

CROX: A Mixed-Methods Framework for Diagnosing Conversion and Customer Experience Breakdowns in Digital Products

Abstract

CROX grew out of a practical problem in product work: analytics could show where people were dropping out of a journey, while research and customer feedback could help explain why, but the two were not always examined together. This paper formalizes that way of working as CROX, a five-stage mixed-methods framework for diagnosing and prioritizing breakdowns in digital products. We examine CROX retrospectively through a tour-package case from a major online travel platform. The archive includes funnel analysis, 782 navigation and contact records, review of 80 users' purchase behavior, 20 in-depth interviews, destination-service surveys, and Outbound Tours NPS reports. The evidence linked a late-stage funnel drop to problems spanning hotel information, passenger-data entry, comparison, pricing, and destination services. NPS improved from -35 to -21 during the initial 2022 period, then returned to -32 by September. Later monitoring from November 2023 to May 2024 showed a materially different state: NPS was positive in five of seven reporting periods and peaked at approximately +13.4. Because the later series was not a controlled evaluation, we do not attribute the recovery to CROX alone. CROX contributes a repeatable process for integrating evidence, choosing actions, and matching claims to the strength of the evaluation design.

Keywords: **mixed methods; conversion rate optimization; customer experience; HCI; funnel analysis; user experience evaluation**

Research Highlights

- CROX links funnel abandonment to evidence regarding users' experiences.
- CROX provides a distinction among fixes, new concepts, and questions requiring research.
- A tour-package case combined behavioral records, surveys, and 20 interviews.
- Later monitoring showed Outbound Tours NPS move from deeply negative to mostly positive.

1 Introduction

Most digital product teams already have both sides of the evidence. They track conversion, events, exits, and funnel performance, and they also run interviews, surveys, support analysis, and satisfaction measurement. The problem is not usually a lack of data. It is that these sources tend to answer different questions and are often reviewed separately. Behavioral evidence is good at showing where a journey breaks down; experience evidence is better at explaining the confusion, expectation, context, or service problem behind that behavior. HCI methods make the same distinction across behavioral traces, usability measures, interviews, surveys, and contextual inquiry (Lazar et al., 2017; Olson & Kellogg, 2014). Large-scale UX measurement work also shows the value of combining behavioral and attitudinal signals when they are tied to explicit user-centered goals (Rodden et al., 2010).

The harder part is connecting these accounts to a specific step in the interaction or to a user segment that matters commercially. When behavioral and experience evidence are analyzed separately, teams can end up fixing the most visible symptom, reacting to the loudest complaint, or overstating the effect of a change. Human-centered design calls for attention to users' needs and context of use throughout the system lifecycle (International Organization for Standardization, 2019), while customer-experience research treats experience as something that develops across a journey rather than at a single touchpoint (Lemon & Verhoef, 2016). CROX brings these ideas together around one practical unit of analysis: a defined segment moving through a defined journey. We make two contributions.

First, we define a repeatable five-stage framework for preparing evidence, locating meaningful conversion drops, explaining the patterns behind them, classifying possible responses, and evaluating what happens next.

Second, we use a retrospective tour-package case to show both what the framework can reveal and where its limits are. The team combined funnel analysis, interviews, surveys, and Net Promoter Score data to understand the journey. The same case also shows why outcome claims have to remain descriptive when key denominators, exposure allocations, or the final controlled-test report cannot be recovered.

The study addressed two research questions:

- RQ1: How can product teams combine funnel measurements and customer experience evidence to identify priority areas for improvement in a digital journey?
- RQ2: What operational findings and inference limitations result from using CROX in an online tour-package case study?

2 Background and Related Work

2.1 Behavioral Optimization and Controlled Experiments

Conversion rate optimization typically uses overall conversion rate, micro-conversion rates between journey steps, and behavioral signals such as exits or failed interactions to identify where performance changes. UX measurement research likewise distinguishes task-performance, navigation, behavioral, and self-reported measures, emphasizing that metric selection should match the decision being made (Albert & Tullis, 2013; Sauro & Lewis, 2016).

However, consistency in event definition, complete instrumentation and an appropriate comparison group are necessary for these measures to be valid. Online controlled experiments offer a robust method for estimating the impact of an action when assignment to groups, metrics used for evaluation, sample ratio checks and statistical assumptions are trustworthy (Kohavi et al., 2020). A difference between before-and-after evaluations can also indicate seasonal trends, traffic composition, inventory levels, price fluctuations, operational changes or concurrent releases.

CROX therefore treats experimentation as one way to evaluate an intervention, not as a label that can be applied to any observed change in conversion.

2.2 Customer Experience Across a Journey

Experiences occur before, during, and after interactions with systems and may encompass interfaces, delivered services, and expectations formed earlier in a journey. Journey research emphasizes both the temporal and cross-touchpoint nature of experience (Lemon & Verhoef, 2016). UX-evaluation research similarly shows that no single method captures the full construct and highlights the value of multi-method approaches across stages of product development (Vermeeren et al., 2010). A complaint voiced after a purchase may stem from incomplete information on a product page, while a low satisfaction score may stem from a hotel stay, transfer experience, or support interaction that the interface did not directly influence.

This makes journey mapping and careful attribution important. Net Promoter Score can be useful as a compact loyalty measure, but it is not a complete measure of experience and should not be treated as a uniquely reliable predictor of growth. Longitudinal evidence has challenged claims that NPS is superior to other customer metrics (Keiningham et al., 2007). In CROX, NPS is one signal among several, interpreted alongside open-text comments, behavioral evidence, and operational data.

2.3 Mixed Methods and Design-Oriented Methodological Artifacts

Mixed-methods research is useful when quantitative or qualitative evidence alone cannot answer the question well enough (Johnson & Onwuegbuzie, 2004). In an explanatory sequential design, quantitative patterns can guide what researchers look for qualitatively and how they interpret it. The key is integration: placing two datasets side by side is not enough to produce a mixed-methods explanation (Fetters et al., 2013). Thematic analysis provides a structured way to identify recurring patterns in interviews and open-text material (Braun & Clarke, 2006). We frame CROX as a design-oriented methodological artifact: a practical method for making better diagnostic decisions in product work. This framing draws on design-science literature about useful artifacts and their evaluation (Hevner et al., 2004; Peffers et al., 2007), although the original organizational work was not conducted as a formal design-science research study.

For that reason, we evaluate CROX as a practical diagnostic process through one documented organizational case, without treating a single case as evidence of universal effectiveness.

3 Research Approach

3.1 Framework Development

CROX emerged from practice rather than from a formal research program. In the product team, conversion analytics and customer-experience research were closely related but still tended to move through separate workflows. We began combining them when a funnel number alone was not enough to decide what to fix. Over time, that working pattern became the five stages described here, shaped by people working across product design, research, analytics, and experimentation. The framework remains deliberately practical: define the decision, align the evidence around the same segment and journey, choose a response that fits the evidence, and decide how that response will be evaluated before implementation.

3.2 Case Setting and Evidence

The retrospective case examines a tour-package product offered by a major online travel platform in 2022. The packages combined flights and accommodation, and the surviving operational material is especially rich for Turkey and Istanbul. Because the records include commercially sensitive and customer data, we do not identify the organization or disclose raw identifiers. We rely on the archival material that remains available rather than trying to reconstruct missing results from memory.

Evidence source	Scope	Use in the case
Funnel records	Tour journey and product-detail stage onward	Located the broad region of abandonment and instrumentation needs
Navigation and contact records	782 records from 21 March–5 April 2022	Defined candidate users and examined subsequent behavior
Purchase-history review	80 users	Distinguished immediate abandonment from delayed or offline purchase

In-depth interviews	20 participants	Explained recurring information, interaction, and service problems
Istanbul surveys	Seven operational workbooks covering 21 March–21 May 2022	Measured destination touchpoints such as transfers and tour leaders
Outbound Tours NPS	21 Mar–20 Apr 2022 n = 876; 21 Apr–21 May 2022 n = 741; 22 May–21 Jun 2022 n = 746	Tracked product-level recommendation scores and detractor reasons

Table 1 Evidence Used in Retrospective Case Study

3.3 Analysis and Integration

The analysis started with the funnel because we needed to know where to look. A large drop region narrowed the problem, but it was only the starting point.

Subsequently, the team reviewed navigation, contact, and purchase histories to narrow the interview sample and conducted interviews to understand why users stopped progressing through the journey. The interview study was collaborative: two members led interview conduct and qualitative analysis, while the supervisor participated across research design, analysis, and integration. The interviews were conducted by telephone. Participants were informed that the calls were for the research and chose to take part. Calls were not recorded; researchers took notes during the interviews, and those notes were coded and grouped into recurring themes. Operational survey and NPS reports were used as contextual outcome indicators. We integrated the evidence by mapping each qualitative or service-related theme back to the journey stage where it could plausibly affect behavior, including upstream problems that only became visible at a later conversion drop.

3.4 Confidentiality, Ethics, and Research Status

The original work was conducted as routine product and service research rather than as prospective academic research. No independent institutional ethics review was sought for the original operational work. Interview participants were informed that the calls were for the research and chose to take part. The calls were not recorded; researchers took notes during the interviews.

This retrospective manuscript introduces no new participant contact and reports only aggregate, de-identified archival information; it does not include user quotations, contact details, order identifiers, or proprietary dashboard graphics. Raw operational files contain personally identifiable and commercially sensitive information and are therefore not publicly accessible.

4 The CROX Framework

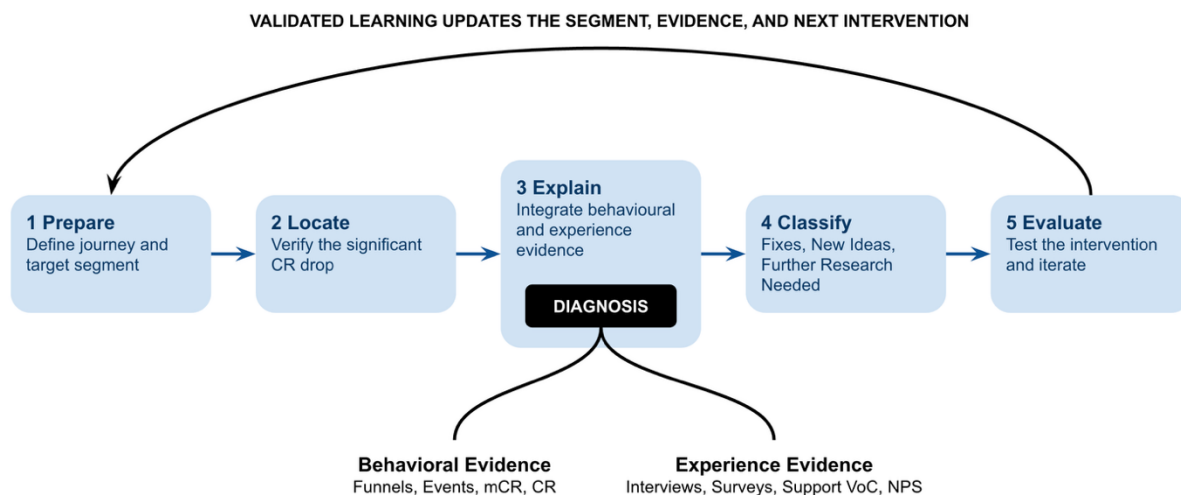


Figure 1 Five-Stage CROX Diagnostic and Evaluation Cycle

4.1 Prepare Evidence and Prioritize a Segment

The first stage begins before analysis: decide what problem is being investigated before opening another dashboard. The team defines the product boundary, journey, population, time frame, and desired outcome, then brings together evidence that belongs to the same scope. Behavioral evidence may include overall conversion, micro-conversion, event completion, error events, exits, latency, or session recordings. Experience evidence may include interviews, surveys, support reasons, service-quality measures, or NPS comments. Before drawing conclusions, the team checks event definitions, missingness, duplicate identities, time windows, and whether the qualitative evidence actually refers to the same population and journey. Segmentation is useful when averages hide important differences or resources force the team to prioritize. Size, value, strategic

importance, behavior, destination, device, lifecycle stage, or RFM can all be reasonable criteria. What matters is defining the segment before the results are known.

4.2 Locate a Significant Drop

For two consecutive funnel steps, i and $i+1$, the micro-conversion rate is the share of users who move from one step to the next. A steep visual drop is not enough for CROX to treat the step as significant. The drop should cross a predefined operational or statistical threshold, and the instrumentation behind it should be credible. In practice, teams compare the step with its historical variation, similar segments, expected error rates, and the business importance of what follows. If the tracking cannot be trusted, the first action is not redesign. It is fixing the measurement.

4.3 Explain Observed Pattern

Finding where people stop is only half of the diagnosis. The next question is whether the reason is actually located at that step. Session recordings may reveal a broken component or a repeated failed action. Interviews and open-text responses may point to missing information, mistrust, confusion, workflow problems, or unmet service expectations. Support contacts and service surveys can show that what looks like an interface issue is partly operational. The goal is not to force everything into one explanation. Themes are coded and mapped back to the journey, and contradictory evidence stays visible when it matters.

4.4 Classify Candidate Action

CROX assigns evidence-based opportunities for improvement to three decision categories:

- Fixes and Improvements address verified defects, misleading text, missing content, accessibility problems, or broken workflows.
- New Ideas change an interaction or service and therefore need an explicit hypothesis: what should improve, why it should improve, and how the team will test it.
- Further Research is the right choice when the cause is still unclear, the evidence conflicts, or acting too quickly could create significant risk.

This classification is meant to slow down a familiar product reflex: turning every complaint into a feature request. A candidate action should make sense for the affected segment, have a plausible

mechanism, and be worth its reach, cost, operational dependencies, and possible downside. Success metrics and guardrails are set before implementation, not after the team has seen the result.

4.5 Evaluate and Iterate

The final stage is about matching the evaluation to the claim. If a team wants to say that an interface change caused an incremental effect, a controlled experiment may be appropriate when random assignment and enough exposure are possible. Operational changes may need a phased rollout, interrupted time series, or a carefully bounded before-after analysis, with correspondingly weaker causal claims. Whatever the design, the report should make the baseline, comparison, unit, time window, sample size, missingness, and type of change clear. Major changes should not be bundled together when attribution matters. CROX then loops: an evaluation may confirm the diagnosis, reveal a segment-specific effect, or send the team back to research.

5 Retrospective Tour Package Case Study

5.1 Funnel Problem and Research Design

CROX review sessions pointed to substantial abandonment from the product-detail page onward. The product combined flights and accommodation, and the surviving records are concentrated on outbound tours and destination services in Istanbul. The selected destinations represented a large share of activity, but the archive does not contain a reliable denominator that would support an exact proportion, so we do not claim one. We analyzed 782 navigation and contact records collected between 21 March and 5 April 2022. The team then reviewed the purchase behavior of 80 users, contacted candidates, and completed 20 in-depth interviews. A separate behavioral analysis from the time reported that roughly 32% of users who abandoned the funnel returned within fifteen days and purchased. The dataset behind that estimate is no longer available, so we treat the figure as an archived operational finding rather than a statistic derived from the 80-user review. Even with that limitation, the pattern is useful: a same-session drop does not necessarily mean the user is lost, and the outcome window can materially change the story the funnel tells.

5.2 Explanatory Themes

The interviews did not point to one clean reason for abandonment. Hotel pages often left users without enough detail about room types, images, or reviews. Passenger-data entry raised practical questions about identification requirements, children, and required fields, especially on mobile. Comparison was also fragile: users could lose filter selections while scrolling and had difficulty comparing packages, flights, rooms, and dates. Inventory and pricing added another layer of friction when choices were limited, prices looked stale, or the final price differed from the one first selected. A late-stage funnel drop was therefore carrying the effect of several different problems at once.

5.3 Destination Service Evidence

Operational surveys measured Istanbul touchpoints including outbound and return transfers, tour leaders, city tours, and optional tours. Results varied across reporting periods.

For example, the proportion of respondents rating outbound transfers as excellent ranged from 43.1% to 78.1% in recovered reports, while dissatisfaction ranged from 8.18% to 30.0%.

Dissatisfaction with city tours reached 28.57% during one reporting period.

Although these measures were not part of controlled experiments, they helped distinguish destination-service problems from interface-related problems and supported continued monitoring of the end-to-end journey.

5.4 Outbound Tours NPS

Reporting period (2022)	Respondents	Promoters	Passives	Detractors	NPS
21 Mar–20 Apr	876	(212) 24%	(147) 17%	(517) 59%	–35
21 Apr–21 May	741	(199) 27%	(154) 21%	(388) 52%	–25
22 May–21 Jun	746	(222) 30%	(147) 19%	(377) 51%	–21

Table 2 Monthly Summary Reports of Outbound Tours Net Promoter Scores

Outbound Tours NPS moved from -35 in the 21 March–20 April reporting period to -21 in the 22 May–21 June period, a 14-percentage-point increase. Over the same period, the share of detractors who attributed dissatisfaction to tour services fell from 40% to 37%, although tour services were still the largest dissatisfaction category in the latter period. Within that subset, recurring complaints concerned tour-leader or destination-operator behavior and the quality of public airport transfers. Later reports recorded NPS scores of -30 for 22 June–22 July, -25 for 23 July–22 August, and -32 for 23 August–22 September 2022. In other words, the spring improvement did not hold over the next three months.

The observed NPS change is a 14-percentage-point descriptive increase. The surviving archive does not support attribution of a specific conversion-rate effect to an A/B test. Although archival documents discuss experimentation and product enhancements, they do not contain the information required to reconstruct a controlled-test estimate, including variant allocation, exposure counts, variant definitions, uncertainty estimates, or final test reports.

Accordingly, the reported outcomes are treated as observational product-level changes rather than causal estimates of an intervention effect.

5.5 Continuous Measurement

Subsequent work in July separated overall Istanbul NPS from a diagnostic score which excluded factors perceived as irrelevant to destination services.

Overall Istanbul NPS remained strongly negative across the reported months, while the restricted diagnostic NPS hovered around zero. The team rejected the restricted score as a replacement for end-to-end NPS because problems with hotels and destination services were still part of the customer's experience. In September, the teams met again to consolidate survey data, revise NPS questions, assign measurement ownership, and define the next CROX agenda. Later operational reports document provider changes in Istanbul following widespread dissatisfaction and show that weekly monitoring continued. One period showed roughly a seven-percentage-point reduction in dissatisfaction, but recurring concerns remained around transfer waiting times, tour-leader conduct, optional-tour pricing, and city-tour quality. These records show an ongoing measurement and service-adjustment process, but they do not provide a controlled estimate of the effect of changing providers.

5.6 Later Operational Evidence (2023–2024)

A separate internal NPS analysis conducted from November 2023 through May 2024 provides a longer-term view of the same product area. It pooled 7,357 tour responses: 4,496 from domestic-tour customers and 2,861 from outbound-tour customers, with 4,824 written comments across the full sample. Within the outbound-tour group, 2,371 respondents left comments, including 1,162 detractor comments; roughly 712 of those detractor comments were associated with Istanbul, Dubai, and Antalya.

The most consequential result is the change in Outbound Tours NPS. The 2022 series remained in negative double digits and fell back to -32 by September. In the November 2023–May 2024 series, NPS was positive in five of seven reporting periods, with approximate values of -6.3, +11.1, +13.4, +11.5, +7.5, -10.8, and +5.7. The difference between the lowest 2022 value (-35) and the later peak (+13.4) is 48.4 NPS points. The series was not monotonic, but the overall shift from deeply negative scores to mostly positive territory is operationally substantial.

The later analysis also coded dissatisfaction into recurring categories including hotel, support, provider, and tour-service issues. In both Istanbul and Dubai, tour services were the largest coded complaint category in the report, while hotel, provider, and support issues also remained visible. This later evidence therefore shows both improvement and continuity: the customer-experience state changed materially, while several of the issue types identified in the original work continued to be monitored. The report was not designed as a controlled evaluation of CROX, so it cannot isolate how much of the recovery was caused by the framework versus subsequent product, supplier, pricing, service, or market changes. That limitation affects causal attribution, not the descriptive magnitude of the observed recovery.

6 Discussion

6.1 Contribution

The case is a useful reminder that a funnel drop is a location, not an explanation. Here, the product-detail and passenger-data stages were where the loss became visible, but the evidence pointed to several causes: missing information, documentation questions, weak comparison, inventory and pricing problems, and destination services. CROX makes that distinction operational. Quantitative evidence narrows the search and helps shape qualitative sampling;

qualitative findings are brought back to the same journey; and the evaluation design limits what can reasonably be claimed afterward. That is the practical contribution. It is also consistent with the mixed-methods principle that integration should happen during design and interpretation, not only when the results are written up (Fetters et al., 2013).

6.2 Implications for Practice

Four lessons from the case are especially useful in practice.

First, teams should define the unit of analysis before collecting additional evidence. Product-level NPS and destination-service surveys may both be valid indicators, but they measure different aspects of experience and answer different questions.

Second, delayed purchases should be considered when a funnel defines success. An archived behavioral analysis reported that approximately 32% of abandoners returned within fifteen days and purchased, suggesting that same-session definitions may be inadequate for some decisions. Because the underlying dataset is no longer available, this figure should be interpreted as an operational finding rather than a recomputed statistic from the retrospective sample.

Third, results must be expressed in the units generated by the underlying measures. A 14-point increase in NPS is not a relative conversion-rate increase, and an observational time trend is not an A/B-test estimate. Teams implementing CROX should maintain a concise evidence register containing segment and funnel definitions, event dictionaries, qualitative samples, theme codebooks, action decisions, and evaluation designs. This makes later review possible when personnel, tools, or dashboards change and creates a practical barrier against post-hoc selection of favorable segments or outcomes.

Fourth, long-term trajectories can reveal change that short windows miss. In 2022, Outbound Tours NPS remained deeply negative even after the initial 14-point improvement. By 2023–2024, five of seven reported values were positive and the series reached approximately +13.4. The observed spread between -35 and +13.4 is 48.4 NPS points. That is not a causal estimate of CROX, but it is an operationally large shift and should not be obscured by the absence of a controlled long-term design. For teams using CROX, the implication is to keep measurement alive after an intervention rather than closing the loop after one experiment or one improved month.

6.3 Limitations

The study has five important limitations.

First, this is one retrospective case in online travel, so we cannot assume the framework will transfer unchanged to other domains.

Second, parts of the archive are missing. Full funnel denominators, complete controlled-experiment reports, and the dataset behind the historical 32% delayed-purchase finding are no longer available, which prevents independent recomputation and causal estimates of conversion effects.

Third, the qualitative sample was selected for an operational purpose, and the surviving documentation is not detailed enough to assess saturation, intercoder agreement, or non-response bias.

Fourth, NPS and destination surveys are self-reported measures. They are vulnerable to response, selection, and timing biases, and the monthly samples and parts of the journey being rated were not constant. A stronger next test of CROX would therefore be prospective, with predefined thresholds, explicit evidence-integration displays, and controlled comparisons where feasible.

Fifth, the 2023–2024 NPS report was produced as routine operational analysis rather than as a prospective replication or controlled evaluation of CROX. It cannot isolate how much of the later recovery is attributable to CROX versus subsequent product, supplier, pricing, service, or market changes. This limits causal attribution, but it does not reduce the descriptive magnitude of the observed shift from deeply negative NPS in 2022 to mostly positive values in the later series.

7 Conclusion

CROX is deliberately simple. Put the evidence around the same segment and journey, find the drop that actually matters, investigate why it is happening, decide whether the answer is a fix, a new idea, or more research, and evaluate the result in a way that supports the claim being made. The 2022 tour-package case shows the value of that approach: a visible funnel drop turned out to involve interface, information, inventory, and service issues at the same time. Later operational monitoring strengthens the practical story. Outbound Tours NPS moved from persistent negative double digits in 2022 to mostly positive values in the 2023–2024 series, peaking at approximately +13.4. We cannot attribute that recovery to CROX alone, but the scale of the change shows that the product's customer-experience state changed substantially over time while systematic measurement and diagnosis continued. The next useful test of CROX is prospective: apply it across different products, compare the decisions it produces with analytics-only and research-only workflows, and test whether integrating evidence in this way leads to better prioritization or better intervention outcomes.

Data Availability

Raw data cannot be disclosed publicly because the source files contain customer identifiers, open-text responses, and commercially sensitive operational information. De-identified aggregate data may be provided to editors for confidential review when organizational and privacy policies permit.

References

- Albert, W., & Tullis, T. (2013). *Measuring the user experience: Collecting, analyzing, and presenting usability metrics* (2nd ed.). Morgan Kaufmann.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qpo630a>
- Fetter, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs: Principles and practices. *Health Services Research*, 48(6 Pt 2), 2134–2156. <https://doi.org/10.1111/1475-6773.12117>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- International Organization for Standardization. (2019). *Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems* (ISO Standard No. 9241-210:2019).
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26. <https://doi.org/10.3102/0013189X033007014>
- Keiningham, T. L., Cooil, B., Andreassen, T. W., & Aksoy, L. (2007). A longitudinal examination of net promoter and firm revenue growth. *Journal of Marketing*, 71(3), 39–51. <https://doi.org/10.1509/jmkg.71.3.039>
- Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy online controlled experiments: A practical guide to A/B testing*. Cambridge University Press. <https://doi.org/10.1017/9781108653985>
- Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction* (2nd ed.). Morgan Kaufmann.
- Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, 80(6), 69–96. <https://doi.org/10.1509/jm.15.0420>
- Olson, J. S., & Kellogg, W. A. (Eds.). (2014). *Ways of knowing in HCI*. Springer. <https://doi.org/10.1007/978-1-4939-0378-8>
- Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>

- Rodden, K., Hutchinson, H., & Fu, X. (2010). Measuring the user experience on a large scale: User-centered metrics for web applications. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 2395–2398). ACM. <https://doi.org/10.1145/1753326.1753687>
- Sauro, J., & Lewis, J. R. (2016). Quantifying the user experience: Practical statistics for user research (2nd ed.). Morgan Kaufmann.
- Vermeeren, A. P. O. S., Law, E. L.-C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: Current state and development needs. In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (pp. 521–530). ACM. <https://doi.org/10.1145/1868914.1868973>

Appendix CROX Operational Checklist

Stage	Required record	Minimum decision
Prepare	Product boundary, segment, period, event dictionary, data-quality checks	The evidence refers to the same population and journey
Locate	CR and mCR by step, baseline variation, error events, threshold	The drop is material and instrumentation is credible
Explain	Interview or survey sample, support/service evidence, themes mapped to journey	The likely cause and competing explanations are documented
Classify	Fix, new idea, or further research; mechanism; risk; owner	The proposed action matches evidence strength
Evaluate	Baseline, unit, exposure, comparison, guardrails, uncertainty	The wording of the result matches the evaluation design

Table A1 Minimum Documentation for CROX Cycle