

Measuring UX Design Outcomes: Overview



Outcome measurement answers: did what we built actually help users?



Outputs are shipped features; outcomes are behavior, attitude, and business change



Measurement loop: goal, signal, metric, instrumentation, analysis, decision, iterate



Watch for vanity metrics, Goodhart's Law, correlation mistaken for causation



You will pick metrics, plan instrumentation, design studies, and report results



Scope: one feature, flow, or design change — not company-wide dashboards



Outputs, Outcomes, and the Measurement Loop



Output: we shipped a redesign.

Outcome: users complete the task unaided.

Impact: support costs fell



Page views and unique users rarely evaluate a UX change — too general, not tied to experience goals



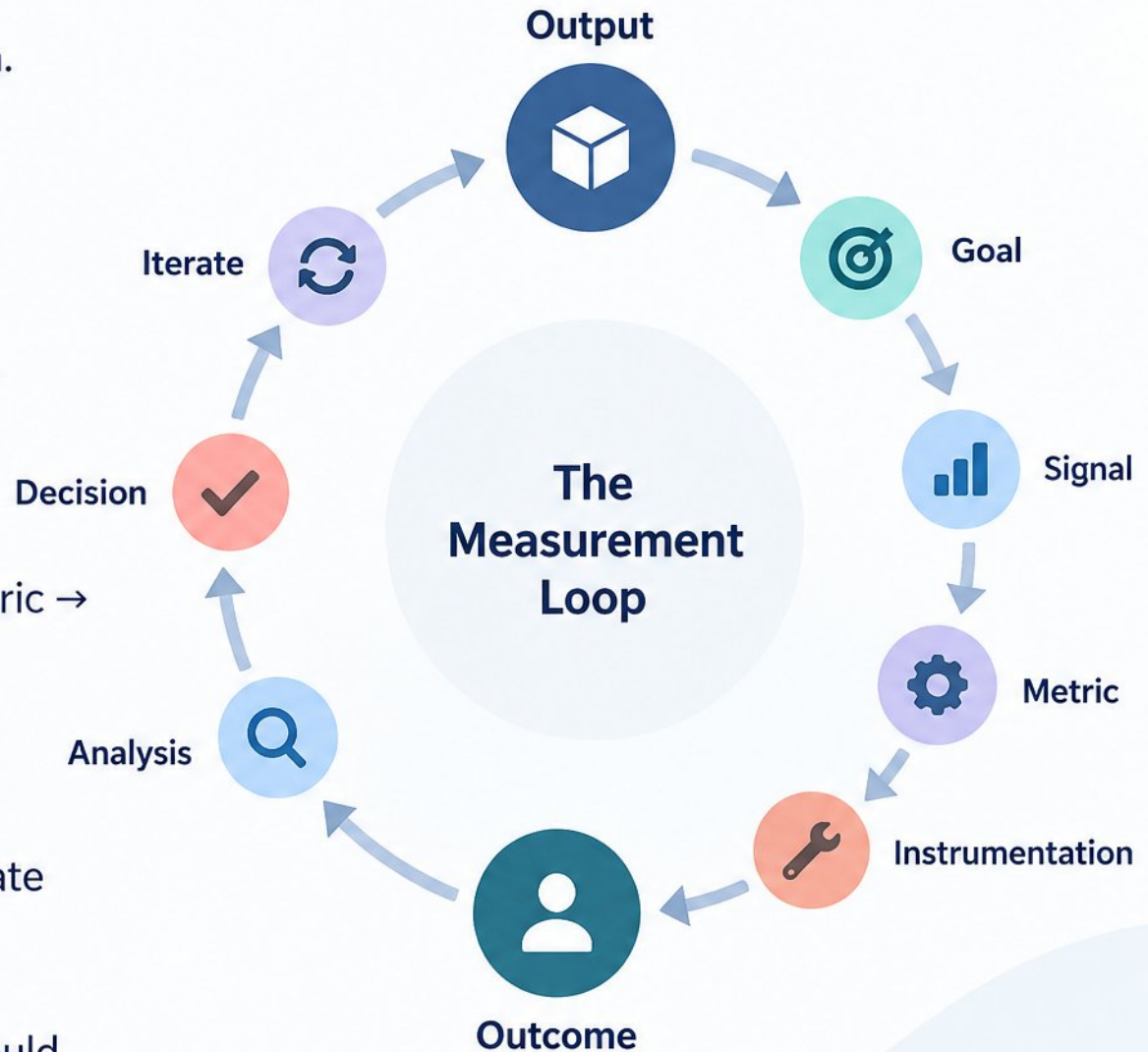
The loop: goal → signal → metric → instrumentation → analysis → decision → iterate



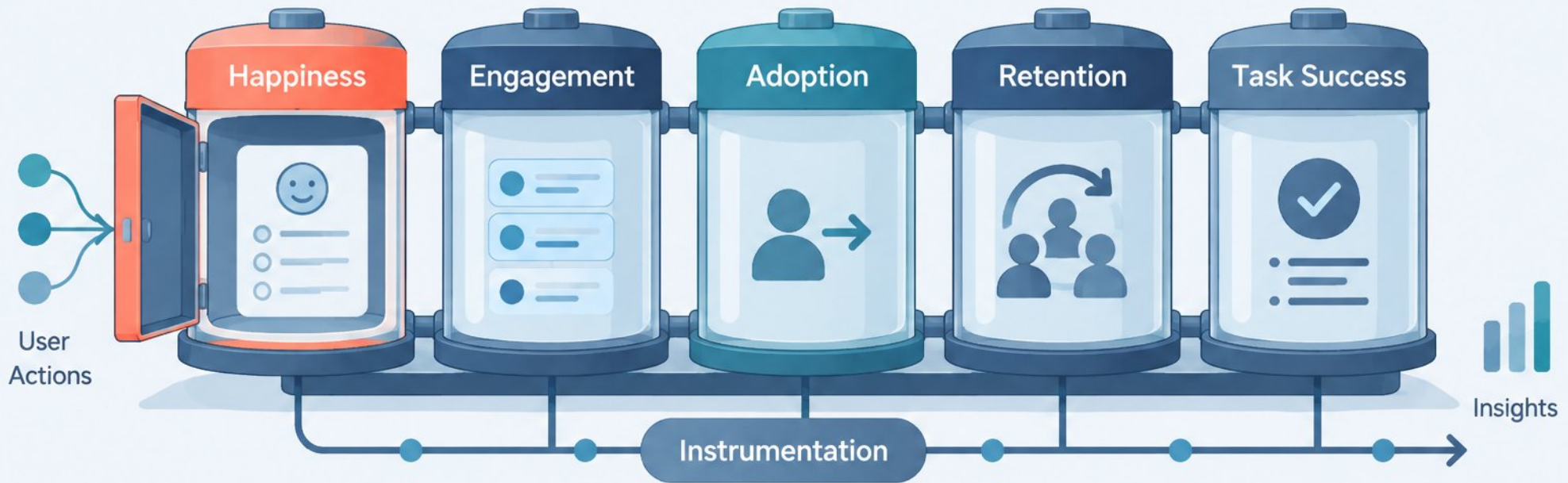
Treat measurement as a UX research method, not a separate analytics function



Decision test: if a number would not change what you do, cut it from the dashboard

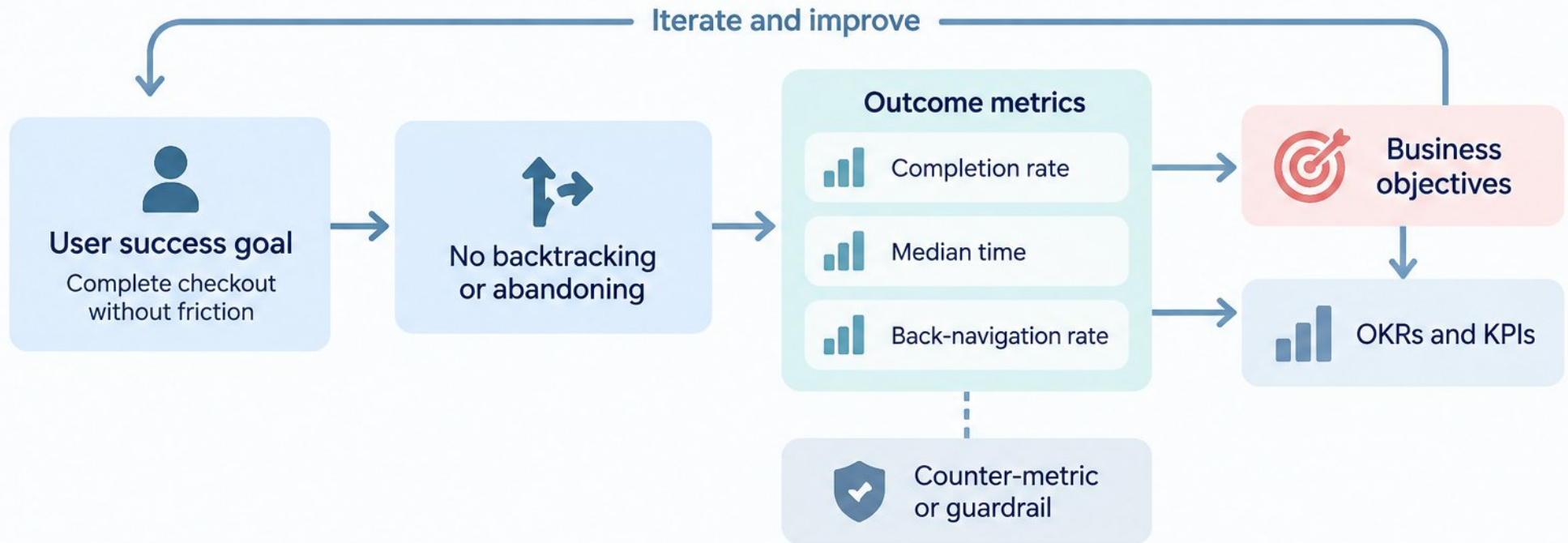


The HEART Framework and Goals-Signals-Metrics



- HEART's five categories: Happiness, Engagement, Adoption, Retention, Task Success
- Happiness is attitudinal — it requires an ask, so surveys count as instrumentation
- Goals-Signals-Metrics: state the goal plainly, name signals, then pick the number
- Choose signals that are easy to track and sensitive to design changes
- Scope HEART to one feature, flow, or launch — not the whole product
- Omit irrelevant categories: new feature launch = Adoption + Task Success

From Design Intent to Measurable Outcomes



Express outcomes as ratios with a stated population, not raw counts



Write goal statements about user success first, then link to business objectives



Worked chain: complete checkout without friction → no backtracking or abandoning → completion rate, median time, back-navigation rate



Avoid goals defined as existing metrics, like "increase traffic"



Ladder metrics to OKRs and KPIs so numbers survive budget talks



For every metric you optimize, name a counter-metric or guardrail

Choosing the Right UX Metrics



Behavioral

Shows what happened



Task success rate



Time on task



Error rate



Attitudinal

Explains how it felt

SUS: 10 items, 0-100
(not %); 68 average,
below 51 serious

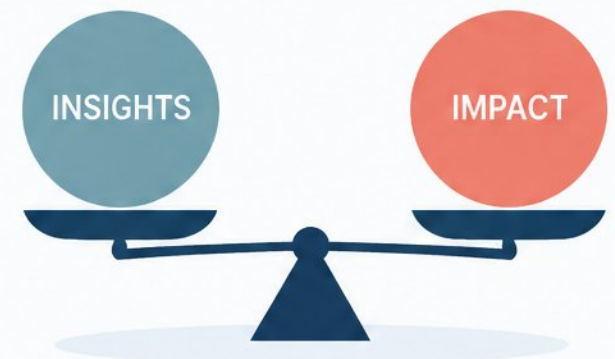


Desirability



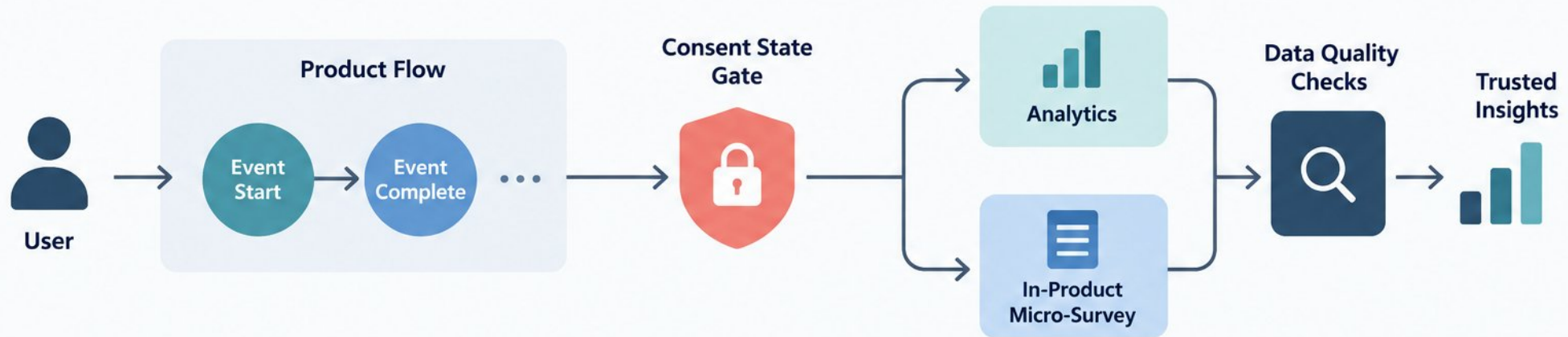
Business outcome

- Four lenses: behavioral, attitudinal, desirability, business outcome
- Behavioral shows what happened; attitudinal explains how it felt
- SUS: 10 items, 0-100 (not %); 68 average, below 51 serious
- NPS tracks relationship trends; CSAT scopes to one moment
- Task success rate plus time on task and error rate
- Match the metric set to product stage and decision context



Instrumentation and Data Collection

From user behavior to trustworthy insights



- Turn hypotheses into an event taxonomy: named start/complete pairs, documented funnels



- Behavioral metrics live in analytics; Happiness and perceptual task success need surveys



- Data minimization by default: collect only events that answer a decision question



- Emit consent state, timestamp, and policy version as events; enforce at every export



- Handle consent gaps by segmenting populations, using aggregate modes, tracking consent trends



- QA checks: missing events, sampling bias, duplicates, instrumentation changes faking product changes

Baselines, Targets, and What to Fix First

Measure with intention. Prioritize what matters. Turn insights into action.



Baseline before shipping: without a pre-redesign reading, the post-launch number is meaningless



Set target bands, not single numbers — e.g., 90-95% task success



Prioritize by impact, confidence, effort, and measurability



Minimum viable scorecard: one outcome, two behavioral, one attitudinal, one guardrail



Keep dashboards to 6-8 metrics, mixed types, trends over snapshots



Pair every reported number with an action, or it won't survive review

Minimum Viable Scorecard

Outcome metric



Behavioral metric 1



Behavioral metric 2



Attitudinal metric



Guardrail metric



Baseline (before shipping) Target band Latest

Prioritization Framework

Impact	Confidence	Effort	Measurability



Use this framework to focus on the highest value opportunities.

Baseline

Target band

Take action

Better outcomes

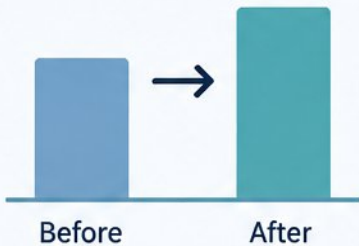
Study and Experiment Designs for Design Changes

- ✓ Pre/post, A/B, and holdout designs each answer different questions
- ✓ No randomisation? Use diff-in-differences or interrupted time series
- ✓ Lock one hypothesis, one primary metric, and guardrails before launch
- ✓ Randomisation unit must match analysis unit; run an SRM check
- ✓ Do not A/B test rare events, obvious bug fixes, or brand trust



COMMON EXPERIMENT DESIGNS

Pre/post design



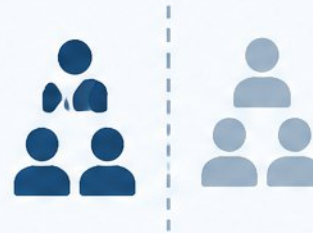
Compare outcomes before and after a change.

A/B design



Randomly assign users to different experiences.

Holdout design



Keep a holdout group to measure overall impact.

QUASI-EXPERIMENTAL TOOLKIT



Difference-in-differences



Interrupted time series



Matched cohorts



Lock hypothesis and primary metric



Run your design



Check guardrails and run an SRM check

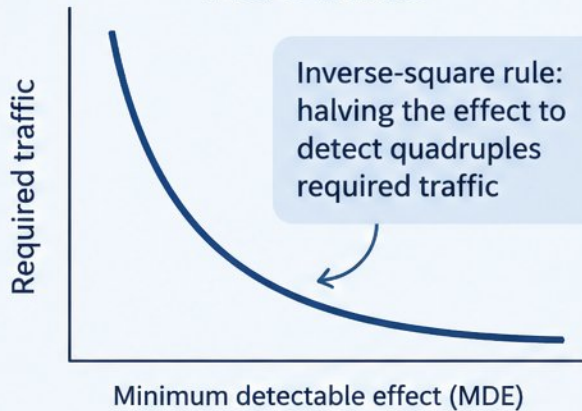


Make a decision

Sample Size, Power, and Practical Significance

From statistical rigor to real-world decisions

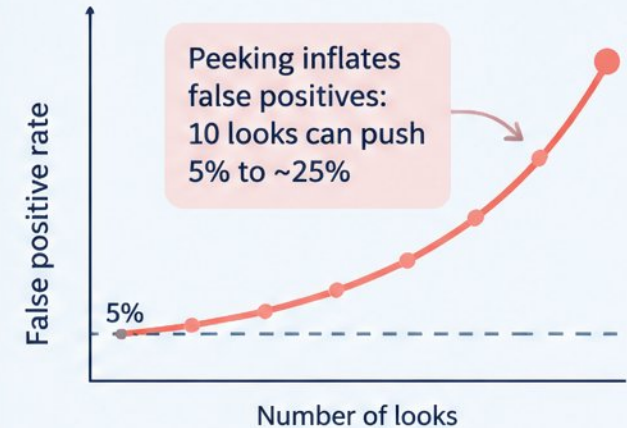
Smaller Effects Require More Traffic



Set MDE from the Decision



The Peeking Problem



Four inputs drive sample size: baseline rate, MDE, 5% significance, 80% power



Inverse-square rule: halving the effect to detect quadruples required traffic



Set MDE from the decision — smallest lift actually worth shipping



Peeking inflates false positives: 10 looks can push 5% to ~25%



Run one to two full business cycles so weekday and monthly patterns average out

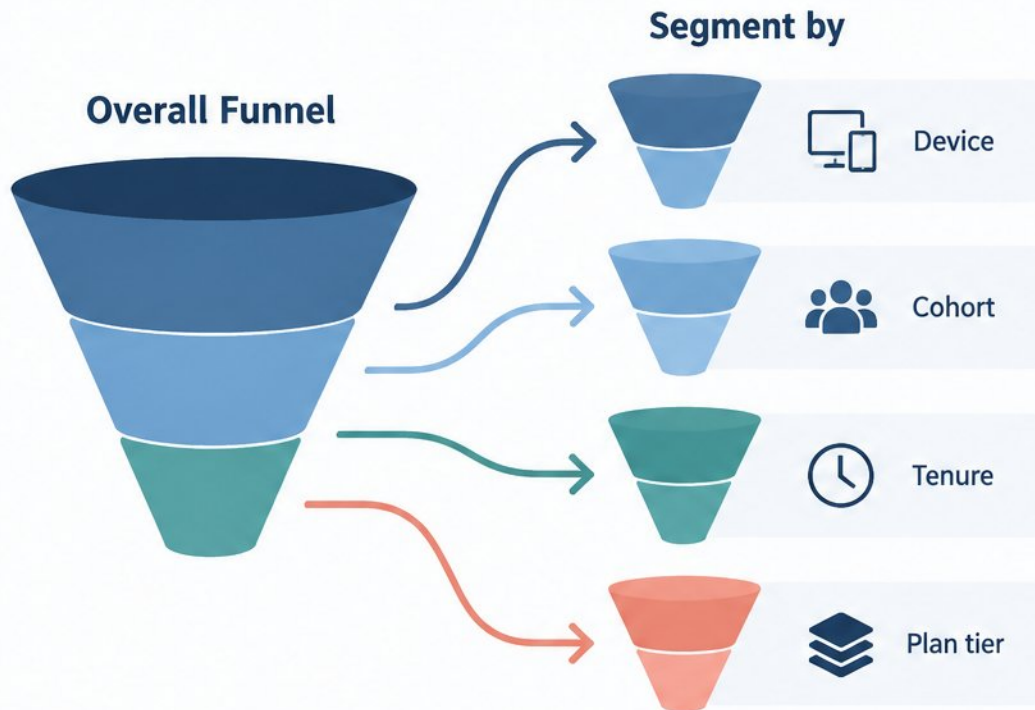


Significance is a gate, not a goal: pair p-values with confidence intervals

Analyzing and Interpreting Results

FROM DATA TO MEANINGFUL INSIGHTS

SAME DATA.
DEEPER ANSWERS.



Describe distributions and funnel steps before running any test



Segment by cohort, device, tenure, and plan tier; aggregates hide signal



Pair behavioral and attitudinal data; divergence is a finding, not noise



High task success with high churn?
You optimized the workaround, not the job



Check novelty, seasonality, selection bias, and regression to the mean



Report what changed, for whom, by how much, with what uncertainty

Behavioral vs. Attitudinal Data



Watch for Validity Threats



Novelty



Seasonality



Selection bias



Regression to the mean

Communicating Outcomes to Stakeholders



Layered report: recommendation, findings, data, methodology appendix



Lead with the headline and its implication, not the method



Translate friction into conversion, churn, support cost, CSAT



Tailor depth: executives one page plus the ask



Show trends and comparison points; annotate consent or tracking shifts



End with named owners and dates, even for negative results



Common Pitfalls and Ethical Considerations

1

Vanity counts always rise; when a measure becomes a target, it stops measuring (Goodhart's Law)

2

Tunnel vision on one metric versus metric overload — both fail; keep a small primary set

3

Dark patterns lift engagement while trust, sentiment, and support tickets deteriorate

4

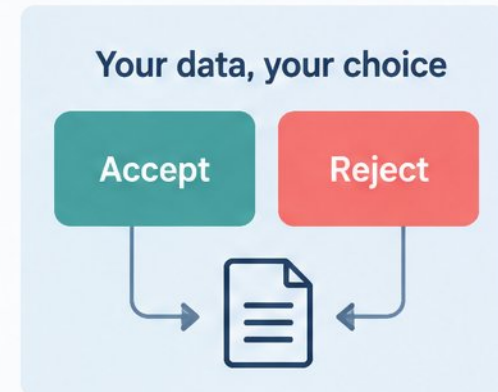
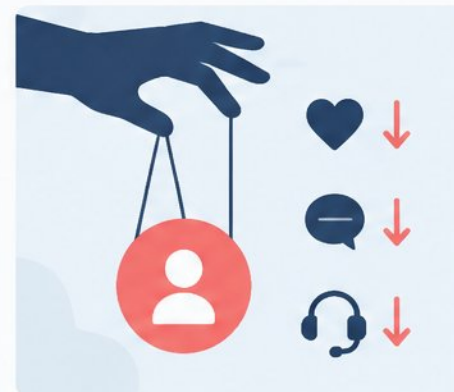
Pair business outcomes with user outcomes: retention beside ease of cancellation

5

Consent telemetry turns choice into an event stream you can audit and replay

6

Consent rate shifts often signal a collection change, not a product change



Operationalizing Outcome Measurement

From ownership to action, with a clear path to scale.



Roles: who owns the goal, events, dashboard, and review



Artifacts: measurement plan template, event naming convention, review cadence



Start with one receptive team; avoid organization-wide dashboards on day one



Maturity levels — foundational, managed, advanced — pick the honest next step



Scale via shared event definitions, roll-up dashboards, consistent metric definitions



Retire indicators that no longer correlate with outcomes



Workshop: Designing a Measurement Plan

- Pick a recent design change; state the decision it should inform
- Draft: goal, signals, metrics with formulas and populations
- Add baseline, target band, instrumentation, analysis, guardrails
- Include one behavioral, one attitudinal, one business metric
- Peer review: population stated? baseline defined? stopping rule? decision attached?
- Action plan: take it back within two weeks; name first instrumentation gap



Measurement Plan Worksheet



Decision



Goal



Signals



Metrics

(with formulas
and populations)



Baseline



Target band



Instrumentation



Analysis



Guardrails

Example Plans



Onboarding
Flow



Checkout
Flow

Peer Review Checklist

- ☐ Population stated?
- ☐ Baseline defined?
- ☐ Stopping rule?
- ☐ Decision attached?



Action plan: take it back within two weeks; name first instrumentation gap

PersonWise.

PersonWise • AI digital-human course platform



Scan the code or click anywhere to watch this course live, with voice Q&A

personwise.ai/t/64ada623/public