorded and transcribed all interviews.

We compiled all available secondary data related to each project, including presentation slide decks, innovation team notes, formal client reports, photos, videos, and news stories. These additional materials ensured triangulation of information (Eisenhardt 2021).

Independently, four members of the research team reviewed all transcripts and secondary materials to identify observations about the experimentation process and practices evidenced in each case. All the coauthors then collaborated to identify dominant codes across individual research team members' observations and cases. Then they shared these combined and consensus observations with the managers and consultants involved in the four cases to ensure consistency in the findings and transparency in the process.

Results

Our comparative analysis revealed that the innovators we studied across all four cases engaged in a similar set of specific activities across each development stage of the innovations being tested. These activities clustered around three key questions: *what* was being tested (specifying the hypothesis and the data needed), *how* it was being tested (selecting a test design), and *what the learnings were* from that test. We report our findings based on this sequence.

What Was Being Tested: Specifying the Hypothesis

When uncertainty is high and projects are in early stages, the managers we observed and interviewed shared that they pay careful attention to *what* is being tested. They treated the innovation concept as a hypothesis and focus on the data needed to test its validity. As one MyHealthEd team member explained, specificity in laying out both the hypotheses and their associated metrics was key: "You start by listing out the hypotheses you're trying to test and how you're going to measure it—getting really quantitative and specific. You're not just waiting to see what happens, you are intentionally measuring certain things, and then are able to reflect back and see whether that hypothesis proved to be true or not. That's how we get into the actual learning."

Whereas managers new to experimentation might start the process with product ideas alone, decoupled from identified target markets, or define targeted markets as broadly as possible, the experienced innovators we observed specified more complete value propositions and preferred narrower rather than broader markets for initial testing. The Nike team leader elaborated on their decision to focus testing on young athletes and their parents, rather than the broader market of the avid runner. "If this works for the mom and her young athletes, does it also work for the avid runner? . . . and wouldn't that equal a bigger, stronger, more valuable business? But in doing that, you neglect the thing that you're actually focused on," the team leader said. "I think it was the right call to focus on the young athlete and serving them and their parents very well instead of heading as quickly as possible to the avid runner."

We also found that participants preferred to move a *portfolio* of ideas into testing. As one PMI project team member explained, "The luxury of taking a portfolio approach is that you're not just saying yes or no to *this* concept. You're saying, 'Amongst these options, which one do we choose'? The tests informed each other. There was real value in seeing how they played with each other." Participants indicated that a portfolio approach discouraged converging prematurely on a single idea, and better managed risk by letting potential users tell them what worked best. This is crucial in early-stage innovation projects where innovation concepts are still under development. Project teams then prioritized that collection of ideas in order of testing urgency. For example, the MyHealthEd team generated many potential ideas to address their adolescent health challenge, including video games, board games, avatars, and social media campaigns, before selecting two—ExploreMe (an avatar-based experience) and xHealth (a storytelling app) for initial testing.

After identifying the subset of ideas that would move into further testing, expert experimenters across the cases consistently summarized the concepts in a short form of some kind. We present a summary prepared by the Nike team on the *Easykicks* concept (Figure 1). The Nike team felt this was critical to ensuring a consistent understanding of what was being tested among both team members and their clients.

What Was Being Tested: Specifying the Data Needed

The next stage ensured that the experienced experimenters we studied explicitly connected their ideas with the data needed to guide the rest of the testing process. Experimentors conducted a set of activities aimed at searching and defining the data needed for the testing process. As SWR's customer service head noted, "We needed to build a business case that this improved things for our customer and improved things for the organization—whether that was through efficiencies or cost savings, or maybe it's less complaints or improved NPS [net promoter score]." They further explained, "It can be difficult to nail these things down, so we broke it into bite-sized chunks to show the benefits, on a quick timescale and on a small budget. We built the business case and the momentum—and it wasn't going to take 18 months and millions of pounds."

When uncertainty is high and projects are in early stages, the experimenters pay careful attention to *what* is being tested.

TEMPLATE 3

Easykicks **Concept Snapshot**

	User 1	User 2
For (target user)	**Young Athletes** (8–12 years)	**Their Parents**
Who want (unmet needs)	• More control, autonomy, agency in their shoe shopping • The best shoes (providing the right fit, comfort, and performance) for various activities	• Convenient, more pleasant shopping experience (smoothing of shopping pain curve; smoothing the spikes) • Confidence that their kids have the right shoes for their activities
We will offer (offering)	A better, more positive shoe-buying experience with: • An accurate fit assessment (digital or in-person) • A sustainable, cyclical buy-return process • A more attractive purchasing mechanism (subscription based)	
That provides (benefits)	• A creative outlet for self-expression • Confidence in their personal style • Special/unique feeling as they perform	• Shoes their young athletes need, when they need them • Less time spent shopping • A reduction in the pain of shopping (i.e., fewer conflicts, less stress)
Uniquely (differentiation)	• A personal (1:1), ongoing relationship between the firm and young athletes (build stronger relationships than created with products only; make young athletes aware of the brand earlier) • A service experience wrapped around shoes • Cyclical shoe experience • A novel business concept	

FIGURE 1. Nike *Easykicks* concept summary

Identifying needed metrics began with surfacing the most critical assumptions—the "make or break" ones, as one of the PMI team members described them. These assumptions, which were the most crucial to the success of the concept, often related to desirability or feasibility. While novice experimenters might focus on testing the concept itself, the innovators we interacted with in our research insisted on focusing on underlying assumptions instead. This focus on testing "make-or-break" assumptions at the outset was sometimes a hard sell to clients, as a member of the PMI team explained: "People [clients] get uncomfortable focusing first only on the make-or-break assumptions. They want to get all the answers to everything at the same time. But we're going to focus on make-or-break assumptions, *not* on everything."

> Identifying needed metrics began with surfacing the most critical assumptions—the "make or break" ones . . . which were the most crucial to the success of the concept, often related to desirability or feasibility.

The SWR team believed testing key assumptions to be the most inexpensive, least resource intensive way to learn—noting they did not require a functional prototype initially, and that this facilitated rapid testing of components of the idea, rather than the entire concept. Once assumptions surfaced, SWR established the nature of the evidence needed to test them and set aspirational and threshold target values. We present a summary of assumptions and evidence prepared by the SWR team for the *Concierge* concept (Figure 2).

In each of the four cases, setting these metrics in advance was explicit and seen as important. SWR's customer service head said, "We set out what our success criteria were going to look like beforehand and shared that and only *then* went into the experiments—so after we went out and gathered the evidence, we could say, 'Yes, it's been successful' or 'No, it hasn't' and this is how it will inform our decisions going forward."

This insistence on agreeing on evidence sources and threshold values served several purposes. It allowed teams to focus on interim measures of directional correctness. "The core idea for setting the threshold metric is that you need to know what is good enough, right? You need to know that you are directionally correct. You don't need to be super

TEMPLATE 7

Assumptions to Evidence for SWR's Concierge Concept

Assumption	Evidence	Threshold Targets	Source
Creates a warm welcome that customers value	Frequency of customers seeking interactions with Concierge staff	Number of customers interacting with Concierge staff goes up 30 percent or more	Observation in ticket hall
	Positive comments from customers expressing appreciation for the service	60 percent of interviews have customers noting appreciation for the service when unprompted	Customer intercept interviews
Provides more efficient service for customers and enables improved interaction with SWR staff	Volume/speed/accuracy of Concierge staff answers	90 percent of Concierge staff answers meet customers'needs in terms of volume/speed/accuracy	Observation in ticket hall
	Length of queues at ticket windows	40 percent (or more) decrease in length of queues	Observation in ticket hall compared with historical record
	Increased volume of tickets purchased at self-service window	40 percent (or more) increase in volume of tickets purchased via self-service machines	Sales data from self-service machines compared with historical record
Increases SWR staff enjoyment of jobs	Staff appreciation for and interest in remaining in new role versus returning to ticket window	60 percent of staff have an appreciation for and express interest in remaining in new role	Staff interviews

FIGURE 2. SWR *Concierge* assumptions and evidence summary

precise—you just need to know that you're going in the right direction," explained a MyHealthEd team member. "We don't need to go all the way to the finish line, just to that next small step. We have to find ways to show progress without waiting for that final outcome." Echoing a similar sentiment, a Nike team member remarked, "We're always looking for just enough volume of similar data from a diverse enough group that we can say 'There's enough strength of signal there that we now feel good about moving forward.' Even if it's somewhat arbitrary, we're going to set a threshold."

These threshold-setting conversations also created an opportunity to bring influential stakeholders into the process of planning the testing journey and collaboratively identifying what success would look like. Doing so helped create consensus between all parties regarding what the important assumptions were and what evidence would confirm or disconfirm them. This action fostered success at later stages. For example, PMI had assumptions around desirability, feasibility, and viability that it wanted to test. For each, PMI articulated explicitly the kind of evidence it needed to gather to prove or disprove the assumption and meet threshold targets. One PMI team member said, "We set a decision-making framework at the onset. We agreed on criteria. They were pretty simple . . . but agreeing on that ahead of time allowed us to structure our tests to get the right kind of data . . . When we agreed to that, we set ourselves up for a really smooth decision-making process later on."

How Are We Testing: Selecting a Test Design

Once the experimenters we observed and interviewed figured out *what* they were testing, they transitioned to making a set of important decisions about *how* to test, including selecting the test design and creating an appropriate prototype (always in that order). The four cases we studied relied on five common test types: cognitive walkthroughs, lemonade stands, smoke tests, simulation, and trials. We summarize test types and outcomes for each case (Table 2) and provide a definition of each test type (Table 3).

Across the four cases, participants described common pitfalls to avoid—namely, inappropriate test types where impatient experimenters rushed to conduct unnecessarily lengthy and sophisticated tests, and prototypes with levels of fidelity in both form and function far higher than they needed to be, which meant more money and resources invested than needed to test the assumptions. As one Nike team member aptly explained, "We always start with 'what's the *least* we can do to return the most learning?' I think that's a critical design imperative that sets your choosing."

The test design selection process progressed through a series of questions that were sometimes articulated, and other times inferred. The innovators we observed all explicitly addressed the question of whether the concept should be tested as a whole versus broken down into its component parts. At MyHealthEd, one team member emphasized that dealing with the complexity of big ideas often required decoupling and then reassembling. "To build an MVP

TABLE 2. Concepts, test types, and outcomes by organizations

Concept to be Tested	Test Types	Outcomes of Tests
Nike *Easykicks:* a shoe subscription service for young athletes based on a technology platform	• Cognitive walkthroughs • Lemonade stands • Smoke tests • Simulations • Trials	The *Easykicks* concept succeeded through multiple rounds of testing and was scaled in the marketplace. Despite generating significant revenue, Nike later terminated it in favor of more promising opportunities.
SWR *Wayfinding Audit:* addressed poor signage with smart customer-led tools. *Concierge* concept: aimed at offering a warm welcome and better infrastructure for ticketing *Cleaner and Cleaner:* a hygiene awareness and nudge campaign	• Cognitive walkthroughs • Trials	All three concepts remain in advanced stages of testing. The *Concierge* concept, for example, moved through two significant initial trials, and is now being tested in SWR's larger station network.
PMI *Snippets:* a microlearning platform *Career Navigator:* a self-assessment tool *Hive:* peer-to-peer connection platform *Spot*: an experiential learning opportunity	• Cognitive walkthroughs • Lemonade stands • Smoke tests • Simulations	The *Snippets* and *Career Navigator* concepts were successful through multiple rounds of testing and launched as tools on the PMI website. They continue to be popular with PMI members. The *Hive* concept was eliminated early in round 2 testing due to desirability concerns on the part of users; *Spot* was also eliminated in round 2 testing due to feasibility concerns associated with its execution.
MyHealthEd *ExploreMe:* an app-based avatar experience with educational missions for youth to learn health content and earn badges *xHealth:* a story-sharing app where youth can write and read stories related to health, respond to others' stories, and discuss health topics	• Cognitive walkthroughs • Lemonade stands • Smoke tests • Simulations • Trials	The *ExploreMe* concept was eliminated early in round 2 testing due to desirability concerns on the part of users and feasibility concerns associated with execution. The *xHealth* concept was successful through multiple rounds of testing and evolved into *Real Talk*, an app where youth can anonymously write and read stories about health topics and be connected with high quality online health resources. The *Real Talk* app went through multiple rounds of testing first as clickable wireframes, then a web prototype, and finally a native iOS app. Real Talk was launched in the Apple App Store in 2016 and has served 35,000+ youth worldwide.

TABLE 3. Five types of field tests

Test type	Description
Cognitive walkthroughs	Cognitive walkthroughs help you find partners and customers by walking them through your concept in detail. They invite the prospective user to help you shape a still malleable concept, using tools like pitch decks, storyboards, or paper prototypes.
Lemonade stands	Lemonade stands are short-term pop-ups, easy to create and then easy to take down. Though they represent a more significant investment than cognitive walkthroughs, they allow the user to come to you, rather than you going to them.
Smoke tests	Smoke tests run competing versions of ads with different value propositions (how we proposed to deliver value) linked to landing pages aimed at "faking a new business fast." They ask prospective customers to act in some way—click a link, send an email, or sign up to be part of a beta test.
Simulations	Simulations can be either virtual experiences that prospective users click through or physical experiences that use a Wizard of Oz "man behind the curtain" delivery approach. Both focus on providing an authentic front end for users to experience but lack functioning back ends.
Trials	Trials represent the most elaborate type of test design and use fully functional prototypes, with authentic front and back ends. They often rely on accompanying methods like control groups and pretest/post-test designs to ensure more rigorous measurement of outcomes.

[minimum viable product] for your very big idea you need to first figure out all the different independent assumptions that you're trying to understand. And then for each of those, you impose constraints that force you to think about how to get to the root of the thing," they explained. "As you start getting data, it makes it easier to imagine then how to reconstruct that complexity, without building something that you don't even understand."

Another MyHealthEd team member emphasized the value that focusing on a single component could contribute versus when to test the whole. "If you're at the stage where you really need to know the nuts and bolts, it makes sense to test one component in greater detail at greater fidelity. "Because then you can really go deep," they said. "If you are more interested in how different components relate to each other and higher level types of questions then testing a full solution might be good."

Another crucial question addressed was the nature of the data to be gathered—specifically, whether it was derived from respondents *saying* ("say" data) something versus *doing* ("do" data) something. "Say" data came from tests that *asked* prospective users what they would do or what they liked or needed; "do" data was based on their *actual behaviors*. Ultimately, experienced experimenters agreed that "do" data were required to establish proof. "We had four promising concepts that we were really excited about and needed to make choices among these four," explained one PMI team member. "With a really important board meeting coming up, we had a lot of 'say' data, which was inspiring the path forward, but didn't give us the confidence we needed to actually invest . . . We needed 'do' data." They also explained that "do" data could be expensive to get, take additional time, and require a more complete prototype. Early in the testing of a new idea, PMI experts preferred "say" data for its speed and low resource use.

Finally, the centrality of triangulation of both approach and data sources emerged as a strong theme across cases. A Nike team leader commented, "You are looking for a 'good enough' level of confidence and directional evidence of whether to stay straight or go left, not right. It's a process of triangulation. You're figuring out where the walls are and what the boundaries are and then trying to head towards center." A PMI team member agreed. "We were sitting in a sea of data because we had run so many tests . . . we had some frames that really synthesized all the data. And, if you squint, you can see across all the tests, and that was really edifying because it kept being the same concepts that were edging out," they said. "When you run eight tests and Snippets and Career Navigator are the top answer in seven of the eight, that's going to give you confidence that these are the two concepts that should move forward. And so, we talk a lot about triangulation of data."

Experienced experimenters planned a careful progression through increasingly challenging levels of test rigor, and resisted the urge to skip ahead to more elaborate tests. One Nike team member elaborated on the reasons why: "There's often a rush to skip steps and say, 'Hey, if I could run a trial here, shouldn't I just start there instead of with simple cognitive walkthroughs? Doesn't that feel faster? And isn't the learning going to be more valuable?' But you don't yet know how to choose if you haven't done the steps (like cognitive walkthroughs) before." They added, "Though it seems a bit counterintuitive, you *need* these steps. They prevent you from making an expensive mistake by rushing to conclusions or having to run a dozen different possible permutations of a trial in the market, which is unaffordable."

Once the experimenters selected a test type, attention turned to the activities involved in creating a visualization of the concept to act as stimulus in the experiment. Innovators described their struggles matching test design with the associated levels of fidelity needed for form, function, and level of interactivity in the needed prototype. For example, when the MyHealthEd team tested the *ExploreMe* and *xHealth* concepts, they first built low-fidelity prototypes, created in PowerPoint, to use in cognitive walkthroughs. As the *xHealth* concept went through additional rounds of testing, it evolved into a new concept called *Real Talk*; the team constructed higher-fidelity prototypes of this concept for use in simulations and trials.

We observed all of the teams striving to keep prototypes as simple as possible, while still allowing them to function as the test type required. One SWR team member explained, "We realized that if you're going to operate a concierge service, you need a stage for your staff. You can't just shove people out onto the ticket office floor and expect things to work. Of course, you also need to have the skills, along with the environment to work in. So the minimum viable concept was a very simple pop-up—because it's cheap. We could do something very simple." A Nike team member concurred: "It's going to be an imperfect experience. That said, we want them to experience it as a real user. They're not thinking it's a perfect solution, but they assume it's fully functional. So you need to build in that functionality. The question is *how?* . . . Maybe eventually you do have to build the technology to go with it, but we prefer to *fake* it for a while until we have to make it."

What was crucial throughout was that whatever prototype was created, it accurately embodied key assumptions in a realistic format. A Nike team member offered an example: a finance department employee argued for simply giving the shoes away, because recording such a small amount of revenue was not worth the headache. "But we couldn't let users in the trial *not* pay for it," they argued. "From a learning standpoint, we really needed to know sooner rather than later if they were willing to pay for it. Who's *not* going to love a free set of shoes every month from Nike? So we couldn't live without charging for it."

What Are We Learning?

To inform future rounds of testing, innovators stressed the importance of synthesizing and documenting the test results and the lessons learned. "In order to actually learn and make

> **A crucial question addressed was the nature of the data to be gathered—specifically, whether it was "say" data derived from respondents *saying* something or "do" data from *doing* something.**

To inform future rounds of testing, innovators stressed the importance of synthesizing and documenting the test results and the lessons learned.

sure that knowledge is sustained, we have to keep accurate records. And that comes in the form of various kinds of documentation," a MyHealthEd team member explained: "I think one challenge I see a lot of teams have is that they have all of the results in their heads. And then those people leave and a new employee repeats the same mistakes, and reinvents the same solutions." They added, "We want good documentation so that if somebody new is trying to come in and learn from previous work, they know exactly why certain decisions were made, and they know exactly what assumptions were already tested. Then, we are able to keep innovations moving forward and building on each other."

Innovators indicated that the credibility of their findings rested on their ability to demonstrate the soundness of their process. In particular, when findings ran counter to what key stakeholders expected, they anticipated attempts to discredit the testing design and execution, rather than accept the findings. A MyHealthEd team member explained, "I think that one of the biggest benefits that we had with *Real Talk* was that pretty much every element of the app, from the name to the branding to the features, were grounded. We could point to where it came from, and why we made that decision in a very robust and rigorous way."

During this stage, concepts were moved forward, iterated, or eliminated. At PMI, for instance, the first-round testing resulted in the elimination of two concepts: *Hive* because it failed to meet threshold levels of desirability, and *Spot*, which suffered from a significant discrepancy between the benefits that users wanted relative to the time they were willing to commit. PMI abandoned these two concepts and only *Snippets* and *Career Navigator* moved on to the next round of testing. The PMI team captured this flow as part of its storytelling to the larger organization (Figure 3).

The Importance of Psychological Safety and Failure

During our research we noticed the presence of psychological safety. Specifically, engaging influential stakeholders had the effect of creating a safe space for expressing and participating in the experimentation process. "We show stakeholders that we are constantly trying to improve and take into account what our frontline employees are saying and what the customer is saying. That's a testament to the success we've had embedding this type of thinking at SWR," a team member explained. "We wanted to change the way the whole organization thinks. That doesn't happen by accident—it's the creation and curation of the events and the process that create a safe space. People know they are safe participating in our process. Participating is a brave thing to do, so we needed to create that safe space."

Given the high rate of uncertainty, the risk and importance of failure, and the high stakes in the front end of innovation, cultivating an environment in which all participants feel safe to express their ideas and opinions in early-stage experimentation is critical. When companies cultivate a psychologically safe environment, innovators do not fear reprisals from failure. A MyHealth team member explained why acknowledging and accepting failure was so critical to a successful project: "We're never going to get it right the first time around . . . we're always going to have some bumps. We're really trying to create a supportive environment, trying to create this environment where we're accepting and receptive of feedback—even when it's critical, even when it means that we're off the mark." They added, "Creating an environment that's really open and receptive to this will then allow those issues to be brought up. And then your team can take that in, in a productive way, to advance your project."

A Process Model for Early-stage Experimentation

All four case studies stress the importance of sequencing. Moving into test design selection (*how it was being tested* from the test) without first surfacing critical assumptions and specifying evidence (*what* was being tested), for instance, risked wasting significant resources on tests that failed to evaluate the most critical aspects of a concept or that were more elaborate than needed. Similarly, building prototypes before surfacing assumptions raised the risk of creating unnecessarily elaborate prototypes—or ones that did not allow testing of the most critical assumptions.

Based on our synthesis of the information gathered during our analysis, we constructed a process model for early-stage experimentation. The process we developed includes five steps that clearly define the aim, approach, and lessons learned in experimenting in early-stage innovation. Steps 1 and 2 focus on identifying a testable hypothesis and the appropriate metrics to test it. Steps 3 and 4 focus on selecting a test design and accompanying prototype. Step 5 describes the execution, documentation, and iteration of the test. Under each step, we delineate the best practices identified in our study associated with that step (Figure 4). To offer further guidance, we have developed a series of templates for each step (Figure 5 and Figure 6). Our process emphasizes the importance of purposeful experimentation.

This process supports the existing academic literature that emphasizes the multifaceted nature of experimentation. It highlights that experimentation involves not only the formulation and testing of hypotheses (von Hippel and Kaulartz 2021) but also the specific and accurate execution of the crafting and testing processes. Additionally, incorporating experimentation in the initial phases of innovation reveals the significance of our process model during the early stages of innovation, when companies have not yet chosen ideas (Sukhov et al. 2021). Our process emphasizes the importance

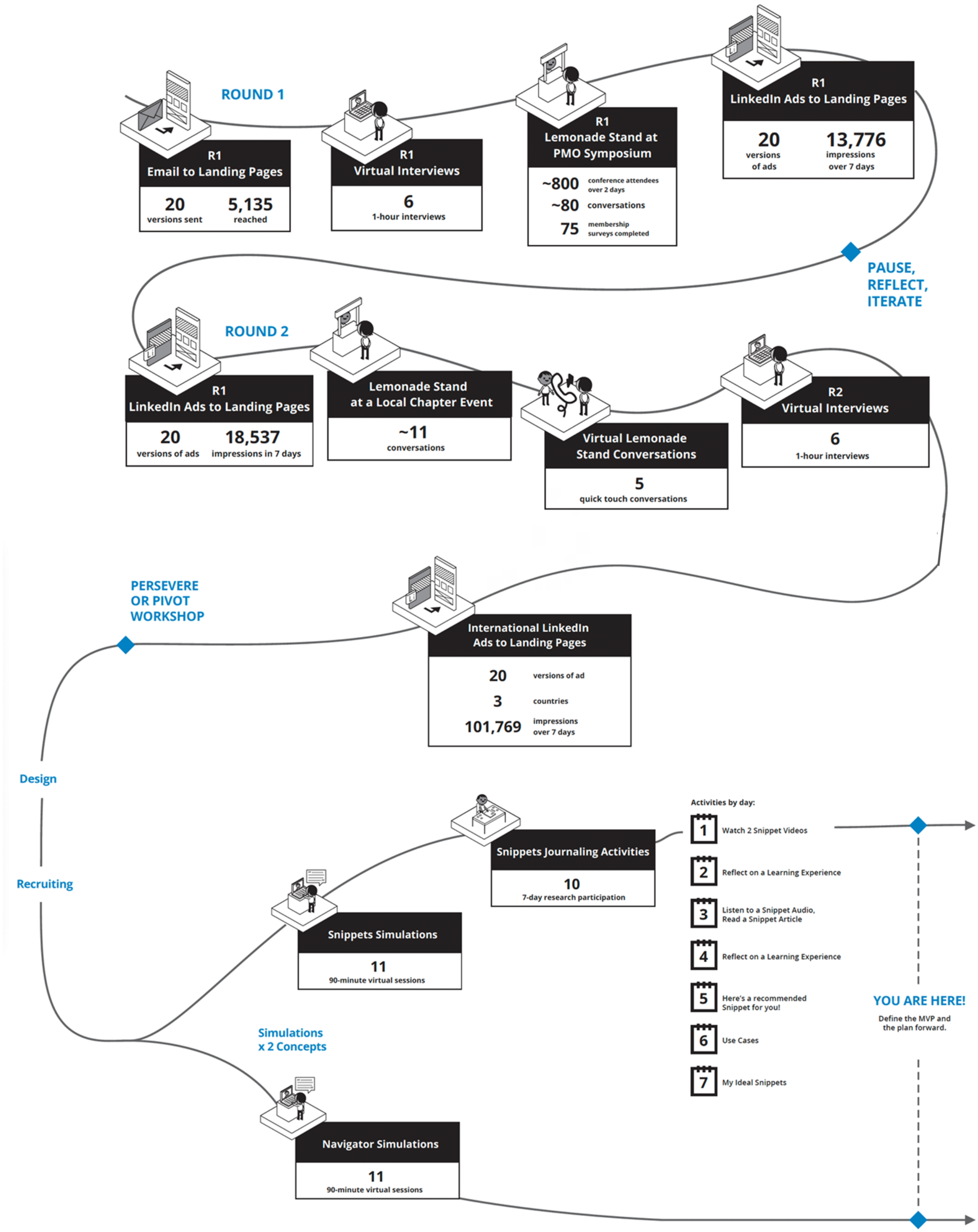

FIGURE 3. The flow of testing of PMI's four concepts

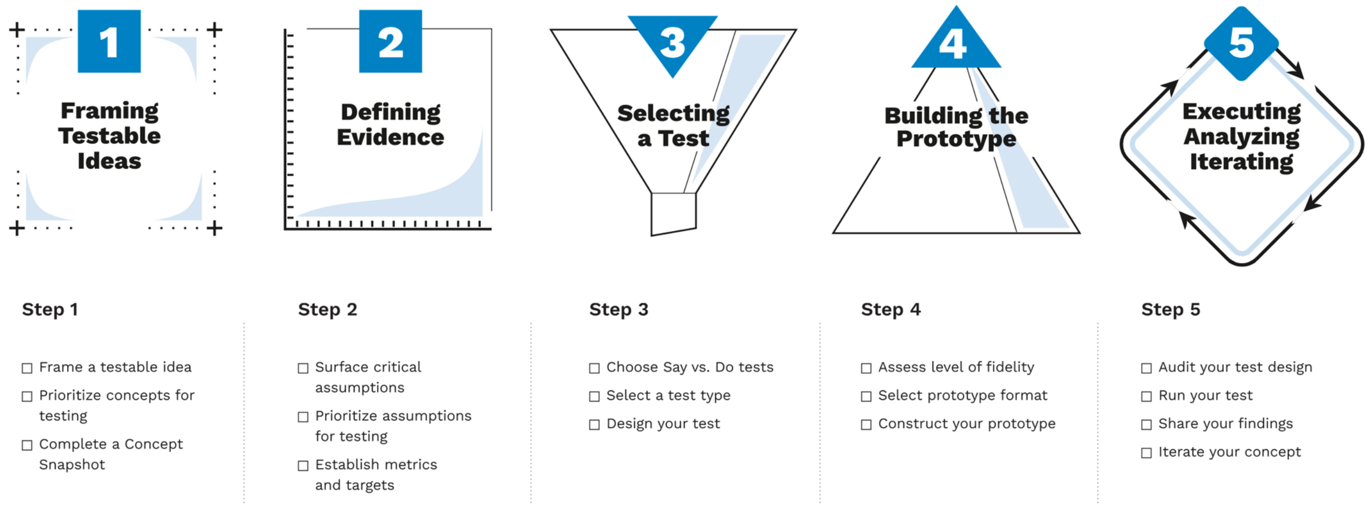

FIGURE 4. Process model overview

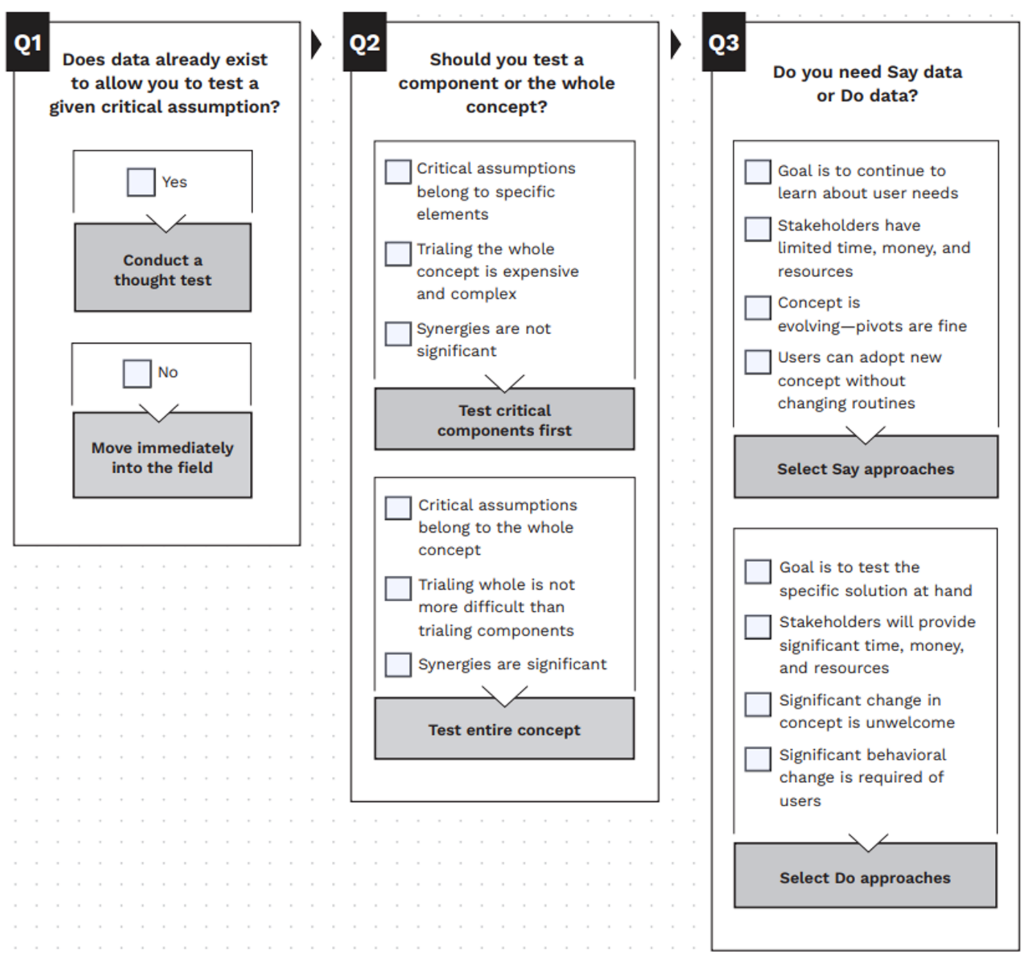

FIGURE 5. Example template of Step 4: Test design decision flow (part 1)

of experimentation from the start, rather than just in later stages as emphasized in design management approaches (Magistretti, Ardito, and Messeni Petruzzelli 2021) to test prototypes and MVPs (Andries et al. 2013). Given the higher uncertainty in this phase (Furr 2022), selecting the priorities and being quick in testing assumptions with a clear objective in mind is even more relevant. Our investigation revealed that experimentation involves more than a mindset—it requires a process and the use of tools, as highlighted in the entrepreneurial literature, which emphasizes the mantra of

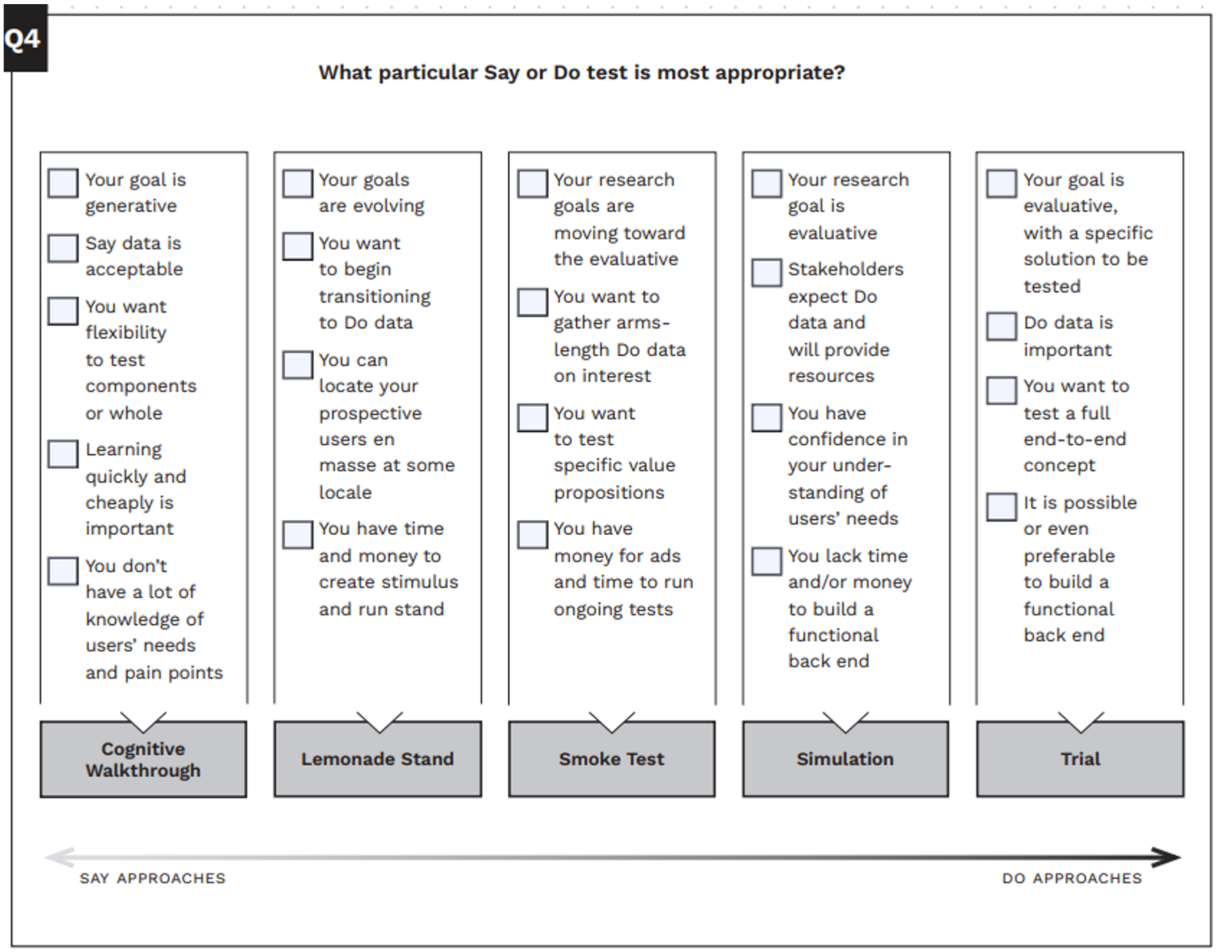

FIGURE 6. Example template of Step 4: Test design decision flow (part 2)

"fail fast to succeed soon" (Hampel, Perkmann, and Phillips 2020). In order to experiment effectively, innovators must have the right tools and tasks to navigate the uncertainties of the early stages of the innovation process.

Managerial Implications

In this article, we present managers and their organizations with guidance in the form of a replicable, teachable, and scalable model of an early-stage experimentation process. We intend our process and associated practices to provide managers with additional structure and rigor for moving their experimentation efforts forward. Our research suggests that this early-stage experimentation process can help practitioners in five keys ways:

1. *For learning:* Experimentation, comprising inexpensive learning-oriented tests that are executed quickly, and which use low-fidelity prototypes, is a valuable way to build capacity and to better manage innovation risks.
2. *As a catalyst:* Experimentation can be codified into a simplified and scalable process that can be practiced across organizational levels and functions. Developing organization-wide capabilities in experimentation can enhance risk management in innovation endeavors.
3. *For inquiry:* By following the steps identified in the process model we have proposed, practitioners should be able to execute the *why*, *what*, and *how* of early-stage experimentation successfully and avoid common pitfalls in the experimentation process.
4. *As a safety net:* The specific practices presented here provide the structure to enhance psychological safety and motivate managers to be more likely to engage in experimentation.
5. *As a dynamic capability:* Leaders looking to reap the benefits of experimentation can implement the process and associated tools in a standardized, scalable way that turns experimentation into a systematic and sustainable practice. In doing so, they can potentially enhance a firm's larger strategic capability to adapt to change and exploit opportunities in an uncertain world.

Conclusion

Through our observations of experimenters in four cases, we identified practices that we used to create a process to enhance early-stage experimentation in innovation. In doing so, we offer practitioners an effective way to accelerate innovation in a cost-effective manner. We contribute to current debate on how to practically enact experimentation in organizations (Thomke 2020; Hampel, Perkmann, and Phillips 2020). We contend that the process and practices identified in our study can make early-stage experimentation tangible, easy to implement, and scalable, and turn it into an embedded practice in organizations.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

References

Andries, P., Debackere, C., and Van Looy, B. 2013. Simultaneous experimentation as a learning strategy: Business model development under uncertainty. *Strategic Entrepreneurship Journal* 7(4): 288–310. doi: 10.1002/sej.1170

Baxter, D., Trott, P., and Ellwood, P. 2023. Reconceptualizing innovation failure. *Research Policy* 52(7): 104811. doi: 10.1016/j.respol.2023.104811

Cannon, M. D., and Edmondson, A. C. 2001. Confronting failure: Antecedents and consequences of shared beliefs about failure in organizational work groups. *Journal of Organizational Behavior* 22(2): 161–177. doi: 10.1002/job.85

Cooper, R. G., and Sommer, A. F. 2018. Agile–Stage-Gate for manufacturers. *Research-Technology Management* 61(2): 17–26. doi: 10.1080/08956308.2018.1421380

Dweck, C. S. 2008. *Mindset: The New Psychology of Success*. New York: Random House.

Durante, I., Dell'Era, C., Magistretti, S., and Pham, C. T. A. 2024. Predictive or imaginative futures? Experimenting with alternative future-making approaches. *Creativity and Innovation Management*, February, 2024. doi: 10.1111/caim.12603

Edmondson, A. C. 1999. Psychological safety and learning behavior in work teams. *Administrative Science Quarterly* 44(2): 350–383. doi: 10.2307/2666999

Eisenhardt, K. M. 2021. What is the Eisenhardt method, really? *Strategic Organization* 19(1): 147–160. doi: 10.1177/1476127020982866

Euchner, J., and Ganguly, A. 2014. Business model innovation in practice. *Research-Technology Management* 57(6): 33–39. doi: 10.5437/08956308X5706013

Furr, N. 2022. *The Upside of Uncertainty: A Guide to Finding Possibility in the Unknown*. Cambridge: Harvard Business Press.

Ganguly, A., and Euchner, J. 2018. Conducting business experiments. *Research-Technology Management* 61(2): 27–36. doi: 10.1080/08956308.2018.1421381

Goffin, K., Åhlström, P., Bianchi, M., and Richtnér, A. 2019. Perspective: State-of-the-art: The quality of case study research in innovation management. *Journal of Product Innovation Management* 36(5): 586–615. doi: 10.1111/jpim.12492

Hampel, C., Perkmann, M., and Phillips, N. 2020. Beyond the lean start-up: experimentation in corporate entrepreneurship and innovation. *Innovation* 22(1): 1–11. doi: 10.1080/14479338.2019.1632713

Higgins, E. T. 1998. Promotion and prevention: Regulatory focus as a motivational principle. *Advances in Experimental Social Psychology* 30:1–46. doi: 10.1016/S0065-2601(08)60381-0

Liedtka, J., Hold, K., and Eldridge, A. 2021. *Experiencing Design: The Innovator's Journey*. New York: Columbia Business Press.

Magistretti, S., and Trabucchi, D. 2024. Agile-as-a-tool and agile-as-a-culture: a comprehensive review of agile approaches adopting contingency and configuration theories. *Review of Managerial Science* 1–31. doi: 10.1007/s11846-024-00745-1

Magistretti, S., Ardito, A., and Messeni Petruzzelli, A. 2021. Framing the microfoundations of design thinking as a dynamic capability for innovation: Reconciling theory and practice. *Journal of Product Innovation Management* 38(6): 645–667. doi: 10.1111/jpim.12586

Magistretti, S., Sanasi, S., Dell'Era, C., and Ghezzi, A. 2023. Entrepreneurship as design: A design process for the emergence and development of entrepreneurial opportunities. *Creativity and Innovation Management* 32(1): 5–21. doi: 10.1111/caim.12529

Micheli, P., Wilner, S., Bhatti, S. H., Mura, M., and Beverland, M. 2019. Doing design thinking: Conceptual review, synthesis, and research agenda. *Journal of Product Innovation Management* 36(2): 124–148. doi: 10.1111/jpim.12466

Narduzzo, A., and Forrer, V. 2024. Nurturing innovation through intelligent failure: The art of failing on purpose. *Technovation* 131:102951. doi: 10.1016/j.technovation.2024.102951

Patel, P. C., Kohtamäki, M., Parida, V., and Wincent, J. 2015. Entrepreneurial orientation-as-experimentation and firm performance: The enabling role of absorptive capacity. *Strategic Management Journal* 36(11): 1739–1749. doi: 10.1002/smj.2310

Sanasi, S., Manotti, J., and Ghezzi, A. 2021. Achieving agility in high-reputation firms: Agile experimentation revisited. *IEEE Transactions on Engineering Management* 69(6): 3529–3545. doi: 10.1109/TEM.2021.3128865

Siggelkow, N. 2007. Persuasion with case studies. *Academy of Management Journal* 50(1): 20–24. doi: 10.5465/amj.2007.24160882

Sukhov, A., Sihvonen, A., Netz, J., Magnusson, P., and Olsson, L. E. 2021. How experts screen ideas: The complex interplay of intuition, analysis and sensemaking. *Journal of Product Innovation Management* 38(2): 248–270. doi: 10.1111/jpim.12559

Thomke, Stefan H. 2020. *Experimentation Works: The Surprising Power of Business Experiments*. Cambridge, MA: Harvard Business Press.

Tversky, A., and Kahneman, D. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5:297–323. doi: 10.1007/BF00122574

von Hippel, E., and Kaulartz, S. 2021. Next-generation consumer innovation search: Identifying early-stage need-solution pairs on the web. *Research Policy* 50(8): 104056. 10.1016/j.respol.2020.104056

Zaharee, M., Abad, L. B., Chandra, P., Krautkramer, C., Mehlman, S., Schall, J., and Trauth Taylor, K. 2021. How companies can benefit from brilliant failures. *Research-Technology Management* 64(2): 31–38. doi: 10.1080/08956308.2021.1865000