Experimentation

Table of contents

Why experiment? . . . 1
Two types, two reasons . . . 1
The scientific method . . . 2
What makes a good hypothesis? . . . 2
Structuring a hypothesis . . . 3
Types of data collection . . . 3
Experiments should . . . 3
More tips . . . 4
Learning from experiments . . . 4
Key takeaways . . . 4

Why experiment?

To gain insights about customers, their problems, and our solutions — before spending the effort to build.

Two types, two reasons

Two types of experiment:

1. **Generative / exploratory** — explore a space.
2. **Evaluative** — find data to support or refute a hypothesis.

Two reasons to experiment:

1. Evaluate our understanding of our customers and market fit.
2. Evaluate products.

The scientific method

```
1. State hypothesis → 2. Design experiment → 3. Test with data → 4. Gather insights
```

The lean startup version of this loop (Steve Blank) expands into seven steps:

1. Extract hypotheses
2. Prioritise hypotheses
3. Design tests
4. Prioritise tests
5. Run tests
6. Capture learnings
7. Make progress — and repeat

Both prioritisation steps are easy to skip and shouldn't be: a team has more hypotheses than it has time, and testing the wrong one first wastes a week.

What makes a good hypothesis?

The full list: **testable, reliable** (not fake), **concise, replicable, specific, ethical**. Condensed to four:

	Property
#1	**Measurable/testable falsifiable**
#2	**Specific**
#3	**Ethical**
#4	**Simple and clear**

Falsifiability is the load-bearing one. An example of a *non*-falsifiable hypothesis:

> "This solution will help this segment of customers improve their health over the long term."

Nothing about it can come back false — "improve", "health" and "long term" are all unmeasurable as stated, so no result could ever refute it. The exercise is to rewrite it until a specific observation could prove it wrong.

Structuring a hypothesis

The test-cards-and-learning-cards[1] format gives hypotheses a fixed shape:

> We believe that [**target**] are [**showing behaviour / display interest in**] [**for reason Y**]. To verify that, we will [**undertake behaviour**] to [**identify behaviour**]. And measure [**particular variable**]. We are right if [**some measurable statement**]. Therefore, we will ask [**specific questions**].

Types of data collection

Six ways to gather evidence, roughly in increasing order of cost:

Method	What it is
Interviews	Talk with potential customers
Observations	Observe people while doing their "jobs"
Data collected by others	Find existing data that tests your hypothesis
Immersion	Work in your customer's environment
Low fidelity prototyping	Paper prototypes or low-fidelity digital prototypes
High-fidelity solutions	Implement new functionality in an existing product

See prototyping[2] for the last two.

Experiments should

- Be as **small and comprehensible** as possible.
- Have **clear expected outcomes**.
- Have a clear definition of **success/failure that everyone agrees to** — agreed *before* the results come in, or the team will argue about interpretation afterwards.
- Be as brief as possible when interviewing — **preferably under 30 minutes**.

[1] courses/comp1100/test-cards-and-learning-cards.pdf
[2] courses/comp1100/prototyping.pdf

More tips

Do: listen; observe people's behaviour, tone and reactions, not just what they say; ask "why" and "why not"; ask who else you should talk to; ask if you can follow up later.

Don't: "correct" people (except to clarify a misunderstood question); convince or sell to them.

Learning from experiments

When results come back, interrogate them rather than just recording them:

- If people would not pay for this, what *would* they pay for?
- If these are not our customers, then who is?
- Did we see any **outliers** that could open up new ideas?
- **Pay attention to why our hypotheses are rejected.**

A refuted hypothesis is more informative than a confirmed one — it tells you something you didn't already believe.

Key takeaways

- Experiments = gain insights about customers and products.
- Hypotheses must be **falsifiable**.
- **Test cards** structure our hypotheses; **learning cards** structure our learnings.
- Refuted hypothesis ask *why*.