

Retail Technology Pilot Playbook

How to structure pilots that produce learning — not just delay

Stern Technology Advisory · sterntechadvisory.com · Morris Stern

The problem with retail technology pilots

Most retail technology pilots are structured to produce confidence — not learning. The vendor gets a favorable store, the timeline is compressed to move quickly, and success is measured by whether the store manager says they liked the new system. Six months later, you have a positive anecdote, no comparative data, and no clear decision framework. The pilot has consumed budget and time and produced almost nothing useful.

This playbook is about structuring pilots differently — to produce actual learning, decision confidence, and a clear go/no-go framework.

Chapter 1 — Before the pilot starts

Define what you are actually testing

Most pilots fail to define hypotheses before they start. A hypothesis is not "we want to see if this works." It is a specific, testable claim:

"Replacing the current tablet-based POS with [Platform X] will reduce average transaction time by 15% and reduce associate onboarding time from 4 days to 2 days."

For every pilot, write down:

- **The specific claim** being tested
- **How you will measure it** (what data, collected how, over what period)
- **The threshold** that constitutes a pass vs. fail
- **The confounding factors** you will need to control for

Select the right pilot environment

Vendor-recommended pilot stores are the wrong starting point. You want stores that represent reality — not best-case conditions.

Good pilot store criteria:

- Representative volume (not your highest or lowest traffic location)
- Representative associate tenure mix (not your most tech-savvy team)
- Representative operational complexity (typical integrations, not a simplified environment)
- Management team willing to report problems honestly

Avoid:

- Flagship stores with atypical staffing or foot traffic
- Stores where the manager is a personal advocate of the vendor
- Stores undergoing other operational changes simultaneously

Set the timeline — and hold it

Pilots need a defined end date. Open-ended pilots drift. If the vendor asks for "a few more weeks," that is information: something is not working and they need more time to fix it.

Define at the start:

- Pilot start date (after full installation and training — not the installation date)
- Minimum run period (for retail, typically 60–90 days of operational data)
- Go/no-go decision date

Chapter 2 — What to measure

Operational metrics

Metric	Why it matters	How to measure
Transaction time	Associate productivity, customer experience	POS transaction logs vs. baseline
System uptime	Operational reliability	Downtime incidents during pilot period
Associate call volume	Support burden, ease of use	IT help desk tickets by category
Error rate	Data quality, integration reliability	Void/return rates, manual corrections
Inventory accuracy	Integration reliability	Physical count vs. system count

Financial metrics

Metric	Why it matters	How to measure
Sales per labor hour	Overall operational productivity	Compare pilot stores vs. control stores
Shrinkage	Operational control	Inventory variance vs. baseline period
Integration cost	Hidden implementation cost	Time spent on manual workarounds

People metrics

Metric	Why it matters	How to measure
Associate satisfaction	Adoption sustainability	Simple pulse survey (5 questions, every 3 weeks)
Training time to proficiency	Onboarding economics	Time to first solo transaction
Manager reported issues	Qualitative operational signal	Weekly manager call with structured agenda

Integration reliability (often ignored in pilots)

- How many times did the integration between the new system and ERP/inventory fail?
- How many transactions required manual reconciliation?

- Were there data integrity issues (duplicate records, sync failures)?
- What is the queue depth in the integration layer during peak periods?

Chapter 3 — Structuring the go/no-go decision

The 80% confidence threshold

Retail technology programs cannot wait for certainty. Waiting for 100% confidence is how pilots turn into permanent projects. The right threshold is 80% confidence on the critical metrics — enough to proceed with a defined risk management plan for the open questions.

Decision framework:

Zone	Criteria	Decision
Green	≥80% of metrics meet threshold; no critical failures	Proceed to rollout planning
Yellow	60–79% of metrics meet threshold; issues are fixable	Expand pilot with specific fixes required
Red	<60% of metrics meet threshold; or critical integration failure	Stop; re-evaluate vendor or scope

What to document before the decision meeting

1. **Metrics summary** — actual vs. target for every metric defined in Chapter 2
2. **Issue log** — every problem encountered, how it was resolved, and whether it is resolved
3. **Open items** — unresolved issues with owner, resolution plan, and timeline
4. **Vendor commitments** — what the vendor has committed to fixing, with dates
5. **Associate sentiment summary** — key themes from pulse surveys and manager calls

Red flags that override green metrics

Even if metrics look good, stop and investigate if you see:

- Integration failures that were manually resolved without the vendor's knowledge
- Associate workarounds that are not captured in transaction data
- Vendor representatives who are embedded in the store managing issues vs. the system running independently
- Data that looks correct in aggregate but has integrity issues at the transaction level

Chapter 4 — Scaling from pilot to rollout

What a pilot proves (and what it doesn't)

A successful pilot at 3–5 stores does not prove the system will work across 200+ locations. It proves the system can work under managed conditions. Rollout introduces:

- **Variability** in store layout, floor plan, and infrastructure
- **Variability** in associate capability and manager engagement
- **Scale pressure** on integrations and infrastructure
- **Support ticket volume** that is 10–50x the pilot period

Before rollout, verify:

- Support SLAs are defined for rollout volume, not pilot volume
- Infrastructure requirements have been validated across store format variability
- Training program has been piloted and produces consistent results
- Rollback procedure is defined and tested
- Integration performance has been load-tested at full rollout scale

Deployment wave structure

Wave 1 (5–10 stores): Controlled rollout with full support coverage. Goal: validate deployment process and training program.

Wave 2 (25–50 stores): Regional rollout. Reduce support coverage. Identify regional variation in performance.

Wave 3 (remaining stores): Full deployment at operational tempo. Support model transitions to steady-state.

Between each wave: Hold a 2-week retrospective. Do not start the next wave until the previous wave's critical issues are resolved.

Common pilot failures and how to avoid them

"The vendor managed the pilot." If the vendor's team is on-site managing issues during the pilot, you are not testing the system — you are testing the vendor's ability to staff an escalation. Structure pilots to run without vendor presence after the first two weeks.

"We measured the wrong things." Customer-facing metrics (satisfaction scores, NPS) are not the right primary metrics for a technology pilot. Measure operational reliability, integration performance, and associate productivity.

"We ran the pilot too short." 60 days of operational data is a minimum. Technology behavior in week 1 is not predictive of week 8 behavior. Seasonal retailers need to ensure the pilot overlaps with a representative traffic period.

"We didn't have a control group." Without a comparable group of non-pilot stores as a baseline, you cannot separate pilot performance from general business performance. Define your control stores before the pilot starts.

About this playbook

This playbook is drawn from direct experience piloting and deploying retail technology across 300+ locations — including POS replacements, digital price tag deployments, and mobile associate tools. The frameworks described here were developed to produce real decisions, not just optimistic case studies.

Morris Stern — Technology Modernization Consultant, Stern Technology Advisory

Former VP of Retail Systems, Ashley Global Retail. Led store technology programs supporting ~\$3B in annual revenue.

sterntechadvisory.com · morris@sterntechadvisory.com