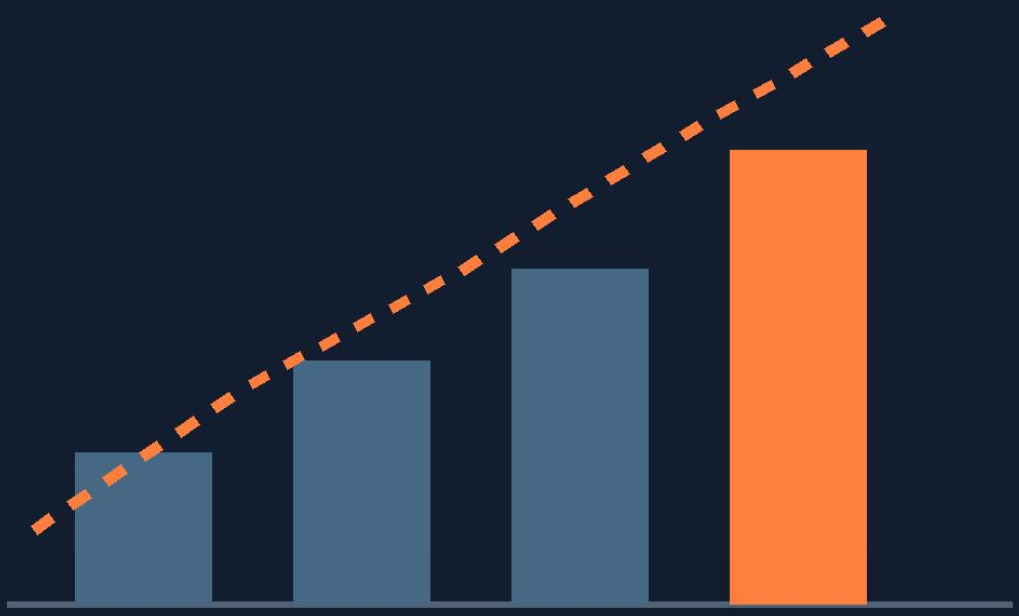
SILICON FOUNDERS

SF / 02

The Startup Validation Playbook

Test Demand Before You Scale



SF PUBLICATIONS

SILICON FOUNDERS / SF PUBLICATIONS

The Startup Validation Playbook

Test Demand Before You Scale

SF PUBLICATIONS / VOLUME 02

Publisher: Silicon Founders

Edition: 1.0 | 30 October 2026

Audience: Founders, product leaders, venture builders, innovation teams, and market researchers

> **Reader note.** This guide provides educational methods for learning about markets and customers. It does not guarantee product-market fit, revenue or investment. All examples are fictional and illustrative. Follow relevant privacy, consent, advertising and consumer-protection requirements in the places where your experiments operate.

About this guide

Startup validation is not a ceremony conducted after a product has been built. It is a repeated effort to reduce the uncertainty surrounding a specific business decision. The most valuable experiment is rarely the most impressive-looking one. It is the smallest ethical test that can distinguish between materially different choices: pursue this segment or another, build the integration or wait, use a subscription or an implementation fee, hire a sales lead or continue founder-led discovery.

This playbook follows the journey from assumptions to trustworthy observations. It emphasizes behavior over compliments, discriminating experiments over polished prototypes, and decision rules over cherry-picked results. It also recognizes constraints: some products have long purchasing cycles, regulated processes, or small accessible markets. A startup should adapt the depth and duration of testing to its real risks rather than blindly copying an internet growth formula.

By the end, the reader should have a customer-segment hypothesis, an interview record, a sequenced evidence plan, at least one demand experiment with an explicit threshold, and a documented decision to continue, change or stop. The worksheets can be reused every time the business enters a new geography, price point or customer category.

Contents

1. Validation Is a Sequence of Decisions
2. Find a Segment with a Costly Job to Be Done
3. Interview for Episodes, Not Compliments
4. Design Experiments That Can Fail Honestly
5. Price, Commitment and Purchasing Reality
6. Activation, Retention and the Signal Problem
7. Interpret the Evidence Without Fooling Yourself
8. The Six-Week Validation Program

Appendix A. Assumption Ledger

Appendix B. Interview Script and Observation Sheet

Appendix C. Experiment Card and Results Log

Appendix D. Sample Validation Case Files

Appendix E. Definitions and Sources

CHAPTER 01

1. Validation Is a Sequence of Decisions

Begin with the decision at risk

Start by asking what the team would do differently if a particular assumption were false. If the answer is 'nothing,' the assumption is not yet worth testing. Teams waste money gathering information that does not change their plans: collecting survey responses that merely echo their product slogan, commissioning broad reports when they need one buyer interview, or optimizing a landing page before confirming the buyer has a budget.

Define validation as an evidence-producing process with four stages: articulate the claim; design a discriminating test; record the actual result; choose the next action. Every stage should be written before the result is known. For example, 'Small clinics will replace a shared spreadsheet with a paid scheduling tool if the tool reduces daily reconciliation time by at least 30 minutes' is a useful hypothesis because it identifies a segment, current alternative, proposed value and measurable behavior. 'Businesses want automation' is not.

An assumption ledger makes the process visible. For each major belief, record its potential impact if wrong, your current evidence, the cost to test, the test owner and the decision date. Prioritize assumptions with the combination of high importance and low evidence. In a marketplace, a critical assumption may concern whether both sides participate at the required time. In enterprise software, it may concern procurement. In a clinical technology venture, lawful use and appropriate safety validation may be threshold conditions rather than ordinary growth experiments.

Do not confuse kinds of evidence

A founder's own experience is a useful hypothesis source, but it is not independent demand data. A customer interview describes reported past behavior. An observed workflow shows current behavior. A signed pilot records a commitment subject to its terms. A collected payment is stronger evidence of payment willingness than a non-binding 'interested' email, but even a paid pilot does not establish sustainable retention or attractive margins. No single evidence type answers every question.

Record evidence along a ladder: desk research; interviews; observed actions; live tests; financial commitments; repeated paid use; reliable retention or expansion. The ladder is not a claim that one rung always invalidates another. Regulatory approval, technical feasibility and procurement policy may require documentary evidence that is more decisive than a fast consumer experiment. Match evidence to the decision.

Fictional case. A team named FlowDesk assumes 'small service firms urgently need an AI proposal writer.' It begins building. Interviews reveal that proposals are not the bottleneck; follow-up approval across departments is. The team revises the job description to 'help service firms move proposals from draft to authorized client delivery.' This change affects the product, buyer, workflow and pricing, and therefore counts as meaningful learning.

Make the learning unit small

One cycle should test one primary assumption, though secondary observations can be recorded. Use a short hypothesis statement, defined target segment, explicit outcome measure and stop rule.

If three major variables change at once, a positive result may be impossible to interpret. That is especially common when teams change price, product and sales channel simultaneously.

Chapter workbook

Deliverable: An assumption ledger ranked by 'damage if false' and 'current evidence.' **Decision rule:** Only greenlight an expensive build when the critical commercial and lawful-operating assumptions have a plausible evidence base. **Team question:** Which assumption will we regret not testing before hiring or integrating more tools?

CHAPTER 02

2. Find a Segment with a Costly Job to Be Done

Narrowing the market accelerates learning

A segment is not simply a demographic category. It is a group of people or organizations with sufficiently similar constraints, alternatives and purchasing behavior that the same offer can be tested with them. 'SMEs' is too broad. 'Independent bookkeeping firms serving 20–100 business clients and preparing monthly reconciliations manually' is more actionable. The more specific the segment, the easier it becomes to compare interview answers and observe repeated patterns.

Use a three-part customer profile: the job the customer is trying to accomplish, the obstacles that make the job costly, and the outcome they value. Strategyzer's Value Proposition Canvas provides a useful reference for jobs, pains and gains, but our purpose here is to create testable claims rather than fill every box for appearance. [R1] Distinguish customer desires from budgeted priorities. A problem can be annoying without being important enough to purchase a solution.

Map the entire purchase situation

If the product is bought by a company, separate the daily user, process owner, financial sponsor, information security reviewer and procurement decision-maker. In some markets one person holds several roles; in others five committees can delay a purchase. Identify where the decision can stall. Talk to the buyer about priorities, not just features. Ask about the last purchase in the category: trigger, approvals, contract length, procurement cycle and reasons deals fail.

A good segment score considers problem severity, frequency, ability to reach interviewees, willingness and ability to pay, competition, technical feasibility, and time to deploy. Scores are prioritization aids. One legal prohibition or inaccessible data dependency can override a high average. Write critical exclusions explicitly: customer types the product cannot currently serve, geographies where it is not authorized to operate, or use cases that require certifications it lacks.

Separate market evidence from market size

A large market does not imply that this startup can enter it. Market sizing belongs in the strategic analysis, but validation starts with the reachable subset. Build a named list of potential accounts from verifiable sources and select interviewees deliberately. Include buyers who recently chose a competitor or refused to solve the problem. The team learns more from explaining disqualification than from counting contacts who merely accepted a meeting.

Fictional case. FlowDesk tests agencies, accountants and property management firms simultaneously. The results cannot be compared because the approval workflow differs dramatically. The team selects agencies that issue 10–30 complex proposals monthly and have a two-person approval chain. Interview quality increases because every conversation now explores the same workflow.

Chapter workbook

Deliverable: One-page segment brief: entry criteria, exclusions, user, buyer, key episode, alternative, hypothesized value, budget event and likely channel. **Decision rule:** Do not blend data from structurally different segments into one conversion metric. **Exercise:** Describe five reasons an apparently ideal customer might still refuse the product.

CHAPTER 03

3. Interview for Episodes, Not Compliments

Build a discovery conversation around memory

Interviewing is the least expensive experiment founders commonly misuse. The main failure is leading the customer: 'Wouldn't it help if our platform automated this?' The question reveals the product concept and invites politeness. Instead ask the participant to reconstruct the most recent episode. A sequence that often works is: context, trigger, current process, workaround, cost, decision roles, alternatives considered, outcome, and next expected occurrence. When relevant, ask to observe or document a process with proper permission.

Invite silence. A hurried founder tends to insert their own interpretation into every pause. Let customers use their own vocabulary; those words may later reveal how they search for solutions. Ask for specific artifacts when appropriate, but never pressure participants to disclose confidential contracts, personal records or privileged internal information. Collect the minimum data needed to answer the research question and follow privacy requirements.

The U.S. National Science Foundation's I-Corps initiative is one example of structured customer discovery applied to commercialization. Its approach emphasizes direct conversations with potential customers and stakeholders rather than relying only on technical enthusiasm. [R2] A startup does not need to mimic a particular interview quota; it should seek enough representative evidence to make the specific decision with appropriate caution.

Code findings before they become anecdotes

After each interview, write observations separately from interpretations. 'The manager re-entered 17 fields from one system into another' is an observation. 'They would pay $500 monthly for automation' is a hypothesis unless an actual commercial response supports it. Quote sparingly and only with permission for public use. In internal notes, tag repeated issues: frequency, severity, workaround, change trigger, who pays, integration obstacle, and reasons not to buy.

Look for disconfirming patterns. If users report different problems, the segment may be too broad. If most buyers cannot remember a recent costly episode, the pain may be less urgent than the founder assumed. If the user loves the concept but the procurement team cannot purchase the vendor category, the go-to-market plan needs to change. The objective is to discover an investable problem if one exists, not to defend the original proposition.

Translate findings into specific next tests

Do not publish a 'validated idea' announcement after an interview series. Synthesize: what problem is strongly supported, what remains contested, what purchase barriers emerged, and which experiment would clarify behavior? A short learning memorandum should lead naturally to a test offer or prototype. Record the number and composition of interviewees, the recruitment source, and any reasons the sample may be biased.

Fictional case. Eight agency managers praise FlowDesk's draft suggestions, but five explain that their real frustration is chasing approvals. Two will share anonymized approval timelines. One agrees to a paid diagnostic engagement. The evidence supports a process-focused pilot, not a claim that the proposed AI product has achieved market fit.

Chapter workbook

Deliverable: An interview guide and coded evidence sheet with date, segment and observed episode. **Decision rule:** Do not count statements of future intent as paid conversions. **Exercise:** Rewrite five leading questions into neutral questions about past behavior; then have someone who dislikes your idea review the script.

CHAPTER 04

4. Design Experiments That Can Fail Honestly

What makes a test informative?

A useful experiment begins with a falsifiable decision hypothesis. 'If we show 20 qualified operations managers a fixed-scope offer with a stated price, at least four will accept a real next step within fourteen days' is informative even if the result is zero. It has a target population, a defined offer, a behavioral metric and a precommitted criterion. The threshold is a managerial assumption, not a statistical guarantee; it should reflect the cost of the next decision.

Choose the least expensive test that has enough realism to be meaningful. A concept interview checks language and problem understanding. A clickable prototype checks task comprehension. A concierge pilot checks whether customers value a manually delivered outcome. A paid pilot tests a narrower commercial behavior. A live product test checks whether the operational system can actually deliver. These are different questions; do not substitute a decorative landing-page click for evidence of successful implementation.

Experiments should be ethical and truthful. Do not pretend a feature exists when it does not. If a pre-launch page accepts interest but not service delivery, say that plainly. If a manual team performs the service, do not describe it as fully automated. Follow relevant advertising standards: U.S. FTC small-business guidance is one source for truthfulness and substantiation principles, though local rules may differ. [R3]

Use a baseline, a decision threshold and a cost cap

Before launch, record the current workflow baseline. Identify what must improve: hours saved, errors avoided, qualified meetings, deposit payments, activation or successful task completion. Define numerator, denominator, exclusions and observation window. Set a time and spending cap. Some tests require sample sizes beyond a startup's means, especially for modest conversion improvements; do not call a small directional result a statistically conclusive finding.

Control confounders where practical. Changing the headline, price and traffic source at once makes attribution difficult. For B2B outreach, comparable accounts and a consistent offer may be more useful than fashionable A/B testing language. Keep a rejection log because rejections often identify the true barrier: timing, authority, existing tools, safety requirements or weak differentiation.

Know when manual delivery is valuable

A concierge experiment can reveal hidden operational work before a team invests in software. A founder can manually perform a service for a few willing clients, recording time, quality and cost. But manual success does not prove product scalability. Map every step that would need automation, delegation or standardized process to preserve margin. The best output may be a redesigned process rather than a software feature.

Fictional case. FlowDesk runs a two-week approval-coordination pilot with three agencies. The founders manually route approvals and measure each stage. All three finish faster, but one requires four hours of founder work per proposal. The test reveals both demand and a delivery-cost problem. The next decision is to automate only the repeated bottleneck, not launch a broad platform.

Chapter workbook

Deliverable: One experiment card, signed-off before launch, and a results log updated after each observation. **Decision rule:** Pre-register the threshold and preserve negative outcomes. **Exercise:** Write a headline that states what the product currently does, without using future tense or unsupported AI claims.

CHAPTER 05

5. Price, Commitment and Purchasing Reality

Pricing is a negotiation with constraints

A quoted price is meaningful only in relation to an offer, scope, buyer, duration and risk. A $2,000 monthly subscription may be inexpensive for a company with an approved budget and expensive for an individual user whose department cannot authorize procurement. Separate willingness to pay from ability to pay, and distinguish contractual intent from operational capacity to deploy.

Start with the value mechanism but do not assume that capturing a fixed percentage of estimated customer savings is always appropriate. A customer's baseline may be uncertain, savings may accrue in a different department, and benefits may be political or risk-related rather than directly monetary. Test the commercial package: what is included, what the customer must supply, how outcomes are measured, how long a trial lasts, what happens on renewal, and what the company will not deliver.

A customer can demonstrate commitment through payment, a budget-approved pilot, an implementation date with assigned staff, or an executed agreement; these signals have different strengths. A non-binding letter of interest can be useful, but should not be counted as revenue or a guaranteed deal. In regulated settings, contract validity and procurement authority may require additional verification.

Measure sales process friction

Keep a deal diary for each qualified opportunity. Record discovery date, buyer roles, procurement requirements, objections, pricing discussions, next step and actual outcome. A slow process can be acceptable if contract economics support it. What matters is whether the business can finance the cycle and whether implementation yields a repeatable margin. The founder should learn which commitments genuinely predict closing and which merely prolong conversation.

Test different pricing mechanisms with caution. Per-seat pricing can penalize broad deployment. Usage pricing can align payment with value but make customer spend unpredictable. Fixed fees simplify budgeting but may subsidize heavy users. Implementation fees compensate onboarding work but may slow adoption. Choose a structure the buyer can understand and the company can meter accurately.

Chapter workbook

Deliverable: A one-page offer sheet containing problem, delivered outcome, price, scope, buyer obligations, timeline, measurable acceptance criteria and exclusions. **Decision rule:** Do not infer repeatable demand from discounted one-off pilots without a path to normal pricing. **Exercise:** Compare how a buyer would approve a monthly subscription, annual contract and fixed pilot fee.

CHAPTER 06

6. Activation, Retention and the Signal Problem

The first value moment must be defined

A user may sign up, log in and click through a tutorial without receiving the product's promised value. Activation should reflect a meaningful first outcome. For FlowDesk it might be 'one proposal is routed through all required approvals and delivered with an auditable history.' For a data product it might be 'a decision-maker uses a verified dashboard to complete a real reporting task.' Record the first-value event, time to value, number of failed attempts and reasons customers stall.

Define a small customer journey. A B2B workflow product may have the states: qualified, contracted, configured, first value, repeat use, renewal and expansion. Track numbers by cohort, not only totals. If ten customers were activated in one month, ask how many still complete the core workflow in month two or three. Retention should be evaluated against the natural frequency of the job. Daily use is not required for a product whose important job happens quarterly.

Understand survivorship bias. Positive interviews with engaged users can ignore customers who never activated. Build a process to study inactivity and churn. Ask what changed in the customer's environment, whether the promised value was obtained, and whether servicing cost undermined the relationship. Track cancellations and refunds honestly; do not retain them as active users in a management dashboard.

Build a minimal metric system

Use only measures tied to decisions: qualified-to-paid conversion, activation rate, time to first value, repeat completion, renewal, contribution margin and customer support load may be sufficient at first. Each metric needs a definition, owner and source. Write the formula next to the number. Stripe provides educational explanations of recurring revenue and retention metrics; apply consistent definitions before using external comparisons. [R4]

A product cannot be declared validated solely on acquisition efficiency when the customer outcome is poor. Paid advertising can create many trial accounts, yet long-term retention may reveal a weak value proposition. Conversely, a small but highly engaged cohort may justify further testing even if initial acquisition is manual. Interpret metrics with sample sizes and commercial context visible.

Fictional case. FlowDesk acquires 25 trial accounts; 11 complete setup and four finish an approval workflow. The team initially celebrates 25 trials, then measures only four value activations. The next experiment removes two configuration steps and asks whether more accounts complete the first value event. This is a focused product decision, not a sweeping market-fit claim.

Chapter workbook

Deliverable: A lifecycle definition with stage event, measurement window, source system and next action when a customer stalls. **Decision rule:** Do not confuse account creation with value delivered. **Exercise:** Interview three users who abandoned onboarding and list what evidence would falsify your favored explanation.

CHAPTER 07

7. Interpret the Evidence Without Fooling Yourself

Evidence is not an argument to win

Founders naturally prefer observations that confirm their thesis. The job of a validation review is to protect the venture from that bias. Present the complete test record: precommitted hypothesis, sample, recruitment method, exclusions, exposure period, measured behavior, costs, contrary evidence and decision. Do not edit the threshold after seeing results without recording that it changed. A result that exceeds a threshold may still have poor external validity if participants were friends, existing partners or unusually motivated early adopters.

Small numbers require appropriate humility. If two out of five prospects commit to a paid pilot, report 'two of five' rather than describing a reliable 40% market conversion rate. The observed ratio is a fact; its transfer to a larger population is an uncertain inference. If the data cannot distinguish two plausible explanations, plan a test that can, or make a reversible decision within the uncertainty.

Use a simple evidence grade: **A** for repeated real behavior with traceable records; **B** for a meaningful but limited live test; **C** for interviews and credible external information; **D** for assumptions or unverified secondary claims. These labels are internal editorial tools, not statistical certifications. A safety, legal or rights issue can halt a venture even with excellent commercial evidence.

Decide: continue, change or stop

A continue decision means the next investment is justified by observed evidence and an explicit cost cap. A change decision means the hypothesis has been revised: perhaps the segment, offer, channel, economics or delivery method. A stop decision means the current experiment does not support more spending on that approach. A stop can preserve resources for a better idea; it does not necessarily imply personal failure.

Write a decision memo that a new team member can audit. Include what was tested, what happened, what was learned, what remains uncertain, the alternatives considered, and the date for review. Forecasts can be included only as assumptions with visible drivers. Never rewrite a past hypothesis to make a failed test appear successful.

Chapter workbook

Deliverable: A two-page validation decision memo. **Decision rule:** If evidence is weak and the next commitment is expensive or hard to reverse, reduce scope or seek stronger evidence first. **Exercise:** Ask a skeptical reviewer to produce the most convincing alternative explanation of your positive result.

CHAPTER 08

8. The Six-Week Validation Program

A program designed around learning cadence

Week 1: define the segment, quantify the problem hypothesis, identify participants and assemble the assumption ledger. Week 2: interview users, buyers and blockers; analyze recent episodes and current alternatives. Week 3: build and test a value proposition with a truthful prototype or manual demonstration. Week 4: present a fixed-scope commercial offer and observe real commitment behavior. Week 5: deliver a small pilot where lawful and operationally feasible; measure activation, time, cost and value. Week 6: review the entire evidence chain, make a decision, and prepare the next constrained experiment.

This schedule is an example, not an obligation. A regulated technology may need months of approvals before live testing; an enterprise buyer might require a longer procurement cycle. The discipline to preserve is weekly documentation and explicit decisions. Do not stretch a weak experiment indefinitely because the answer is disappointing. Distinguish delay caused by process from delay caused by lack of urgency.

Use three documents throughout: the ledger, the evidence log and the decision record. Do not build an elaborate research portal before the team has an effective habit of using these documents. Assign one accountable owner for evidence quality. Every published claim must be more conservative than the underlying evidence, not more ambitious.

What happens after validation?

Validation does not end with one go/no-go meeting. If customers demonstrate willingness to pay, the next questions concern delivery reliability, gross margin, recurring use and distribution. These connect this book to SF / 01, *The Venture-Ready Startup*, which covers the wider company readiness system. If evidence contradicts the starting idea, the next move is not necessarily to abandon the market. Return to the customer's costly job and test the most promising revised solution.

Illustrative closing decision. FlowDesk finds a repeatable process problem, two paying pilots, a clear approval buyer and several manual-delivery costs. It chooses a narrow workflow automation release rather than broad proposal-generation software. It still has not established scalable retention. That conclusion is more useful than claiming the product has been 'fully validated.'

Chapter workbook

Deliverable: A six-week experiment calendar with one owner, one decision per week, and resource caps. **Decision rule:** Allow ongoing experiments only when they continue to resolve a material uncertainty. **Exercise:** Record the one sentence you hope to be able to say truthfully after the next six weeks, and design your tests backward from the evidence that sentence requires.

REFERENCE MATERIAL

Appendix A. Assumption Ledger

A01

Hypothesis: Customer experiences problem weekly
Impact if wrong: High
Evidence today: Interviews only
Test: Process diary
Owner: TBD
Decision date: TBD

A02

Hypothesis: Buyer can authorize budget
Impact if wrong: High
Evidence today: Unknown
Test: Buyer interviews
Owner: TBD
Decision date: TBD

A03

Hypothesis: Pilot can be delivered profitably
Impact if wrong: High
Evidence today: Internal estimate
Test: Concierge pilot
Owner: TBD
Decision date: TBD

A04

Hypothesis: Buyers accept proposed price
Impact if wrong: High
Evidence today: Unverified
Test: Fixed-price offer
Owner: TBD
Decision date: TBD

A05

Hypothesis: Customers repeat the core task
Impact if wrong: High
Evidence today: No cohort
Test: Live usage cohort
Owner: TBD
Decision date: TBD

All entries above are **illustrative placeholders**, not findings from a real company. Replace with observed records and document the evidence standard.

REFERENCE MATERIAL

Appendix B. Interview Script and Observation Sheet

Opening and context

Explain the educational/research purpose and intended use of notes; request appropriate consent. 'Please walk me through your most recent attempt to [job].' 'What triggered it?' 'What happened first?' 'What did you use to complete the task?' 'Who was consulted or asked to approve?' 'What went wrong?' 'How did you resolve it?' 'What did it cost you?' 'When will this probably happen again?'

Buying and alternatives

'How was the current solution chosen?' 'Which alternatives were reviewed?' 'Who owns the budget?' 'What security, data or procurement reviews apply?' 'What would cause you to look for a new option?' 'Have you funded a solution to this problem before?' 'What would make a pilot impossible?'

Observation versus interpretation

Exact behavior observed: _____ **Participant's statement (if appropriately retained):** _____ **Founder interpretation:** _____ **Reason this may be misleading:** _____ **Additional test required:** _____

REFERENCE MATERIAL

Appendix C. Experiment Card and Results Log

Hypothesis:

Segment inclusion and exclusion criteria:

Commercial offer or product behavior tested:

Single primary measure and formula:

Minimum acceptable signal / threshold:

Sampling and recruitment method:

Start and end date:

Time, cash and legal/ethical constraints:

Observed numerator / denominator:

Dropout and non-response:

Contrary findings:

Decision and responsible owner:

Results table

Qualified response

- **Expected:** Precommitted value
- **Actual:** TBD
- **Evidence link:** TBD
- **Interpretation limit:** Sample may be biased

Meaningful commitment

- **Expected:** Precommitted value
- **Actual:** TBD
- **Evidence link:** TBD
- **Interpretation limit:** Contract conditions matter

Successful first value

- **Expected:** Precommitted value
- **Actual:** TBD
- **Evidence link:** TBD
- **Interpretation limit:** May require manual effort

Repeat use

- **Expected:** Precommitted value
- **Actual:** TBD
- **Evidence link:** TBD
- **Interpretation limit:** Cohort may be too small

REFERENCE MATERIAL

Appendix D. Sample Validation Case Files

Case 1 — Fictional B2B workflow product. The problem is delayed approval in proposals. The assumption is that a manager will pay to reduce approval time. A useful test is a priced two-week coordination pilot with a baseline. Evidence is paid participation, actual approval-cycle reduction, and implementation workload. What remains uncertain is whether the workflow can be automated affordably and renewed.

Case 2 — Fictional training platform. The problem is low course completion among professional learners. The team hypothesizes that structured weekly feedback increases completion. Run a pilot with informed participants, consistent learning outcomes and a comparison group where appropriate. Measure completed work and support time, not only login counts. Avoid claiming improved educational outcomes from an uncontrolled small trial.

Case 3 — Fictional agricultural marketplace. The problem is fragmented ordering for small wholesalers. Interviews show both buyers and sellers desire easier ordering, but one side demands credit the platform cannot lawfully provide. The appropriate decision is to test a non-credit ordering workflow and obtain specialist review before offering financial services. Positive interest does not override regulatory constraints.

REFERENCE MATERIAL

Appendix E. Definitions and Sources

Hypothesis: A falsifiable claim whose evidence may change a business decision. **Validation:** Adequate evidence for a specified decision under stated uncertainty; not a universal certification. **Activation:** First observable delivery of the product's core promised value. **Retention:** Continued relevant behavior or payment over meaningful intervals. **Concierge pilot:** A service test delivered partly or fully by people before automation exists. **Precommitment:** Recording test criteria before obtaining results. **Customer segment:** A group with sufficiently similar job, circumstances and buying behavior to support meaningful comparison.

Sources checked **30 October 2026**. They inform the methods, but no source endorses Silicon Founders or the fictional cases.

- [R1] Strategyzer, 'The Value Proposition Canvas': https://www.strategyzer.com/library/the-value-proposition-canvas
- [R2] U.S. National Science Foundation, 'About I-Corps': https://www.nsf.gov/funding/initiatives/i-corps/about-i-corps
- [R3] U.S. Federal Trade Commission, 'Advertising FAQs for Small Business': https://www.ftc.gov/business-guidance/resources/advertising-faqs-guide-small-business
- [R4] Stripe, 'Net Revenue Retention for SaaS Businesses': https://stripe.com/resources/more/net-revenue-retention
- [R5] U.S. Small Business Administration, 'Plan your business': https://www.sba.gov/counseling/plan-your-business/
- [R6] Y Combinator, 'Startup Library': https://www.ycombinator.com/library

Publication and rights note

PRACTICAL WORKSHEET / 01

Customer Segment Brief

Use this page as a reusable internal working tool. Record dates, sources and accountable owners; do not invent entries.

Prepared by: _____ Date: _____

Segment inclusion and exclusion rules

_____

_____

Core job and most recent costly episode

_____

_____

Current alternative and economic buyer

_____

_____

Reachable recruitment channel

_____

_____

Evidence gaps and next decision

_____

_____

PRACTICAL WORKSHEET / 02

Customer Interview Field Notes

Use this page as a reusable internal working tool. Record dates, sources and accountable owners; do not invent entries.

Prepared by: _____ Date: _____

Date, role, segment and consent record

_____

_____

Last episode in participant's own words

_____

_____

Actions observed versus reported

_____

_____

Workarounds, cost and decision roles

_____

_____

Contradictions and next test

_____

_____

PRACTICAL WORKSHEET / 03

Experiment Design Card

Use this page as a reusable internal working tool. Record dates, sources and accountable owners; do not invent entries.

Prepared by: _____ Date: _____

Single falsifiable hypothesis

_____

_____

Eligible sample and recruitment method

_____

_____

Offer, baseline and primary behavioral measure

_____

_____

Precommitted threshold, time and resource cap

_____

_____

Ethical/privacy issues and responsible owner

_____

_____

PRACTICAL WORKSHEET / 04

Evidence and Results Review

Use this page as a reusable internal working tool. Record dates, sources and accountable owners; do not invent entries.

Prepared by: _____ Date: _____

Actual numerator, denominator and exclusions

_____

_____

Preplanned threshold and achieved result

_____

_____

Alternative interpretations and sample bias

_____

_____

Money and time spent; operational burden

_____

_____

Evidence grade and decision memo link

_____

_____

PRACTICAL WORKSHEET / 05

Six-Week Learning Decision

Use this page as a reusable internal working tool. Record dates, sources and accountable owners; do not invent entries.

Prepared by: _____ Date: _____

Most important original assumption

_____

_____

What changed in the customer understanding

_____

_____

Validated behaviors and unproven claims

_____

_____

Continue / change / stop decision and why

_____

_____

Next experiment, budget cap and owner

_____

_____