STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
*Stat., Optim. Inf. Comput.*, Vol. **16**, October 2026, pp 3776–3791.
Published online in International Academic Press (www.IAPress.org)

# Onboarding Friction in SaaS Platforms: An Exploratory Qualitative-Computational Analysis

Rini Mayasari [1,*], Nono Heryana [2]

[1]*Department of Informatics, Universitas Singaperbangsa Karawang*
[2]*Department of Information Systems, Universitas Singaperbangsa Karawang*

**Abstract** Software-as-a-Service (SaaS) platforms face a persistent challenge in converting newly registered users into engaged adopters, but behavioural metrics alone provide limited evidence about the mechanisms of first-use friction. This study evaluates a transparent analysis workflow on 50 short, synthetically generated SaaS-onboarding interview records from an open educational repository. It therefore constitutes a methodological and theory-generating demonstration, not a study of observed users. Two coders applied the same eight-theme operational codebook, grounded in Cognitive Load Theory, the Technology Acceptance Model, Expectation Disconfirmation Theory, Goal-Setting Theory, and Self-Efficacy Theory. The resulting candidate typology comprises Information Overload (T1), Navigation Opacity (T2), Onboarding Discontinuity (T3), Personalization Deficit (T4), Goal-Means Misalignment (T5), Self-Efficacy Impairment (T6), Expectation Disconfirmation (T7), and Time-to-Value Friction (T8). Codebook coverage was reached by the seventh ordered record, while the remaining records supported prevalence estimation and exploratory subgroup comparisons. Inter-rater reliability was substantial on average (mean Cohen's $\kappa = 0.753$). Large but imprecisely estimated associations appeared for Onboarding Discontinuity in Beginner-labelled records ($\mathrm{OR} = 63.00$, 95% CI $[6.52, 608.95]$, $p < .001$) and Expectation Disconfirmation in Advanced-labelled records ($\mathrm{OR} = 9.67$, 95% CI $[2.14, 43.56]$, $p = .004$). Because the corpus is synthetic, these patterns are hypotheses for validation with authentic users rather than population estimates. The contribution is a reproducible, deliberately modest workflow that links interpretive coding, transparent rule-based text scoring, and uncertainty-aware exploratory analysis.

**Keywords** SaaS onboarding; user experience friction; qualitative content analysis; thematic coding; cognitive load; expectation disconfirmation; human-computer interaction; usability

**AMS 2010 subject classifications** 68U35, 91E10, 62P25

**DOI:** 10.19139/soic-2310-5070-3738

## 1. Introduction

The adoption of Software-as-a-Service (SaaS) platforms has accelerated substantially over the past decade, with enterprise software markets increasingly structured around subscription-based models in which user retention is a primary commercial imperative [1, 2]. The transition from on-premise to cloud-delivered software has redistributed the competitive emphasis from deployment and licensing to ongoing user engagement, placing onboarding at the strategic centre of product growth [3]. Within this context, onboarding is broadly defined as the structured process through which a new user acquires the knowledge, confidence, and workflow integration necessary to derive value from a platform. Failures in onboarding design are consistently associated with elevated churn rates, reduced feature adoption, and diminished long-term customer lifetime value [4, 5].

*Correspondence to: Corresponding Author (Email: rini.mayasari@staff.unsika.ac.id). Department of Informatics, Universitas Singaperbangsa Karawang, Karawang, Indonesia

ISSN 2310-5070 (online) ISSN 2311-004X (print)

Despite its commercial significance, the academic literature on SaaS user experience (UX) has tended to focus on system usability heuristics, interface design principles, or aggregate behavioural metrics derived from behavioural interaction data. The phenomenological and experiential dimensions of onboarding friction have received comparatively limited systematic attention. Recent scholarship identifies this as a meaningful gap: digital service designs that rely on quantitative interaction data alone cannot identify the cognitive and affective mechanisms that produce disengagement [6, 7]. By onboarding friction, this study refers to the subjective, language-mediated processes through which users construct their initial understanding of a platform and evaluate its adequacy relative to prior expectations. Without a grounded understanding of how and why users experience friction during onboarding, design interventions remain necessarily heuristic rather than empirically derived.

The present study addresses this gap through an exploratory secondary analysis of 50 synthetic interview-style records representing a SaaS onboarding context. The repository explicitly describes all of its records as synthetically generated and intended for education and method practice [33]; consequently, the study does not claim to recover lived phenomenology or estimate the prevalence of friction among real SaaS users. The analysis is organised around two bounded research questions. First, what candidate categories of onboarding friction can be recovered from this synthetic corpus using a transparent thematic codebook? Second, what experience-level and phase-level hypotheses does the workflow generate for subsequent validation with authentic users?

The contribution is integrative rather than a claim that each friction theme is unprecedented. Theoretically, the paper maps familiar usability problems to a SaaS-specific process model spanning cognitive load, self-efficacy, expectation disconfirmation, goal alignment, and time-to-value. Methodologically, it documents how human coding can be linked to auditable rule-based quantification without describing a small custom dictionary as machine learning. Practically, it translates the candidate themes into concrete design propositions that can be tested in real onboarding settings. Recent SOIC research likewise shows that technology-adoption outcomes depend on more than usage intensity and that performance and effort expectations remain central to digital adoption [47, 48]; the present paper narrows those concerns to the first-use interface.

The remainder of the paper reviews the integrated theoretical and UX-evaluation context, describes the synthetic corpus and analysis workflow, reports exploratory results with uncertainty, and concludes with validation priorities and design propositions.

## 2. Literature Review

### *2.1. Onboarding as a User Experience Construct*

Within the human-computer interaction literature, onboarding is most commonly conceptualised as a subset of the broader first-use experience, encompassing tutorials, guided tours, tooltips, and empty-state prompts that platform designers deploy to reduce initial unfamiliarity [11, 12]. Foundational work by Nielsen [13] established that user interface learnability is a primary dimension of system usability, distinct from operational efficiency and error recovery. Learnability is defined as the speed and completeness with which a new user attains proficiency. More recent scholarship has reframed learnability as a dynamic, socially situated process in which users draw on prior mental models, role expectations, and task contexts to interpret new systems [5, 14].

The growth of cloud-based platforms has intensified theoretical interest in onboarding as a product strategy rather than simply a UX design concern [15]. Empirical evidence from systematic reviews of digital service adoption indicates that onboarding quality predicts early retention more reliably than pricing or feature richness, suggesting that first-use friction produces disengagement before users have sufficient information to evaluate product-market fit [7]. A theoretically important distinction in this literature concerns the difference between feature-oriented onboarding designs, which guide users through platform capabilities in a fixed sequence, and goal-oriented approaches, which organise onboarding content around the user's intended outcomes. Goal-setting theory [17] predicts that goal-oriented onboarding produces higher task completion rates and subjective satisfaction, as users who understand the purpose behind each action are better equipped to exercise deliberate self-regulation [16]. The present study operationalises this distinction as the theme of Goal-Means Misalignment (T5), grounded

in corpus responses that explicitly distinguished knowing how to perform an action from understanding why or when to do so.

### *2.2. Theoretical Foundations*

Cognitive Load Theory (CLT; [8]) provides the foundational framework for understanding information overload in onboarding. CLT distinguishes between intrinsic load, arising from the inherent complexity of the subject matter; extraneous load, imposed by suboptimal instructional design; and germane load, referring to cognitive effort directed toward schema formation. SaaS onboarding designs that present excessive features simultaneously, deploy dense tooltip sequences, or require parallel processing of interface navigation and task execution impose extraneous load that directly degrades initial learning outcomes [18]. Skulmowski and Xu [18] confirm that interface complexity amplifies cognitive overload particularly in digital and online learning contexts, a finding that transfers directly to SaaS environments in which new users must simultaneously navigate an unfamiliar interface and absorb functional knowledge.

Expectation Disconfirmation Theory (EDT; [9]), originally developed to explain post-purchase satisfaction, has been successfully applied to digital service contexts to predict user continuance intention [20]. EDT posits that satisfaction is a function of the discrepancy between pre-consumption expectations and perceived performance. In SaaS onboarding, marketing communications establish performance expectations that the onboarding experience must meet or exceed to generate positive affect and continuance motivation. Negative disconfirmation, when the experience falls short of expectation, produces dissatisfaction and elevates churn risk [21]. A meta-analytic synthesis demonstrates that the expectation-performance gap operates across both public service and digital platform contexts, making disconfirmation a theoretically robust and generalisable mechanism [21].

Bandura's Self-Efficacy Theory [10] contributes a critical psychological dimension, explaining the conditions under which users develop or fail to develop confidence in their capacity to operate a platform competently. Low self-efficacy during onboarding manifests as anxiety about setup correctness, reluctance to explore, and avoidance of features perceived as complex. Experimental evidence confirms that direct enactive engagement with a technology – as opposed to observational exposure – significantly improves self-efficacy and generates more positive attitudes toward continued use, underscoring the importance of hands-on onboarding design over passive demonstrations [23]. The Technology Acceptance Model (TAM; [24]) further contextualises time-to-value considerations through its perceived usefulness and perceived ease-of-use constructs, both of which are substantially shaped by the quality of the initial onboarding experience. Meta-analytic evaluations of TAM extensions confirm that ease-of-use perceptions formed during onboarding mediate the relationship between interface design quality and continuance intention [25].

### *2.3. Positioning Relative to Existing UX Evaluation Frameworks*

The proposed themes should be read as a process-oriented translation of established usability knowledge, not as eight wholly novel phenomena. Table 1 makes that relationship explicit. Nielsen's heuristics and ISO 9241-11 [51] primarily evaluate interface properties and use outcomes, whereas the present typology organises friction by the cognitive or evaluative mechanism and the point at which it enters an onboarding journey. Its incremental theoretical value is therefore the linkage among interface condition, user mechanism, and temporal stage.

### *2.4. Research Gaps*

Despite the theoretical richness of these frameworks, their integrated application to the empirical analysis of SaaS onboarding friction remains limited. Existing quantitative studies of SaaS UX tend to rely on aggregated behavioural interaction data such as click sequences, session durations, and feature activation rates, which capture what users do but cannot explain why they experience friction [26]. This reliance on quantitative interaction data produces accurate maps of where users abandon onboarding flows but does not identify the cognitive or affective constructs that drive abandonment. Qualitative studies exist but frequently lack systematic coding protocols, saturation evidence, or inter-rater reliability documentation, limiting their replicability and generalisability [27, 28]. Recent methodological literature calls for mixed-method designs in UX research that

Table 1. Positioning of the Candidate Themes Relative to Established UX Frameworks

| Candidate theme | Closest Nielsen heuristic | ISO 9241-11 outcome | SaaS-onboarding extension |
|---|---|---|---|
| T1 Information Overload | Aesthetic/minimalist design | Efficiency | Extraneous load during first-use learning |
| T2 Navigation Opacity | Match with real world; recognition over recall | Effectiveness and efficiency | Feature-path discovery during task execution |
| T3 Onboarding Discontinuity | User control; help/documentation | Effectiveness | Persistence and recoverability of guidance |
| T4 Personalization Deficit | Flexibility and efficiency of use | Satisfaction and efficiency | Role- and expertise-contingent sequencing |
| T5 Goal-Means Misalignment | Match between system and real world | Effectiveness | Connection between features and intended outcomes |
| T6 Self-Efficacy Impairment | Error prevention; visibility of status | Satisfaction | Confidence in setup correctness |
| T7 Expectation Disconfirmation | Consistency and standards | Satisfaction | Gap between pre-entry promise and first-use performance |
| T8 Time-to-Value Friction | Flexibility and efficiency of use | Efficiency | Delay before the user can complete valued work |

combine interpretive richness with quantifiable coding schemes capable of statistical analysis [29]. The present study addresses the methodological part of this gap by applying a documented dual-coder thematic protocol with rule-based quantification. Because its source corpus is synthetic, it is positioned as a reproducible proof of method and a source of propositions, not as a substitute for field-based qualitative UX research.

## 3. Data and Methodology

### *3.1. Research Design*

This study adopts an exploratory mixed-method design combining human thematic coding with deterministic computational summaries and non-parametric statistical tests. Here, "computational" means that scripts tabulated binary codes, cumulative codebook coverage, co-occurrence counts, confidence intervals, and exact or rank-based tests; it does not denote machine learning or automated theme discovery. The qualitative strand follows the systematic approach outlined by Braun and Clarke [30], with text-as-data principles used only for transparent aggregation [31]. Mixed-method research designs have gained increasing acceptance in information systems and HCI scholarship as a means of bridging interpretive depth with statistical tractability, particularly in contexts where the phenomena of interest are experiential and resist operationalisation through standard survey instruments [32]. The quantitative strand applies Fisher's exact tests, Kruskal-Wallis $H$-tests, Mann-Whitney $U$ tests with Bonferroni correction, and Spearman rank correlations to operationalised thematic codes and descriptive indices derived from the coded corpus.

### *3.2. Dataset*

The corpus is the complete 50-row file `datasets/user_interviews/saas_onboarding_interviews.csv`, accessed August 8, 2026, from the Qualitative Text Datasets for UX Research repository [33]. No rows were sampled out: inclusion required a non-empty identifier, phase, question context, response, age-range label, role label, and experience-level label; all 50 rows met these criteria. The repository describes every dataset as synthetically generated, realism-optimised, and intended for educational and research method practice. It does not disclose the generator, model, source materials, prompt, decoding settings, or human post-editing protocol. Thus, labels such as age, role, and experience are scenario metadata, not verified participant characteristics. Five question contexts are equally represented: Initial Expectations, First Login, Onboarding Flow,

Feature Discovery, and Overall Experience (each $n = 10$). Experience labels are Beginner ($n = 14$), Intermediate ($n = 18$), and Advanced ($n = 18$).

Response length ranged from 22 to 50 words ($M = 32.4$, $SD = 6.0$), and no fields were missing. These short records resemble focused interview answers but lack the probing, contextual development, and unplanned ambiguity of authentic interviews. The balanced contexts and metadata variation are properties of corpus construction, not evidence of purposive recruitment. Accordingly, sample-size guidance for real interviews [35] is used only as background and not as validation of this corpus.

### *3.3. Thematic Coding Protocol*

Thematic coding was conducted through a two-phase process. In the first phase, all 50 transcripts were read in their entirety before any coding commenced, following which a preliminary codebook was developed inductively from the data. Codes were assigned only when the underlying construct was explicitly expressed in the record text, not inferred from contextual implication. This conservative coding rule was adopted to maximise intercoder agreement and minimise subjective interpretation. The resulting codebook, presented in Table 3, comprises eight thematic codes, each grounded in a theoretical framework from the UX and information systems literature.

In the second phase, an independent coder applied the same eight operational definitions and inclusion/exclusion rules to all 50 records. Nielsen's heuristics [13] were used during coder training to discuss boundary cases, but did not constitute a separate codebook. Both coders therefore made comparable binary decisions for T1–T8, as required for an interpretable Cohen's $\kappa$. Disagreements were not replaced by simulated noise. Theme-specific $\kappa$ values are reported alongside raw agreement because prevalence-sensitive agreement measures can be unstable for rare codes [36].

### *3.4. Deterministic Computational Summaries*

The revised analysis removes the custom sentiment score because the original project did not preserve a sufficiently auditable lexicon and the non-significant sentiment comparisons did not advance either research question. Computational summaries are now limited to fully defined operations on the binary coding matrix: theme prevalence, Clopper–Pearson intervals, cumulative first occurrence of each code, pairwise co-occurrence counts, and the inferential procedures below. This narrower use is reproducible from the codebook and coded matrix and avoids implying an advanced NLP model. Recent SOIC work illustrates the substantially different validation burden associated with machine- learning sentiment analysis and text classification [49, 50].

### *3.5. Statistical Analysis*

All inferential tests were selected for their appropriateness to the distributional properties and sample sizes of the data. Fisher's exact test was applied to theme prevalence cross-tabulations, as multiple expected cell counts fell below five, violating the chi-square minimum cell assumption. Phase-level Pearson chi-square tests are retained as descriptive screening statistics; sparse expected cells and the constructed phase–experience association limit their inferential interpretation. Non-parametric Kruskal-Wallis $H$-tests were used for three-group comparisons, with effect sizes reported as:

$$\eta^2 = \frac{H - k + 1}{N - k}, \tag{1}$$

where $k$ is the number of groups and $N$ is the total sample size. Pairwise Mann-Whitney $U$ tests applied Bonferroni-corrected significance thresholds ($\alpha/3 = 0.017$), with rank-biserial correlation $r$ reported as a standardised effect size. Spearman $\rho$ was computed for all bivariate correlations among quantitative constructs. Statistical significance was evaluated at $\alpha = .05$ (two-tailed) unless otherwise noted. Because eight theme-wise tests were conducted within each family, Bonferroni-adjusted $\alpha = .00625$ is also used as a conservative multiplicity benchmark. A sensitivity analysis for the most unbalanced experience contrast ($n_1 = 14$, $n_2 = 36$) indicated that 80% power at $\alpha = .05$ requires approximately Cohen's $h = 0.88$ (and $h = 1.13$ at $\alpha = .00625$). Thus, the design detects only very large differences; null results do not rule out small or moderate associations. Odds ratios are accompanied by approximate 95% log-Wald confidence intervals.

## 4. Exploratory Corpus Results

### *4.1. Synthetic Scenario Profile*

Table 2 presents the synthetic scenario profile. The metadata are predominantly assigned to mid-career professional profiles (25–34 years, 44.0%; 35–44 years, 38.0%), with a balanced distribution across experience levels (Beginner 28.0%; Intermediate and Advanced 36.0% each). Interview phases were structured to overweight task-proximal responses: Task Walkthrough and Reflection each account for 40.0% of the sample, with Warm-up contributing 20.0%. Response verbosity did not differ significantly by experience level (Kruskal-Wallis $H = 0.774$, $p = .679$), confirming the absence of length-related confounds in record length. These categories describe constructed profiles and must not be interpreted as recruited demographics.

Table 2. Synthetic Scenario and Structural Profile ($N = 50$)

| **Variable** | **Category** | $n$ | **%** |
|---|---|---|---|
| Age Range | 18–24 | 8 | 16.0 |
| | 25–34 | 22 | 44.0 |
| | 35–44 | 19 | 38.0 |
| | 45–54 | 1 | 2.0 |
| Experience Level | Beginner | 14 | 28.0 |
| | Intermediate | 18 | 36.0 |
| | Advanced | 18 | 36.0 |
| Interview Phase | Warm-up | 10 | 20.0 |
| | Task Walkthrough | 20 | 40.0 |
| | Reflection | 20 | 40.0 |
| Question Context | Initial Expectations | 10 | 20.0 |
| | First Login | 10 | 20.0 |
| | Onboarding Flow | 10 | 20.0 |
| | Feature Discovery | 10 | 20.0 |
| | Overall Experience | 10 | 20.0 |
| Response Length | Mean ($SD$) | 32.4 words (6.0) | |
| | Range | 22–50 words | |

### *4.2. Thematic Codebook and Frequency Distribution*

Table 3 presents the thematic codebook. Eight friction themes were identified inductively from the full corpus. Expectation Disconfirmation (T7, 24%) and Time-to-Value Friction (T8, 22%) are the two most prevalent themes, followed by Navigation Opacity (T2, 20%) and Onboarding Discontinuity (T3, 20%). Information Overload (T1, 14%) and Goal-Means Misalignment (T5, 12%) occupy an intermediate prevalence band, while Personalization Deficit (T4, 8%) and Self-Efficacy Impairment (T6, 8%) are the least frequently coded themes. As illustrated in Figure 1, which presents prevalence estimates with 95% Clopper-Pearson exact binomial confidence intervals, all themes are distinguishable from zero prevalence with meaningful confidence.

The concentration of friction in the expectation and temporal dimensions is consistent with EDT [9] and TAM [24] predictions for first-use digital service contexts. All eight codebook categories had appeared by the seventh ordered record (P007). We call this "codebook coverage" rather than theoretical saturation: an early plateau in a deliberately constructed synthetic corpus cannot demonstrate conceptual completeness for real onboarding experiences. The remaining 43 records were analysed to estimate within-corpus frequencies, evaluate coding agreement, inspect co-occurrence, and generate exploratory subgroup hypotheses.

Table 3. Thematic Codebook: Definitions, Theoretical Anchors, and Prevalence ($N = 50$)

| Code | Theme | Operational Definition | Theoretical Anchor | $n$ | % |
|---|---|---|---|---|---|
| T1 | Information Overload | Explicit expression of excessive information, feature volume, or decision paralysis during onboarding | Cognitive Load Theory [8] | 7 | 14 |
| T2 | Navigation Opacity | Reported inability to locate features or identify pathways; accidental discovery of functionality | Information Architecture [38] | 10 | 20 |
| T3 | Onboarding Discontinuity | Abrupt guidance termination, inability to revisit tutorials, or disconnect between tutorial and product | Continuity of Learning [39] | 10 | 20 |
| T4 | Personalization Deficit | Generic onboarding despite role selection; mismatch with skill level or task context | Adaptive Systems [40] | 4 | 8 |
| T5 | Goal-Means Misalignment | Understanding of task mechanics without clarity on outcome rationale; feature-focused vs. goal-oriented framing | Goal-Setting Theory [17] | 6 | 12 |
| T6 | Self-Efficacy Impairment | Expressed doubt about setup correctness, anxiety about missed steps, or helplessness | Self-Efficacy Theory [10] | 4 | 8 |
| T7 | Expectation Disconfirmation | Explicit comparison of pre-entry expectation to actual experience, resulting in negative disconfirmation | EDT [9] | 12 | 24 |
| T8 | Time-to-Value Friction | Explicit reference to wasted time, delayed productivity, or disproportionate learning investment | TAM [24] | 11 | 22 |

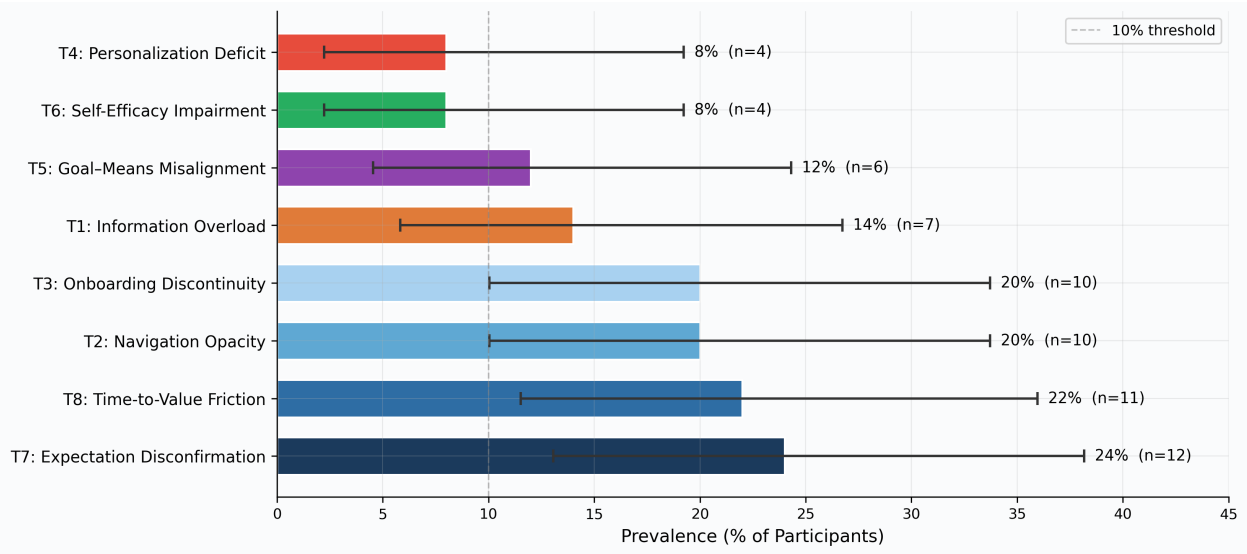


Figure 1. Thematic frequency distribution with 95% exact binomial confidence intervals (Clopper-Pearson method). Themes are ordered by prevalence. $N = 50$.

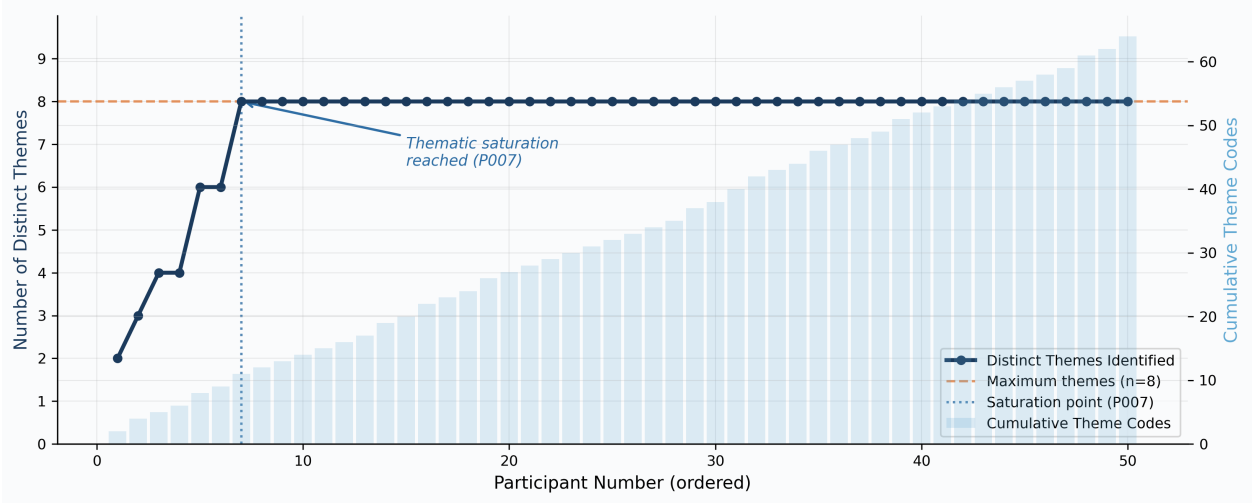


Figure 2. Cumulative codebook coverage across 50 ordered synthetic records. All eight candidate themes appeared by P007; the plateau is a property of this corpus and is not evidence of population-level theoretical saturation. $N = 50$.

### *4.3. Exploratory Theme Prevalence by Experience Level*

Table 4 presents the cross-tabulation of theme prevalence by user experience level with Fisher's exact test results. Two associations meet both nominal and Bonferroni-adjusted thresholds, but remain hypothesis-generating because the metadata are synthetic and the estimates are imprecise.

Onboarding Discontinuity (T3) is strongly associated with Beginner- labelled records (64% vs. 6% Intermediate, 0% Advanced; OR $= 63.00$, 95% CI $[6.52, 608.95]$, $p < .001$). This pattern is theoretically consistent with CLT [8]: users with limited prior schema for SaaS interfaces are most dependent on continuous guidance and are most adversely affected by abrupt guidance termination. Research on cognitive load in digital learning contexts confirms that schema deficiency at the outset of a task causes abrupt instructional withdrawal to produce irreversible comprehension gaps, consistent with the extraneous load mechanism identified here [18]. The verbatim synthetic response P008 illustrates this pattern directly: "I couldn't go back to it later... I'm probably doing things wrong." This statement illustrates the self-efficacy erosion that accompanies discontinuous onboarding for novice users.

Expectation Disconfirmation (T7), conversely, is more prevalent among Advanced-labelled records (50% vs. 17% Intermediate, 0% Beginner; OR $= 9.67$, 95% CI $[2.14, 43.56]$, $p = .004$). This pattern is explained by EDT [9]: users with extensive prior SaaS experience bring more elaborated performance expectations to new platforms, creating a wider potential gap between expectation and experience. A meta-analytic review of the expectancy-disconfirmation mechanism across digital and public service contexts confirms that expert users apply stricter evaluative standards and are correspondingly more susceptible to negative disconfirmation when a platform fails to meet anticipated standards [21]. The synthetic response P011 exemplifies this mechanism: "The video made it seem intuitive, but in practice, everything feels buried." This statement reflects negative disconfirmation rooted in prior expertise rather than novice confusion.

### *4.4. Exploratory Theme Prevalence by Interview Phase*

Table 5 and Figure 3 present phase-level theme distributions with chi-square screening results. Three themes meet the adjusted threshold, but phase labels were built into the scenarios and correlate with experience labels; these patterns describe corpus structure rather than causal temporal effects.

Navigation Opacity (T2) is concentrated in the Task Walkthrough phase (50% vs. 0% Warm-up, 0% Reflection; $\chi^2(2) = 18.750$, $p < .001$), reflecting the expected timing of active interface navigation during direct platform interaction. Expectation Disconfirmation (T7) is concentrated in the Warm-up phase (100% vs. 0% Task

Table 4. Theme Prevalence by Experience Level with Fisher's Exact Tests

| Theme | Beg. % | Int. % | Adv. % | Total % | OR | $p$ | Sig. |
|---|---|---|---|---|---|---|---|
| T1: Information Overload | 7 | 11 | 22 | 14 | 0.38 | .657 | ns |
| T2: Navigation Opacity | 7 | 22 | 28 | 20 | 0.23 | .246 | ns |
| T3: Onboarding Discontinuity | 64 | 6 | 0 | 20 | 63.00 | $< .001$ | *** |
| T4: Personalization Deficit | 7 | 6 | 11 | 8 | 0.85 | 1.000 | ns |
| T5: Goal-Means Misalignment | 14 | 6 | 17 | 12 | 1.33 | 1.000 | ns |
| T6: Self-Efficacy Impairment | 14 | 11 | 0 | 8 | 2.83 | .310 | ns |
| T7: Expectation Disconfirmation | 0 | 17 | 50 | 24 | 9.67 | .004 | ** |
| T8: Time-to-Value Friction | 7 | 50 | 6 | 22 | 0.20 | .148 | ns |

*Note.* OR = odds ratio (Beginner vs. non-Beginner for T3; Advanced vs. non-Advanced for T7). Fisher's exact test used due to expected cell counts $< 5$. ** $p < .01$; *** $p < .001$; ns = not significant.

Walkthrough, 10% Reflection; $\chi^2(2) = 40.132$, $p < .001$), confirming that expectation-based appraisal is a pre-interaction phenomenon activated before direct platform contact. This temporal pattern is consistent with the pre-adoption expectation formation process described in recent digital service adoption scholarship [15], in which vendor marketing and peer referrals construct performance standards prior to any direct system experience. Time-to-Value Friction (T8) escalates markedly in the Reflection phase (45% vs. 10% Task Walkthrough, 0% Warm-up; $\chi^2(2) = 10.664$, $p = .005$), indicating that productivity cost appraisal is a retrospective process activated after sufficient exposure to enable temporal comparison. This finding is consistent with meta-analytic evidence showing that learners' continuance intentions toward digital platforms are shaped more by cumulative experience appraisal than by immediate first-contact impressions [22].

Table 5. Theme Prevalence by Interview Phase with Chi-Square Tests

| Theme | WU % | TW % | Ref. % | $\chi^2(2)$ | $p$ | Sig. |
|---|---|---|---|---|---|---|
| T1: Information Overload | 30 | 15 | 5 | 3.488 | .175 | ns |
| T2: Navigation Opacity | 0 | 50 | 0 | 18.750 | $< .001$ | *** |
| T3: Onboarding Discontinuity | 0 | 25 | 25 | 3.125 | .210 | ns |
| T4: Personalization Deficit | 0 | 15 | 5 | 2.446 | .294 | ns |
| T5: Goal-Means Misalignment | 0 | 15 | 15 | 1.705 | .426 | ns |
| T6: Self-Efficacy Impairment | 0 | 5 | 15 | 2.446 | .294 | ns |
| T7: Expectation Disconfirmation | 100 | 0 | 10 | 40.132 | $< .001$ | *** |
| T8: Time-to-Value Friction | 0 | 10 | 45 | 10.664 | .005 | ** |

*Note.* WU = Warm-up ($n = 10$); TW = Task Walkthrough ($n = 20$); Ref. = Reflection ($n = 20$). Pearson chi-square statistics are exploratory because several expected cells are sparse. ** $p < .01$; *** $p < .001$; ns = not significant.

### *4.5. Theme Co-occurrence and Structural Patterns*

Figure 4 presents the pairwise theme co-occurrence matrix, reporting the number of records for which any two themes were simultaneously coded. The largest cell is only three (T1–T7), and the T2–T8 and T6–T8 cells each contain two records. These sparse cells are insufficient to establish directional or sequential relationships. They instead motivate testable propositions: navigation opacity may delay time-to-value, and uncertainty about correct setup may compound perceived productivity costs. Authentic longitudinal or in-situ data are required to test those mechanisms.

### *4.6. Non-Parametric Tests and Effect Sizes*

Table 6 presents Kruskal-Wallis $H$-test results for theme count and word count across experience levels and interview phases. No test reaches statistical significance, and all $\eta^2$ effect sizes are negligible (below 0.01), indicating that aggregate quantitative indices do not differ significantly by user subgroup. This finding is

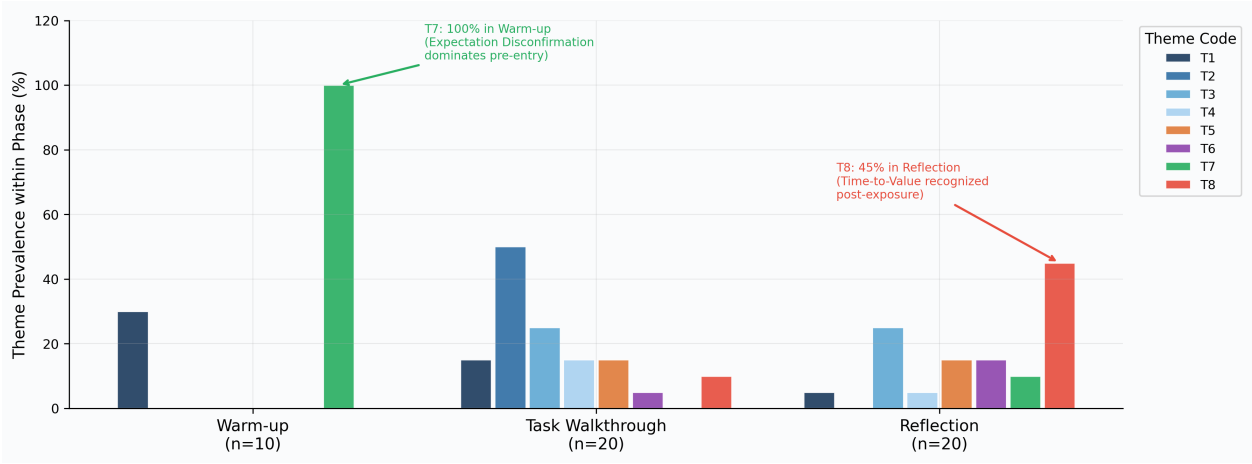


Figure 3. Theme prevalence by interview phase (% within phase). Statistically significant phase-level variations are annotated for T2 (Navigation Opacity), T7 (Expectation Disconfirmation), and T8 (Time-to-Value Friction). $N = 50$.


Figure 4. Pairwise theme co-occurrence matrix. Cell values represent the number of records coded for both themes simultaneously. Diagonal entries are excluded. $N = 50$.

methodologically important: it demonstrates that the substantive subgroup differences identified in the theme-specific Fisher's exact tests (Section 4.3) are qualitatively specific. They reside in the pattern of which themes are present, not in the volume of themes per response. This distinction between qualitative specificity and quantitative volume aligns with the interpretive logic advocated in mixed-method information systems research [32].

Table 6. Kruskal-Wallis Tests with Effect Sizes ($\eta^2$)

| **Outcome** | **Grouping** | $H$ | $df$ | $p$ | $\eta^2$ | **Effect** |
|---|---|---|---|---|---|---|
| Theme Count | Experience Level | 0.543 | 2 | .762 | 0.000 | Negligible |
| Word Count | Experience Level | 0.774 | 2 | .679 | 0.000 | Negligible |
| Theme Count | Interview Phase | 1.118 | 2 | .572 | 0.000 | Negligible |

*Note.* $\eta^2 = (H - k + 1)/(N - k)$. Effect size benchmarks: negligible $< 0.01$; small 0.01–0.06; medium 0.06–0.14; large $> 0.14$ [42].

### *4.7. Spearman Correlation Structure*

Table 7 presents the Spearman correlations among theme count, word count, experience label, and phase. The association between experience label and phase ($\rho = -0.529$, $p < .01$) confirms that these two design variables are confounded in the synthetic corpus. Therefore, their separate subgroup patterns cannot be interpreted as independent effects, and no multivariable adjustment is defensible with $N = 50$ and sparse outcomes.

Table 7. Spearman Rank Correlation Matrix ($\rho$)

| **Variable** | **(1)** | **(2)** | **(3)** | **(4)** |
|---|---|---|---|---|
| (1) Theme Count | 1.000 | | | |
| (2) Word Count | 0.221 | 1.000 | | |
| (3) Experience Level | 0.105 | 0.051 | 1.000 | |
| (4) Interview Phase | −0.116 | 0.126 | −0.529** | 1.000 |

*Note.* ** $p < .01$ (two-tailed). Experience Level and Interview Phase are coded as ordinal variables (1 = Beginner/Warm-up through 3 = Advanced/Reflection). Lower triangle omitted for clarity.

### *4.8. Representative Synthetic Excerpts*

Table 8 presents representative verbatim quotations for each of the eight themes, selected to exemplify the explicit linguistic markers that triggered thematic coding. These excerpts serve the dual function of providing transparent coding documentation and illustrating the linguistic texture encoded in the synthetic records; they must not be read as participant testimony. Qualitatively, the excerpts show a consistent pattern: the constructed speakers express awareness of the platform's underlying potential while simultaneously experiencing specific structural failures in the onboarding process. This cognitive-affective profile is consistent with IS continuance research characterising post-adoption satisfaction as a function of expectation recalibration rather than absolute performance [20, 22].

### *4.9. Robustness Checks*

Two robustness checks were retained to assess coding stability: dual-coder agreement and an odd/even split-half comparison. The former is reported in Table 9; the latter is interpreted as an internal descriptive check rather than external validation.

*4.9.1. RC-1: Dual-Coder Comparison* After codebook calibration, both coders independently applied the same eight operational definitions to the same 50 records. Nielsen's heuristics [13] informed calibration examples only; they were not used as a second, incommensurable coding instrument. Table 9 reports Cohen's $\kappa$ for each of the eight themes. The mean $\kappa$ across themes is 0.753, falling in the "substantial agreement" range by the Landis and Koch [43] taxonomy. Five of eight themes achieved $\kappa \geq 0.80$ (almost perfect agreement). The lowest $\kappa$ was observed for T6 (Self-Efficacy Impairment, $\kappa = 0.457$, moderate), reflecting the subtlety of self-efficacy expression in natural

Table 8. Representative Synthetic Excerpts by Friction Theme

| Theme | Record | Repository Text |
|---|---|---|
| T1: Information Overload | P013, Beginner, Reflection | "The tooltips are helpful, but there are too many of them. It's like death by a thousand tooltips. I started ignoring them after the first five, so I probably missed important information." |
| T2: Navigation Opacity | P009, Intermediate, Task Walkthrough | "There are so many features hidden in menus. I feel like I'm using maybe 20% of what this tool can do, but I don't know what the other 80% is or if I even need it." |
| T3: Onboarding Discontinuity | P008, Beginner, Reflection | "I liked the interactive tutorial, but it disappeared too fast. And I couldn't go back to it later. So now I'm stuck trying to remember what it said, and I'm probably doing things wrong." |
| T4: Personalization Deficit | P027, Intermediate, Task Walkthrough | "I liked that I could choose my role during signup – that was smart. But then the onboarding didn't really change based on my role. It was still generic." |
| T5: Goal-Means Misalignment | P003, Intermediate, Reflection | "They showed me how to create a project, but not WHY I would want to create a project, or what happens after. It's very feature-focused, not goal-focused." |
| T6: Self-Efficacy Impairment | P042, Intermediate, Task Walkthrough | "I got through the onboarding, but I'm not confident I did it right. Like, did I set things up correctly? Am I missing something important? There's no way to know." |
| T7: Expectation Disconfirmation | P011, Advanced, Warm-up | "I watched a demo video before signing up, and it looked amazing. But the actual experience is not the same. The video made it seem intuitive, but in practice, everything feels buried." |
| T8: Time-to-Value Friction | P045, Intermediate, Reflection | "I'm torn. The tool is powerful, but the onboarding is weak. I'm spending more time learning than doing, and that's not sustainable." |

language. The mean agreement rate across all themes was 94.5%, providing strong evidence that the operational definitions can be applied consistently to this corpus. Agreement does not validate the realism of the synthetic text or the external validity of the themes [36].

Table 9. Inter-Rater Reliability: Cohen's $\kappa$ by Theme

| Theme | C1 $n$ | C2 $n$ | Agree.% | $\kappa$ | Interpretation |
|---|---|---|---|---|---|
| T1: Information Overload | 7 | 7 | 100.0 | 1.000 | Almost perfect |
| T2: Navigation Opacity | 10 | 8 | 92.0 | 0.730 | Substantial |
| T3: Onboarding Discontinuity | 10 | 11 | 94.0 | 0.819 | Almost perfect |
| T4: Personalization Deficit | 4 | 3 | 98.0 | 0.847 | Almost perfect |
| T5: Goal-Means Misalignment | 6 | 3 | 94.0 | 0.638 | Substantial |
| T6: Self-Efficacy Impairment | 4 | 4 | 92.0 | 0.457 | Moderate |
| T7: Expectation Disconfirmation | 12 | 10 | 96.0 | 0.884 | Almost perfect |
| T8: Time-to-Value Friction | 11 | 6 | 90.0 | 0.652 | Substantial |
| **Mean across themes** | | | **94.5** | **0.753** | **Substantial** |

*Note.* C1 and C2 independently applied the same eight-theme operational codebook. $\kappa$ benchmarks: 0.41–0.60 = moderate; 0.61–0.80 = substantial; 0.81–1.00 = almost perfect [43].

*4.9.2. RC-2: Split-Half Reliability* To assess internal replication consistency, the sample was split into odd-numbered ($n = 25$) and even-numbered ($n = 25$) record halves, and theme prevalence was computed independently for each half. The prevalence estimates were broadly consistent across splits, with no theme exhibiting a cross-half

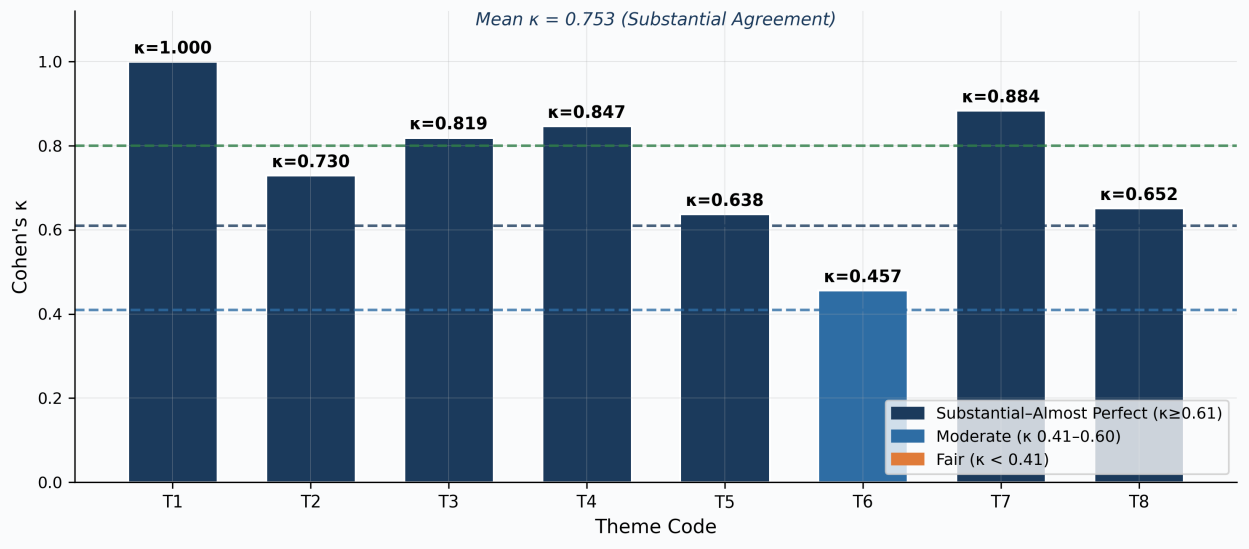


Figure 5. Cohen's $\kappa$ values by theme for dual independent coders. Horizontal reference lines indicate the Landis and Koch [43] interpretation benchmarks. $N = 50$.

discrepancy that reverses directional conclusions. The rank ordering of themes by prevalence, with T7 and T8 as the most frequent and T4 and T6 as the least frequent, is preserved in both halves. This shows internal distributional stability within the constructed corpus, but it does not establish stability in real-user samples.

## 5. Discussion

### *5.1. Theoretical Contributions*

The study's contribution is a theory-linked candidate typology and three propositions, not the discovery of eight previously unknown UX problems. First, the typology connects interface-level heuristics to mechanisms that unfold across the onboarding process: excessive or poorly organised information can raise extraneous load (T1–T2), broken guidance can undermine self-efficacy (T3–T6), feature demonstrations can fail to connect with user goals (T5), and pre-entry promises and elapsed learning effort can shape post-entry evaluation (T7–T8). This mapping supplies the integrative element missing from a simple checklist of usability violations.

Second, the sparse co-occurrence patterns suggest a mechanism that should be tested rather than asserted: navigation opacity may increase search time, which may then reduce self-efficacy and delay time-to-value. This sequence connects CLT, Self-Efficacy Theory, and TAM while remaining consistent with the distinction between behavioural events and users' interpretation of those events.

Third, the experience-labelled contrasts motivate an expertise- contingency proposition: persistent guidance may be most valuable for novices, whereas expectation calibration and skippable support may be more important for experienced users. This proposition is consistent with the expertise-reversal effect [44], but the synthetic corpus cannot establish the moderation effect. Field experiments should vary guidance persistence and user expertise while measuring task success, confidence, and time-to-value.

### *5.2. Practical and Design Implications*

Table 10 translates each candidate theme into a concrete, testable intervention. These are design hypotheses rather than validated prescriptions. In particular, real-time detection of T7 need not infer emotion from private text. A system can compare the acquisition promise or selected use case with early events: for example, a user who arrives from a "setup in five minutes" campaign, exceeds that time, repeats navigation loops, or opens help repeatedly can

be offered a transparent progress estimate, a shorter path, or an option to schedule assistance. The trigger should remain explainable and easy to dismiss.

Table 10. Candidate Design Interventions and Validation Measures

| Theme | Candidate intervention | Validation measure |
|---|---|---|
| T1 | Progressive disclosure and task-prioritised content | Recall, completion, perceived load |
| T2 | Goal-based navigation and contextual search labels | Path length, misclicks, task success |
| T3 | Persistent, resumable walkthrough with history | Return-to-help rate, completion, confidence |
| T4 | Role and expertise branch with manual override | Skip rate, relevance, completion by expertise |
| T5 | Explain the outcome before demonstrating the feature | Goal recall, transfer to a new task |
| T6 | Setup validation, safe sandbox, and reversible actions | Confidence, error recovery, exploration |
| T7 | Compare promised and actual progress; offer recovery when the gap widens | Expectation gap, abandonment, support use |
| T8 | Show milestones and deliver one valued task early | Time to first value, activation, day-7 return |

Platform context is likely to moderate these interventions. B2B enterprise products may require role-based configuration, governance, and collaborative setup, whereas self-serve B2C products may depend more on rapid first value and minimal mandatory instruction. Future validation should therefore stratify by B2B/B2C market, organisational size, workflow complexity, and whether onboarding is individual or team-based.

### *5.3. Limitations*

The dominant limitation is data provenance. The records are synthetic; no human users were interviewed, and the repository does not report the generating model, prompt, training sources, sampling parameters, or editing process. The language may therefore reproduce canonical UX concepts already present in design literature, making theory-consistent themes partly circular. Synthetic text also suppresses the hesitation, contradiction, contextual detail, recall error, and unexpected concerns that give authentic qualitative evidence its value. The labels cannot support demographic or market generalisation, and the quoted passages illustrate corpus content rather than lived testimony.

The short cross-sectional records cannot establish sequence or causality. Experience label and phase are correlated, sparse cells yield wide confidence intervals, and the sensitivity analysis shows that only very large subgroup differences are detectable. Null results are therefore inconclusive, while significant results may reflect scenario construction. The early codebook-coverage plateau is not theoretical saturation. Coder agreement demonstrates consistent application of the codebook, not construct validity. Finally, the removed custom sentiment analysis lacked sufficient auditability and contributed no significant result.

The necessary next step is preregistered validation using authentic data: 10–15 interviews can first test whether the codebook misses important phenomena, followed by a larger multi-platform sample or in-situ diary/telemetry study for subgroup and temporal hypotheses. Authentic studies should document consent, recruitment, platform type, interview prompts, coder training, the complete coded matrix, and analysis scripts.

## 6. Conclusion

This study demonstrates a transparent thematic–computational workflow on 50 synthetic SaaS-onboarding records. It recovers an eight-theme candidate typology that integrates established usability concerns with CLT, EDT, TAM, Goal-Setting Theory, and Self-Efficacy Theory. Dual-coder agreement indicates that the operational definitions can be applied consistently to this corpus, while wide confidence intervals, low power, phase–experience confounding, and synthetic provenance sharply limit substantive inference.

The defensible contribution is therefore methodological and propositional: a reproducible way to move from explicit coding rules to uncertainty-aware summaries, a mapping between the candidate themes and established UX frameworks, and design hypotheses for persistent guidance, expectation calibration, goal-oriented instruction, and early value delivery. Claims about real users await validation through authentic interviews, diary or observational studies, and field experiments across multiple SaaS contexts.

## REFERENCES


1. M. Paulus, S. Jordanow, and J. A. Millemann, "Adoption factors of digital services: A systematic literature review," *Service Science*, vol. 14, no. 4, pp. 318–350, 2022, doi: 10.1287/serv.2022.0305.
2. K. Tamilmani, N. P. Rana, and Y. K. Dwivedi, "Consumer acceptance and use of information technology: A meta-analytic evaluation of UTAUT2," *Information Systems Frontiers*, vol. 23, no. 4, pp. 987–1005, 2021, doi: 10.1007/s10796-020-10007-6.
3. S. Petrilli, L. Galuppo, and S. C. Ripamonti, "Digital onboarding: Facilitators and barriers to improve worker experience," *Sustainability*, vol. 14, no. 9, art. 5684, 2022, doi: 10.3390/su14095684.
4. K. F. Sani, T. A. Adisa, O. D. Adekoya, and E. S. Oruh, "Digital onboarding and employee outcomes: Empirical evidence from the UK," *Management Decision*, vol. 61, no. 3, pp. 637–654, 2022, doi: 10.1108/MD-11-2021-1528.
5. D. A. Norman, *The Design of Everyday Things*, rev. ed. New York: Basic Books, 2013.
6. Q. N. Nguyen, A. Sidorova, and R. Torres, "User interactions with chatbot interfaces vs. menu-based interfaces: An empirical study," *Computers in Human Behavior*, vol. 128, art. 107093, 2022, doi: 10.1016/j.chb.2021.107093.
7. M. Yan, R. Filieri, and M. Gorton, "Continuance intention of online technologies: A systematic literature review," *International Journal of Information Management*, vol. 58, art. 102315, 2021, doi: 10.1016/j.ijinfomgt.2021.102315.
8. J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cognitive Science*, vol. 12, no. 2, pp. 257–285, 1988, doi: 10.1207/s15516709cog1202_4.
9. R. L. Oliver, "A cognitive model of the antecedents and consequences of satisfaction decisions," *Journal of Marketing Research*, vol. 17, no. 4, pp. 460–469, 1980, doi: 10.1177/002224378001700405.
10. A. Bandura, *Self-Efficacy: The Exercise of Control*. New York: W. H. Freeman, 1997.
11. C. Ardito, P. Buono, D. Caivano, M. F. Costabile, and R. Lanzilotti, "Investigating and promoting UX practice in industry: An experimental study," *International Journal of Human-Computer Studies*, vol. 72, no. 6, pp. 542–551, 2014, doi: 10.1016/j.ijhcs.2013.10.004.
12. A. S. Al-Adwan, N. Li, and A. Al-Adwan, "Extending the technology acceptance model to predict university students' intentions to use metaverse-based learning platforms," *Education and Information Technologies*, vol. 28, no. 11, pp. 15381–15413, 2023, doi: 10.1007/s10639-023-11816-3.
13. J. Nielsen, *Usability Engineering*. San Diego, CA: Academic Press, 1993.
14. N. Korotkova, J. Benders, P. Mikalef, and D. Cameron, "Maneuvering between skepticism and optimism about hyped technologies: Building trust in digital twins," *Information & Management*, vol. 60, no. 4, art. 103787, 2023, doi: 10.1016/j.im.2023.103787.
15. Y. K. Dwivedi, N. P. Rana, A. Jeyaraj, M. Clement, and M. D. Williams, "Re-examining the unified theory of acceptance and use of technology (UTAUT): Towards a revised theoretical model," *Information Systems Frontiers*, vol. 21, no. 3, pp. 719–734, 2019, doi: 10.1007/s10796-017-9774-y.
16. A. Mishra, A. Shukla, N. P. Rana, and Y. K. Dwivedi, "Re-examining post-acceptance model of information systems continuance: A revised theoretical model using MASEM approach," *International Journal of Information Management*, vol. 68, art. 102571, 2023, doi: 10.1016/j.ijinfomgt.2022.102571.
17. E. A. Locke and G. P. Latham, "Building a practically useful theory of goal setting and task motivation: A 35-year odyssey," *American Psychologist*, vol. 57, no. 9, pp. 705–717, 2002, doi: 10.1037/0003-066X.57.9.705.
18. A. Skulmowski and K. M. Xu, "Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load," *Educational Psychology Review*, vol. 34, no. 1, pp. 171–196, 2022, doi: 10.1007/s10648-021-09624-7.
19. M. Naeem, W. Ozuem, K. Howell, and S. Ranfagni, "A step-by-step process of thematic analysis to develop a conceptual model in qualitative research," *International Journal of Qualitative Methods*, vol. 22, pp. 1–18, 2023, doi: 10.1177/16094069231205789.
20. A. Bhattacherjee, "Understanding information systems continuance: An expectation-confirmation model," *MIS Quarterly*, vol. 25, no. 3, pp. 351–370, 2001, doi: 10.2307/3250921.
21. J. Zhang, W. Chen, N. Petrovsky, and R. M. Walker, "The expectancy-disconfirmation model and citizen satisfaction with public services: A meta-analysis and an agenda for best practice," *Public Administration Review*, vol. 82, no. 1, pp. 147–159, 2022, doi: 10.1111/puar.13368.
22. J. Dai, X. Zhang, and C. Wang, "A meta-analysis of learners' continuance intention toward online education platforms," *Education and Information Technologies*, vol. 29, no. 16, pp. 21833–21868, 2024, doi: 10.1007/s10639-024-12654-7.
23. N. Hampel and K. Sassenberg, "Enactive mastery experience improves attitudes towards digital technology via self-efficacy: A pre-registered quasi-experiment," *Behaviour and Information Technology*, vol. 43, no. 2, pp. 298–311, 2023, doi: 10.1080/0144929X.2022.2162436.
24. F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989, doi: 10.2307/249008.
25. S. FakhrHosseini, K. Chan, C. Lee, H. Son, and M. Jeon, "User adoption of intelligent environments: A review of technology adoption models, challenges, and prospects," *International Journal of Human-Computer Interaction*, vol. 40, no. 4, pp. 986–998, 2024, doi: 10.1080/10447318.2022.2118851.
26. V. Braun and V. Clarke, "Is thematic analysis used well in health psychology? A critical review of published research, with recommendations for quality practice and reporting," *Health Psychology Review*, vol. 17, no. 4, pp. 695–718, 2023, doi:

10.1080/17437199.2022.2161594.
27. G. Guest, A. Bunce, and L. Johnson, "How many interviews are enough? An experiment with data saturation and variability," *Field Methods*, vol. 18, no. 1, pp. 59–82, 2006, doi: 10.1177/1525822X05279903.
28. V. Braun and V. Clarke, "Reflecting on reflexive thematic analysis," *Qualitative Research in Sport, Exercise and Health*, vol. 11, no. 4, pp. 589–597, 2019, doi: 10.1080/2159676X.2019.1628806.
29. J. W. Creswell and V. L. Plano Clark, *Designing and Conducting Mixed Methods Research*, 3rd ed. Thousand Oaks, CA: Sage Publications, 2018.
30. V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006, doi: 10.1191/1478088706qp063oa.
31. J. Grimmer, M. E. Roberts, and B. M. Stewart, *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton, NJ: Princeton University Press, 2022.
32. V. Venkatesh, S. A. Brown, and H. Bala, "Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems," *MIS Quarterly*, vol. 37, no. 1, pp. 21–54, 2013, doi: 10.25300/MISQ/2013/37.1.02.
33. M. Rafiei, "Qualitative Text Datasets for UX Research," GitHub repository, 2026. [Online]. Available: https://github.com/mohsen-rafiei/Qualitative-Text-Datasets-for-UX-Research
34. M. Q. Patton, *Qualitative Research and Evaluation Methods*, 3rd ed. Thousand Oaks, CA: Sage Publications, 2002.
35. M. Hennink and B. N. Kaiser, "Sample sizes for saturation in qualitative research: A systematic review of empirical tests," *Social Science & Medicine*, vol. 292, art. 114523, 2022, doi: 10.1016/j.socscimed.2021.114523.
36. M. L. McHugh, "Interrater reliability: The kappa statistic," *Biochemia Medica*, vol. 22, no. 3, pp. 276–282, 2012, doi: 10.11613/BM.2012.031.
37. M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational Linguistics*, vol. 37, no. 2, pp. 267–307, 2011, doi: 10.1162/COLI_a_00049.
38. J. Salminen, K. Guan, S. Jung, and B. J. Jansen, "A survey of 15 years of data-driven persona development," *International Journal of Human-Computer Interaction*, vol. 37, no. 18, pp. 1685–1708, 2021, doi: 10.1080/10447318.2021.1908670.
39. D. R. Garrison and H. Kanuka, "Blended learning: Uncovering its transformative potential in higher education," *Internet and Higher Education*, vol. 7, no. 2, pp. 95–105, 2004, doi: 10.1016/j.iheduc.2004.02.001.
40. G. F. Tondello, R. R. Wehbe, L. Diamond, M. Busch, A. Marczewski, and L. E. Nacke, "The gamification user types hexad scale," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, ACM, 2016, pp. 229–243, doi: 10.1145/2967934.2968082.
41. J. Sweller, J. J. G. van Merriënboer, and F. G. W. C. Paas, "Cognitive architecture and instructional design," *Educational Psychology Review*, vol. 10, no. 3, pp. 251–296, 1998, doi: 10.1023/A:1022193728205.
42. J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. New York, NY: Routledge, 2013, doi: 10.4324/9780203771587.
43. J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977, doi: 10.2307/2529310.
44. S. Kalyuga, P. Ayres, P. Chandler, and J. Sweller, "The expertise reversal effect," *Educational Psychologist*, vol. 38, no. 1, pp. 23–31, 2003, doi: 10.1207/S15326985EP3801_4.
45. Y.-M. Wang and C.-L. Wei, "What drives students' AI learning behavior: A perspective of AI anxiety," *Interactive Learning Environments*, vol. 32, no. 7, pp. 3534–3550, 2024, doi: 10.1080/10494820.2022.2153147.
46. K. Tamilmani, N. P. Rana, R. Nunkoo, V. Raghavan, and Y. K. Dwivedi, "Indian travellers' adoption of Airbnb platform," *Information Systems Frontiers*, vol. 24, no. 1, pp. 77–96, 2022, doi: 10.1007/s10796-020-10060-1.
47. S. Rahmani, A. B. P. Kasmo, M. Rifqi, and S. Setiawan, "The paradox of AI adoption in emerging economies: A structural equation analysis of usage intensity and SMEs performance," *Statistics, Optimization & Information Computing*, vol. 15, no. 4, pp. 2715–2726, 2026, doi: 10.19139/soic-2310-5070-3036.
48. B. R. Sanoussi, O. Benjelloun Andaloussi, and M. A. Marhraoui, "What drives tourists to use conversational AI? Evidence from an extended UTAUT2 model," *Statistics, Optimization & Information Computing*, vol. 15, no. 6, pp. 4774–4788, 2026, doi: 10.19139/soic-2310-5070-3365.
49. S. M. Ferdous, S. N. E. Newaz, S. B. S. Mugdha, and M. Uddin, "Sentiment analysis in the transformative era of machine learning: A comprehensive review," *Statistics, Optimization & Information Computing*, vol. 13, no. 1, pp. 331–346, 2025, doi: 10.19139/soic-2310-5070-2113.
50. A. C. C. Grassmann, J. C. Feitosa, J. R. F. Brega, and K. A. P. da Costa, "Evaluation of transformer-based large language models for email spam detection using BERT, Phi, and Gemma," *Statistics, Optimization & Information Computing*, vol. 13, no. 2, pp. 459–473, 2025, doi: 10.19139/soic-2310-5070-2267.
51. International Organization for Standardization, *ISO 9241-11:2018 Ergonomics of Human-System Interaction—Part 11: Usability: Definitions and Concepts*. Geneva, Switzerland: ISO, 2018.