16 / BOTTOM-UP EXPERIMENTATION

Intuit

From pilot to nationwide

When does a promising local pilot justify a broader launch?



Intuit headquarters, Mountain View, 2008. Photograph: Coolcaesar.
Wikimedia Commons · CC BY-SA 3.0. Cropped and reproduced in black and white. This image adaptation is also CC BY-SA 3.0.

A mobile tax-preparation idea must move beyond convenience to demonstrate eligibility, accuracy and support across a larger customer population.

The SnapTax product leader and executive sponsor | Late 2010 • reconstructed SnapTax decision

35-45 minutes to prepare / 75 minutes to discuss
Original public-source teaching case • First member edition • 1 November 2026

PRINCIPLED BUSINESS IN PRACTICE

Introduction

About Principled Business

Principled Business (PB) is a nonprofit organization that brings together business leaders to advance capitalism through education, relationships and action. It helps founders, executives and investors understand free enterprise, apply its principles in their organizations, and explain the contribution business makes to people's lives.

The PB manifesto grounds capitalism in individual agency, private ownership and voluntary production and trade. These commitments take practical form in everyday business decisions.

Bottom-up experimentation

PB recognizes that useful knowledge often emerges through local initiative and trial. Leaders must make room for experiments while defining the evidence and safeguards that govern further investment.

SnapTax tests the discipline required to scale an experiment. A local pilot can reveal a useful idea while leaving important questions about different users and operating conditions unanswered. Students must decide what evidence is sufficient for expansion when errors can impose real costs on customers.

PB presents this case to make the principle usable in a real business decision. Your recommendation should connect the principle to a workable policy, account for the people affected, and identify evidence that could change your mind.

What you will learn

01 Compare completion and support measures using consistent denominators.

02 Distinguish a promising point estimate from evidence sufficient to meet a quality gate.

03 Design a staged rollout that tests generalization and protects customers.

> **Apply it to your own business**
> Which pilot in your organization is being considered for wider rollout? What differs between the pilot users and the next population, and which difference could matter most?

Source: About Principled Business; Capitalists for Capitalism Manifesto, Section 2: Bottom-Up Experimentation.

NOTES FOR STUDENTS AND TEACHERS

Using this case study

This packet includes the case, exhibits, worksheet, sources, answers and teaching notes. Allow 35-45 minutes to prepare and 75 minutes for discussion. Have a calculator or spreadsheet available.

Pages	What to use and when
4-8	Decision, evidence, exercise and alternatives. Read before discussion.
9	Decision worksheet. Complete before reading the answers.
10	Sources and reading notes. Check the evidence and its limits.
11-12	Worked answers, outcome and teaching plan. Read after deciding.

For students

Read once to identify the decision, then work through the exhibits. Separate historical facts from the invented classroom exercise. Use the case's decision boundary; write down assumptions when evidence is missing. Complete the calculation and worksheet on page 9. Be ready to defend your recommendation and the strongest objection to it.

For individual study, decide before turning to pages 11-12. Compare your reasoning with the worked analysis and write a short reflection on what you would change. Choose one action to try in your own organization.

For teachers and discussion leaders

Read the full packet and timed plan before teaching. Assign pages 1-10 first and reveal the final two pages after students commit to a choice. Record the alternatives, evidence and disagreements on the board.

Treat accuracy and appropriate eligibility as launch conditions. The exercise is not tax advice, and its invented audit rates cannot be used to characterize Intuit's historical product quality.

> **How to assess the work**
> Look for a clear choice, accurate analysis, stated assumptions, attention to affected people, a fair objection, and an implementation plan with an owner and review point. Assess reasoning rather than agreement with the historical decision.

LATE 2010 • AFTER A CALIFORNIA INTRODUCTION

The decision

SnapTax introduced a mobile approach to simple tax returns in California in 2010. Users could photograph a W-2, review the information and complete filing through the application. The later nationwide announcement described a limited initial return scope. [N1]

Take the role of a product leader and executive sponsor preparing the next tax season. The pilot suggests that mobile entry could reduce effort, but a national release would reach people with different circumstances, devices and support needs. The team must decide what to expand and what to keep deliberately narrow.

Choose the rollout boundary

Recommend a nationwide release for a tightly defined eligible group, a staged expansion or another limited pilot. Explain the accuracy evidence, eligibility checks and support capacity required before more people use the product.

No internal Intuit pilot results are available in this case. The next exhibits provide a fictional dataset for practicing the decision. The scenario should not be read as a claim that the real product had the errors or support rates shown.

> **Decision boundary**
> Convenience and revenue cannot compensate for a failure to meet the required accuracy and eligibility conditions. Define those conditions before evaluating the rollout’s commercial appeal.

THE BUSINESS LOGIC

A pilot has a population

The next users may be different

A local pilot can underrepresent harder cases, less familiar devices or people who need more help. Expanding geographically is therefore more than multiplying a conversion rate. Identify which user characteristics affect the service and whether the next group is represented in the evidence.

Completion can conceal friction

A completion rate among people who start excludes those screened out or discouraged before starting. Both views can be useful, but they answer different questions. Support demand and abandonment reasons help explain why a seemingly simple process remains difficult for some users.

Quality requires its own evidence

A convenient interface does not establish accurate results. Independent checks, clear review steps and a way to correct errors are part of the operating model. A small audit provides a sample estimate, not certainty about all users. The team must decide how much evidence is appropriate for the consequences of a failure.

A staged rollout can vary one important dimension at a time, preserve a route to a suitable alternative and reveal support bottlenecks. It should have an explicit stopping rule; otherwise “staged” may merely describe a slower version of an already predetermined national launch.

> **The scaling question**
> What changed between the pilot and the next cohort, and what evidence shows that the service still works under those changed conditions?

EXHIBIT 1 • HISTORICAL CONTEXT

What the public record establishes

Historical product feature [N1]	Question before expansion
California introduction in 2010	Which users and circumstances did the first offer represent?
Photo entry of W-2 information	How will users review and correct captured information?
A limited simple-return scope	How will ineligible users be identified and directed appropriately?
A mobile filing workflow	What support and recovery process is available when a user cannot finish?

Design the next cohort

Specify the user groups, device conditions and support channels to include. Record exclusions explicitly. A narrow eligible population can be a responsible product boundary if users understand it and have a suitable next step when they do not qualify.

Separate three gates

A commercial gate asks whether enough users value the offer at its price. An operational gate asks whether the team can support them. A quality gate asks whether the service meets its required accuracy and appropriateness conditions. Passing one does not establish the others.

> **No hidden pilot evidence**
> The following counts are invented. They are designed to show why a larger number of completed filings can coexist with weaker conversion and more support work.

EXHIBIT 2 • INVENTED CLASSROOM EXERCISE

Read the pilot carefully

Compare fictional cohorts A and B. B represents a broader population. An independent review labels each audited filing either accurate or requiring correction. Assume audits are randomly selected from completed filings, but do not treat these small samples as conclusive population estimates.

Measure	A	B
Eligible people offered the service	1,000	2,000
Started	800	1,600
Completed filings	600	1,000
Audited filings / accurate	120 / 118	200 / 196
Support tickets per completed filing	0.25	0.40

For the economic model only, charge $15 per completed filing, incur $3 variable processing cost and $8 per support ticket. Each cohort costs $10,000 to run. These rounded prices and costs are invented. A hypothetical launch gate requires at least 99.5% observed audit accuracy, plus adequate statistical evidence and operational review.

01 Calculate completion among eligible people and among starters.

02 Calculate observed audit accuracy. Does either cohort pass the stated point-estimate gate?

03 Calculate contribution before pilot cost and the net pilot result.

04 Recommend the next test or rollout action.

ALTERNATIVES • IMPLEMENTATION • ASSIGNMENT

Scale only what the evidence supports

Path	Potential benefit	Required discipline
Nationwide, narrow eligibility	Reaches more suitable users while limiting product scope.	Evidence that accuracy and support hold across the included population.
Staged expansion	Tests differences between populations with bounded exposure.	A meaningful gate and the authority to pause.
Another limited pilot	Resolves a specific quality or usability uncertainty.	A defined question rather than indefinite delay.

Your assignment

01 Choose the next action and identify the first unresolved gate.

02 Specify how eligible and ineligible users will be treated.

03 Design independent quality checks, correction procedures and support capacity.

04 State what evidence would justify a larger cohort and what result would stop it.

Write an executive recommendation that includes both the attractive result and the most important weakness. If you propose a further test, specify the population and uncertainty it will address. Do not use additional volume to hide a quality shortfall.

> **A decision to resist**
> “The larger cohort produces more total revenue, so it is ready to scale.” Explain the missing reasoning before approving the next commitment.

WORKSHEET • COMPLETE BEFORE READING THE TEACHING NOTES

Your decision

Write as the decision-maker. Use the historical evidence and the classroom model for their stated purposes. A recommendation needs a boundary, an owner and a test.

My recommendation and the principle it puts into practice

The calculation or evidence that matters most; its main limitation

Who bears the risk, and how their choice or voice is protected

The strongest objection and my response

First action, accountable owner, review date and stop/change trigger

EVIDENCE • METHOD • LIMITS

Sources and reading notes

A reconstructed late-2010 product decision informed by the January 2011 launch release, which confirms a prior California introduction. The decision roles are generic; the pilot data and quality gate are PB inventions.

N1 Intuit TurboTax, Taxes on Your Mobile Phone? It’s a Snap, 14 January 2011

Confirms California introduction in 2010, nationwide iPhone/Android availability in 2011, W-2 photo entry and the limited initial return scope.

Open source document →

N2 Principled Business, Capitalists for Capitalism Manifesto

Bottom-up experimentation provides the interpretive framework.

Open source document →

How to read the evidence

The release is a product announcement, not an independent evaluation of accuracy or pilot results. Historical tax forms and prices describe the period only. No current filing instructions or tax eligibility advice is offered.

All quantities in the classroom exercise are invented unless explicitly identified as historical. The models omit factors deliberately for discussion; they are not company forecasts. Photographs are documentary context, with dates and licenses on the cover.

Original Principled Business teaching material. No company approval or affiliation with Harvard Business School is implied. No private interviews or unpublished company records were used. Public sources checked 1 November 2026.

TEACHING NOTES • READ AFTER COMPLETING THE DECISION

Worked analysis

The denominator changes the story

Result	A	B
Completion / eligible	60%	50%
Completion / starters	75%	62.5%
Observed audit accuracy	98.33%	98.00%
Contribution before pilot cost	$6,000	$8,800
Net pilot result	−$4,000	−$1,200

A produces 150 support tickets and B produces 400. Contribution is completed filings × ($15 − $3), less tickets × $8. Contribution per completed filing is $10 in A and $8.80 in B. B's larger total does not mean it serves each customer more efficiently.

Neither cohort passes the stated gate

118 ÷ 120 and 196 ÷ 200 both fall below 99.5%. Even a sample at or above the point threshold would still require adequate statistical evidence and operational review under the stated rule. The case does not prescribe a universal sample size or a real tax-product standard.

A defensible next step

Investigate correction types, improve the relevant process and run a targeted, independently reviewed test before expansion. Preserve suitable alternatives for users who do not qualify. The next cohort should represent the conditions the team intends to support, with a pre-agreed pause rule.

> **What not to infer**
> These invented rates say nothing about Intuit's actual product quality. They illustrate why growth, convenience and positive contribution cannot substitute for a required quality condition.

TEACHING NOTES • READ AFTER COMPLETING THE DECISION

Outcome and teaching plan

What happened next

On 14 January 2011, Intuit announced nationwide SnapTax availability for iPhone and Android after the prior California introduction. The release described simple returns and a review-before-filing workflow. It did not publish the pilot dataset used in this exercise. [N1]

A 75-minute discussion

Minutes	Facilitation plan
0-10	Choose a rollout boundary and state the first gate.
10-25	Compare pilot and target populations.
25-40	Calculate conversion, audit accuracy and support economics.
40-55	Design the next test, correction route and pause authority.
55-65	Reveal the national announcement; separate a launch from proof of every performance claim.
65-75	Apply the scaling questions to another pilot under review.

> **Closing question**
> What evidence must survive a change in population before a successful pilot becomes a responsible rollout?

Assess the decision at its stated boundary. Later success cannot prove the original choice was sound; later failure cannot, by itself, prove it was unreasonable.