Evaluating the Quality of Customer Education

LEARN
MEASURE
IMPROVE
TOGETHER

- Only 19% of programs meet maturity benchmarks; 51% claim maturity
- Most measure reach and completion; just 37% see the full learner journey
- Education leads own strategy; CS owns adoption; L&D owns measurement rigor
- Roadmap: define quality, build framework, select metrics, report, improve
- Evaluation is a continuous system with owners and cadence



BETTER LEARNERS
BRIGHTER OUTCOMES

The 2026 Context: Why Measurement Is the Bottleneck

- Customer education now treated as a retention and revenue lever
- Formalized programs show ~7.4% retention lift and large adoption gains
- 52% of leaders are "very confident" they can prove ROI—yet
- 31% lack analytics capability, 20% run disconnected systems, 21% have no agreed success measures
- Only ~3% consistently link learning programs to business outcomes
- AI shifts measurement from counting activity to predicting outcomes
- Core pain points: disconnected LMS/CRM/product data and unclear data ownership



Defining Quality: Output Quality vs. Outcome Quality



Output quality

The crafted lesson

- Output quality: content accuracy, instructional design, currency, accessibility, learner experience



Outcome quality

The changed customer behavior

- Outcome quality: skills acquired, in-product behavior change, adoption, retention, cost



Define standards per program type—tutorial, onboarding path, certification differ



Borrow from L&D: Kirkpatrick levels, Phillips ROI, instructional design criteria



Shared vocabulary aligns education, CS, product, and L&D on "good"

The Four Levels: Applying Kirkpatrick to External Customer Learners

Design tip: start at Level 4, work backward from the business problem



Level 4 Results:
renewal, retention, expansion, support cost, NPS

Level 3 Behavior:
trained vs. untrained tickets, feature adoption, correct administration

Level 2 Learning:
assessments, scenario performance, certification pass rates

Level 1 Reaction:
satisfaction, engagement, relevance—not just a smile sheet

Common Failure Modes in Quality Evaluation



Vanity metrics:

10,000 completions or 85% satisfaction reveal nothing about capability



Completion-rate illusion:

completion measures compliance, not capability—high completion can mask rising tickets and churn



Satisfaction-only evaluation:

CSAT and NPS are leading indicators, not proof of learning



Attribution overreach:

correlating training with renewal without addressing selection bias or segment mix



Blind spots:

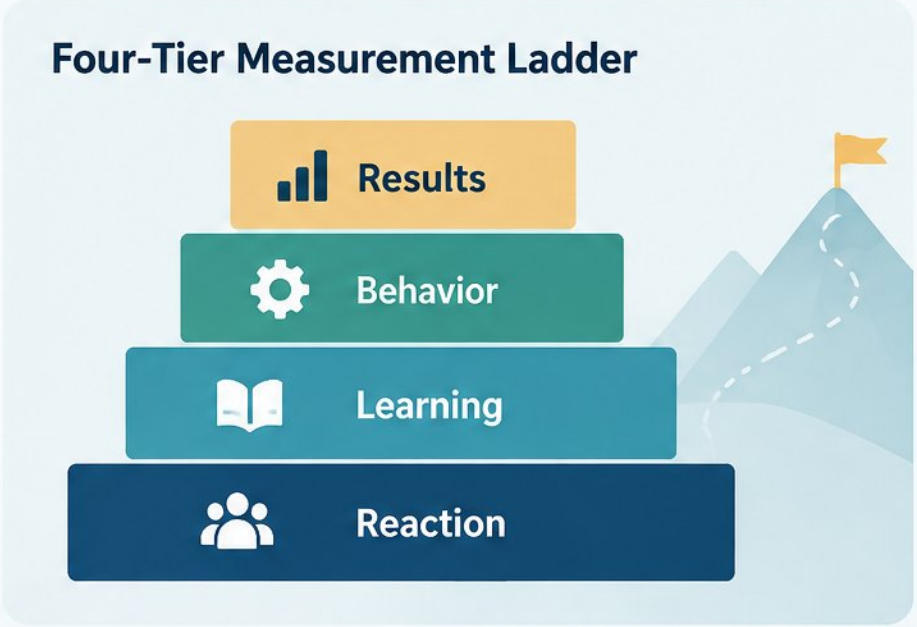
neglecting accessibility, localization, and learner data privacy

Building the Measurement Framework: A Logic Model



Validate each metric: available | affordable | high-quality | timely | benchmarkable

- Logic model: inputs → activities → outputs → short/long-term outcomes → impact
- Build top-down from the business goal you own, not from existing data
- Map quality questions to four tiers; pick at least one metric per tier
- Validate each metric: available, affordable, high-quality, timely, benchmarkable
- Align with CS health scores, product telemetry, and support systems



Core Metrics by Tier: Learning, Behavior, and Business

Learning

- assessment scores,
- certification pass rates,
- time-to-competence,
- scenario performance

Behavior

- feature adoption,
- activation milestones,
- portal engagement,
- self-service deflection,
- correct admin configuration

Business

- retention,
- churn,
- NRR/GRR influence,
- time-to-value,
- ticket reduction,
- support cost

Experience

- CSAT, NPS, and qualitative feedback as leading signals, not verdicts

Leading vs. lagging:
completion predicts renewal, renewal confirms it—use both to forecast

Leading → Lagging

Reaction → Learning → Behavior → Results

Method Choice: Diagnose, Prove, or Improve

Diagnose

Quantitative

Quantitative: pre/post assessments, surveys, A/B and holdout tests, segment comparisons

Qualitative

Qualitative: learner interviews, ride-alongs, ticket theme analysis, community listening

Prove

Trained vs. untrained cohort

Trained vs. untrained cohort: define 'trained,' label records, pick one outcome (e.g., 92% vs. 84% renewal)



Improve

Defensibility

Defensibility: compare within segment/role, consistent eligibility window, training before outcome

Match method to decision

Match method to decision: diagnose drop-off, prove impact, or improve a lesson

Designing Assessments and Certification That Verify Capability

Instrument quality: validity, reliability, and response-bias awareness for surveys and checks

Assessment design: align items to objectives; use scenario items; separate engagement from comprehension

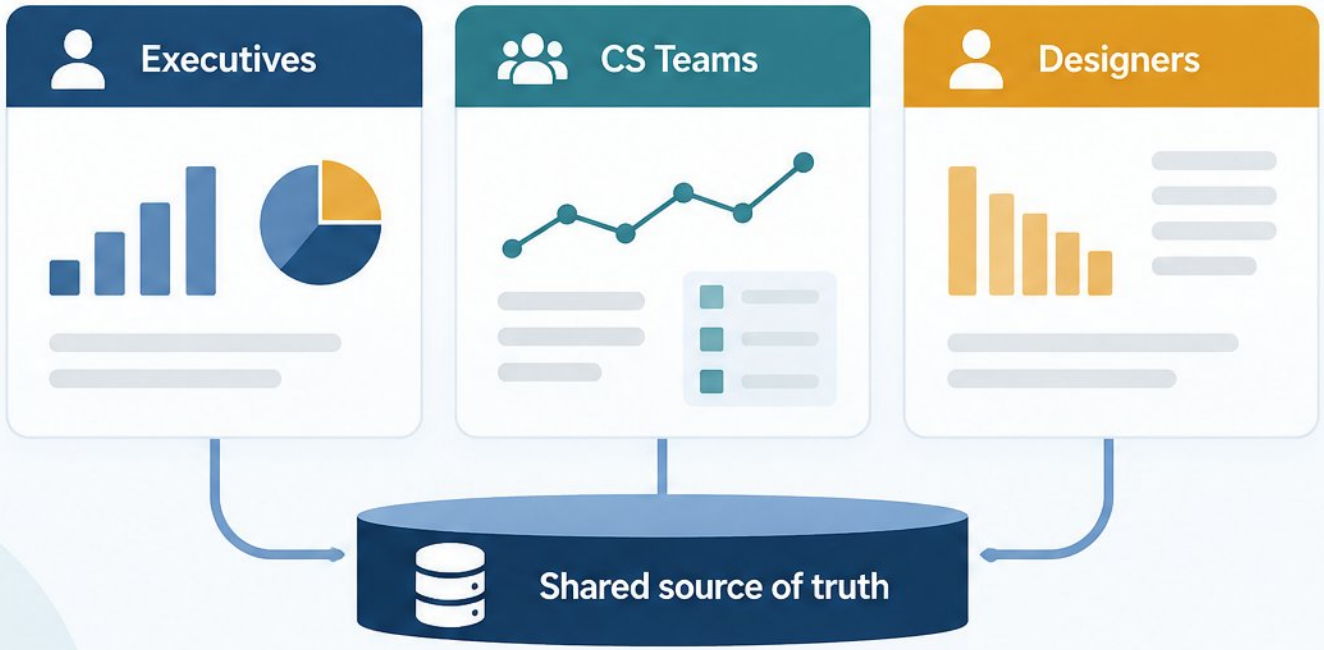
Certification strategy: 51% of high-return programs offer tiered certifications vs. 44% of low-return programs offering none

Benchmarks: certified users show 21–27% higher product usage; certified partners close 18% more deals

Guardrails: prevent answer sharing and refresh items when the product changes



From Data to Decisions: Dashboards and Reporting



Rituals

- Monthly operating reviews
- Quarterly business reviews
- Annual audit

Design views by audience: execs see retention, NRR, support cost; CS sees account health; designers see drop-off

Pair every metric with learner stories and business context

Avoid dashboard sprawl: one metric per decision, one named owner

Governance: written definitions, source-conflict tie-breakers, periodic reviews

Case Examples: Turning Findings into Program Changes

Gainsight PX

- Gainsight PX: completing six or more courses lifted returns 17%; one course drove 87% feature adoption



Mindbody

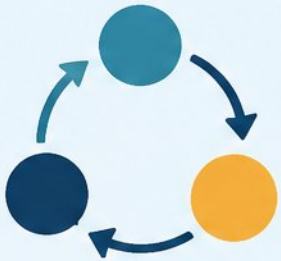
- Mindbody: trained users contacted support less than half as often; lesson-level NPS sped rework



Cin7

- Cin7: usage up 11x, completion 20% → 50%+, enrolled customers retained 7% higher





- Iteration pattern: find friction, pilot a change, measure the delta, codify the fix



- Keep content fresh: 60% of high-return programs refresh continuously vs. 31% of low-return

Closing the Loop: Continuous Improvement Practices

Translate findings into three action types: content fixes, delivery changes, CS enablement plays

Run fast cycles: pilot, measure, refine—feed results into next content refresh



Build feedback loops with CS, product, and support to catch ticket themes early

Keep definitions and metrics stable while improving; otherwise comparisons lose validity

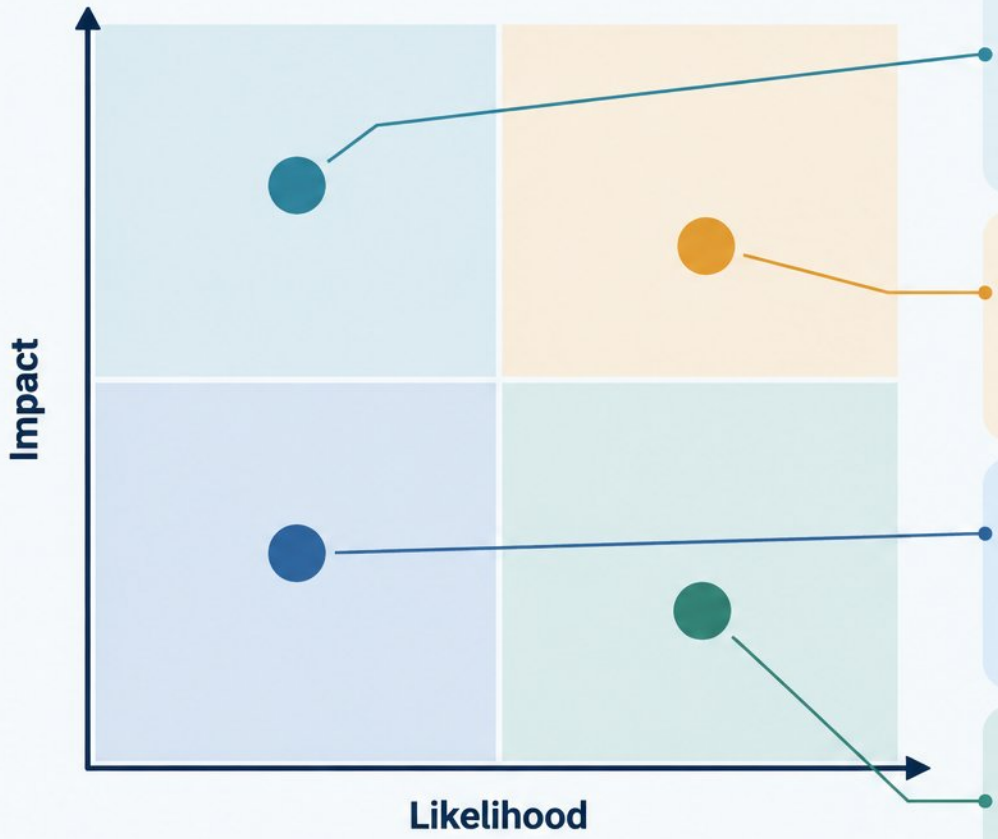
Content fixes

Delivery changes

CS enablement plays

Quality Risks: Attribution, Accessibility, and Data Privacy

Identify key risks, understand their impact, and apply practical safeguards.



Attribution risk: correlation isn't causation—use control groups, matched cohorts, consistent windows

Accessibility and localization: include captions, language coverage, reading level, regional variants

Privacy and consent: learner data is personal data—apply minimality, consent, and residency rules

Data hygiene: watch definitional drift, ownership vacuums, and stale context across systems

Reporting risk: avoid precise ROI without caveats—present ranges and assumptions honestly

90-Day Action Plan and Next Steps

From alignment to impact — a focused 90-day path forward.

Days 1–30

- Days 1–30: define quality criteria for top 2–3 programs; agree 3–4 core metrics with owners

Days 31–60

- Days 31–60: baseline performance, connect LMS and customer data for one program

Days 61–90

- Days 61–90: run one pilot evaluation, share findings in leadership review, launch improvement cycle

Role actions

- Role actions: education leads set framework; CS shares account data; designers instrument assessments; L&D owns method rigor

Success criteria

- Success criteria: one credible business-outcome finding, agreed data definitions, scheduled next review

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/6b5a004e/public