

PAPER

An Interactive Mobile Learning Platform for Digital Forensics Education: Design, Development, and Pedagogical Evaluation

Krommavut Nongnuch ,
Saowalak
Leelawongsarote  (✉),
Tawan Khunarsa, Suthin
Kasetratanachai 

Rajamangala University of
Technology Suvarnabhumi,
Ayutthaya, Thailand

saowalak.l@rmutsb.ac.th

ABSTRACT

Demand for digital forensics (DF) graduates is rising, yet undergraduate DF teaching still depends on costly desktop tools, fixed laboratory slots, and limited time for repeated practice—barriers that weigh most heavily on resource-constrained universities. This paper presents DFLearn-Mobile, a cross-platform mobile application for undergraduate DF education built with Flutter, Node.js, and MongoDB. The app delivers six short modules aligned with the NIST DF process model, each closing with an interactive task and quiz, plus a capstone branching case for end-to-end procedural reasoning. We evaluated the platform with $N = 80$ undergraduates at Rajamangala University of Technology Suvarnabhumi using a quasi-experimental design combining pre- and post-tests, the system usability scale (SUS), an adapted technology acceptance model (TAM) instrument, and server-side analytics. DF knowledge improved significantly (Wilcoxon $V = 3,240$, $p < .001$; all 80 learners improved; mean normalized gain $\langle g \rangle = 0.586$, medium band). The SUS score ($M = 61.03$) fell significantly below the 68 benchmark, identifying usability priorities for the next iteration. TAM construct means (3.85–3.95) indicated mild positive acceptance, with $PU \leftrightarrow BI$ the strongest correlation ($\rho = 0.458$, $p < .001$). The framework, instruments, and analytic toolkit offer a replicable basis for mobile DF teaching in resource-constrained settings. These are preliminary pilot findings; causal claims require a randomized controlled comparison.

KEYWORDS

mobile learning, digital forensics (DF) education, interactive learning platform, cybersecurity education, gamification, higher education

1 INTRODUCTION

Digital forensics (DF) is the disciplined process of identifying, preserving, analyzing, and presenting digital evidence [1], [2]. It is a core skill for cybersecurity,

Nongnuch, K., Leelawongsarote, S., Khunarsa, T., Kasetratanachai, S. (2026). An Interactive Mobile Learning Platform for Digital Forensics Education: Design, Development, and Pedagogical Evaluation. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(20), pp. 25–48. <https://doi.org/10.3991/ijim.v20i20.62646>

Article submitted 2026-06-15. Revision uploaded 2026-07-20. Final acceptance 2026-08-13.

© 2026 by the authors of this article. Published under CC-BY.

law enforcement, and incident-response practitioners. Demand for DF-capable graduates keeps rising as evidence sources multiply across mobile, cloud, and IoT, and as attacks grow more sophisticated [3], [4]. Universities have responded by adding DF tracks to computer science, IT, and cybersecurity degrees [5].

The teaching of DF, however, faces three persistent obstacles. Industry-standard tools—EnCase, FTK, X-Ways—are expensive and license-bound, limiting access at universities with tight budgets [6]. DF is a procedural craft: students need repeated, scaffolded practice on realistic artefacts and chain-of-custody routines, yet time-tabled labs offer too few hours for that [7]. Today's undergraduates also study on phones. They expect short, modular, gamified content [8], [9]. Desktop-only DF instruction sits awkwardly against this reality, especially at multi-campus universities like the one in this study.

Mobile learning (m-learning) helps close the gap. Across STEM and professional fields, mobile delivery supports ubiquity, situated practice, and short bursts of focused work [10], [11]. In cybersecurity, gamified mobile apps have lifted engagement, recall, and self-efficacy [12], [13]. Yet mobile platforms built specifically for *digital forensics* are rare. The few that exist tend to stop at awareness, terminology, or recall quizzes, and seldom train the procedural and analytical reasoning that DF actually demands [14], [35].

This paper addresses that gap by presenting *DFLearn-Mobile*, an interactive mobile learning platform purpose-built for undergraduate DF education. The platform integrates: (i) structured theory modules covering the DF process model [1]; (ii) interactive artefact-analysis simulations (file-signature inspection, metadata reading, log timeline reconstruction); (iii) gamified case investigations with adaptive difficulty; and (iv) scenario-based assessments aligned with course learning outcomes. We position *DFLearn-Mobile* within the m-learning and cybersecurity-education literature, describe its design rationale and implementation, and present a quasi-experimental evaluation protocol with undergraduate students at a Thai multi-campus university.

This study contributes to the literature in four major aspects:

1. It proposes a **mobile-first pedagogical framework** for DF education that integrates gamification, micro-learning, and interactive forensic simulations aligned with the NIST SP 800-86 process model [1].
2. It develops **DFLearn-Mobile**, a cross-platform interactive learning environment supporting Android and iOS deployment, specifically designed for use in resource-constrained higher education settings.
3. It introduces **scenario-based forensic investigation activities** aligned with the four NIST DF process phases—Collection, Examination, Analysis, and Reporting—to support procedural competency development beyond declarative recall.
4. It empirically evaluates the platform using learning-performance analysis (pre-/post-test), usability evaluation (SUS), technology acceptance assessment (adapted TAM), and platform analytics.

The remainder of the paper is organized as follows: Section 2 reviews related work on DF education, m-learning in cybersecurity, and interactive learning platforms. Section 3 describes the design and implementation of *DFLearn-Mobile*. Section 4 presents the evaluation methodology. Section 5 reports results. Section 6 discusses implications and limitations, and Section 7 concludes.

2 RELATED WORK

2.1 Digital forensics education

The DF process model formalized in NIST SP 800-86 [1]—*Collection, Examination, Analysis, Reporting*—has long served as the curricular spine of undergraduate DF instruction. Garfinkel [3] and subsequent reviews [4] argue that DF education must extend beyond procedural compliance to cultivate analytic reasoning over heterogeneous artefacts (file systems, logs, memory, mobile data, cloud traces). Reviews of curricular practice highlight persistent gaps between course content and operational practice, particularly in artefact diversity and case authenticity [5], [7]. Cost and licensing of commercial tools further constrain hands-on practice, motivating use of open-source tool-chains (e.g., Autopsy, Sleuth Kit, Volatility) and bespoke educational platforms [6]. Recent work has begun to explore immersive and simulation-based formats for DF instruction, including virtual-reality reconstructions of crime-scene seizure procedures [23].

2.2 Mobile learning in cybersecurity education

Meta-analyses of m-learning [10], [11] consistently report positive effects on engagement and outcome attainment when mobile experiences are designed around active interaction rather than passive content consumption. In cybersecurity-specific contexts, gamified mobile apps for phishing recognition, password hygiene, and security awareness have demonstrated measurable improvements in user behavior and self-efficacy [12], [13]. Scenario-based and narrative learning approaches have proven particularly effective in cultivating decision-making competencies relevant to incident response and forensics [15]. More recent bibliometric evidence confirms the sustained growth of mobile learning in higher education and its role in supporting flexible access, engagement, and scalable delivery across institutional contexts [32], [36]. Evidence from higher-education deployments similarly indicates that mobile learning techniques can improve learning outcomes and student perceptions when integrated into existing curricula [37].

2.3 Interactive and gamified platforms for cybersecurity skills

Capture-the-Flag (CTF) platforms, cyber-ranges, and serious games occupy an established place in cybersecurity skills training [16], [17], [24], [31]. Recent meta-analyses confirm that gamification produces small-to-medium effects on cognitive, motivational, and behavioral learning outcomes [29], with effect sizes moderated by game fiction and the combination of competition with collaboration. A systematic theoretical review of 118 underlying theories in gamification research has further clarified that self-determination theory and goal-setting theory are most frequently invoked to explain why gamified designs work [30]. However, most existing CTFs are desktop-bound and assume substantial prerequisite knowledge, limiting their suitability for foundational undergraduate DF instruction. Mobile-first gamified platforms designed for novices remain comparatively rare, and few systematically scaffold the DF process model from collection through reporting. More broadly, recent syntheses emphasize that gamification's benefits for engagement and achievement depend on design quality and should not be treated as automatically effective [33].

2.4 Pedagogical and acceptance frameworks

Evaluation of educational technologies routinely draws on the technology acceptance model (TAM) [18] and its extensions (UTAUT, UTAUT2) [19], the system usability scale (SUS) [20], and Kirkpatrick's four-level training evaluation model [21]. These instruments enable triangulated assessment of usability, perceived usefulness, behavioral intention, and learning outcomes, and are widely used in m-learning research [11]. Recent TAM-based structural analyses of mobile learning applications in higher education confirm perceived usefulness and perceived ease of use as key drivers of student satisfaction and adoption [38].

2.5 Theoretical framing

The pedagogical design of DFLearn-Mobile is grounded in four complementary learning theories. Constructivist learning [25] motivates the key-term highlights, formative quizzes with explanatory feedback, and scaffolded prerequisite chains that prompt learners to integrate each new artefact concept with the NIST process-model schema. Experiential learning [26] is realized as a cycle of interactive artefact inspection (concrete experience), feedback-mediated quizzes (reflection), module summaries (conceptualization), and the capstone branching scenario (active experimentation). Situated learning [27] underpins the case-based design of M2–M6, which situates chain-of-custody simulation, log timeline reconstruction, artefact matching, and report drafting within realistic investigative narratives. Finally, microlearning theory [28] motivates decomposing each module into ≤ 5 -minute offline-capable cards suited to the fragmented, mobile-first study patterns of multi-campus undergraduates. Together, these lenses provide the pedagogical justification for the platform's interaction design, assessment structure, and gamification choices described in Section 3.

2.6 Comparative analysis of existing platforms

To position the proposed work, Table 1 compares representative cybersecurity and DF learning platforms across five dimensions: platform type, mobile support, gamification, coverage of the DF process model, and evaluation methodology.

Table 1. Comparison of existing cybersecurity and digital forensics learning platforms

Study	Platform Type	Mobile Support	Gamification	DF Process Coverage	Evaluation
Chothia & Novakovic [16]	Offline CTF VM (desktop)	No	Partial (CTF score)	Examination only	Pilot survey
Vykopal et al. (KYPO) [17]	Cyber range (web)	No	Yes (scoring)	Investigation phases	Case studies
Giannakas et al. (CyberAware) [14]	Mobile awareness app	Yes	Yes (mini-games)	Awareness only (not DF)	Pre/post + survey
Karagiannis & Magkos [15]	CTF-based skills mapping	Partial	Yes	Selected skills	Skills mapping review
Hassenfeldt et al. [23]	VR DF (Bagging & Tagging)	No (VR HMD)	Limited	Collection only	Quasi-experimental
McDaniel et al. [13]	CTF course intervention	No	Yes (CTF)	Investigation only	Course outcomes
DFLearn-Mobile (this work)	Mobile DF learning platform	Yes (Android + iOS)	Yes (XP, badges, leaderboard)	Full NIST workflow (C–E–A–R)	Quantitative evaluation (pre/post + SUS + TAM + analytics)

2.7 Research gap

Synthesizing the literature and the comparison in Table 1, three gaps motivate the present work:

- **Domain gap:** mobile platforms dedicated to *digital forensics*—as distinct from generic cybersecurity awareness—remain scarce; most mobile cybersecurity apps target awareness or phishing recognition rather than DF procedural skills.
- **Pedagogical gap:** existing tools rarely scaffold the *full* NIST DF process model (Collection → Examination → Analysis → Reporting) with interactive artefact analysis suited to mobile form factors. CTFs typically address only investigation; VR platforms address only collection.
- **Empirical gap:** few studies report rigorous quantitative evaluations of mobile DF platforms in resource-constrained, multi-campus higher education contexts. DFLearn-Mobile is designed specifically to address these three gaps.

3 DFLEARN-MOBILE: DESIGN AND IMPLEMENTATION

3.1 Design principles

Five design principles guided the platform development:

1. **Process-model fidelity:** Content is organized around the NIST DF process model [1], reinforcing transferable conceptual structure.
2. **Active interaction over passive consumption:** Every module concludes with an interactive task (drag-and-drop ordering, artefact inspection, timeline construction) rather than only multiple-choice review.
3. **Scaffolded difficulty:** Modules progress from guided practice to open scenario investigations, with adaptive hints.
4. **Mobile-first ergonomics:** Content is decomposed into ≤ 5 -minute micro-lessons; interactions are optimized for one-handed phone use; offline mode is supported for low-connectivity contexts.
5. **Pedagogical transparency:** Each module declares its learning outcomes, estimated time, and prerequisite modules, supporting self-regulated learning.

3.2 System architecture

DFLearn-Mobile is implemented as a three-tier cross-platform mobile application:

- Front-end (Mobile Client): Flutter (Dart), targeting Android (API 24+) and iOS (13+). State management uses Riverpod.
- Back-end (API): Node.js with Express, exposing a RESTful API for content delivery, progress tracking, and assessment scoring.
- Data: MongoDB stores module content, user profiles, progress records, and event logs. Sensitive credentials are hashed (bcrypt), and JWT is used for session authentication.
- Analytics: event logs are aggregated to a dashboard (administrator view) for instructor monitoring.

A high-level architecture diagram is shown in Figure 1.

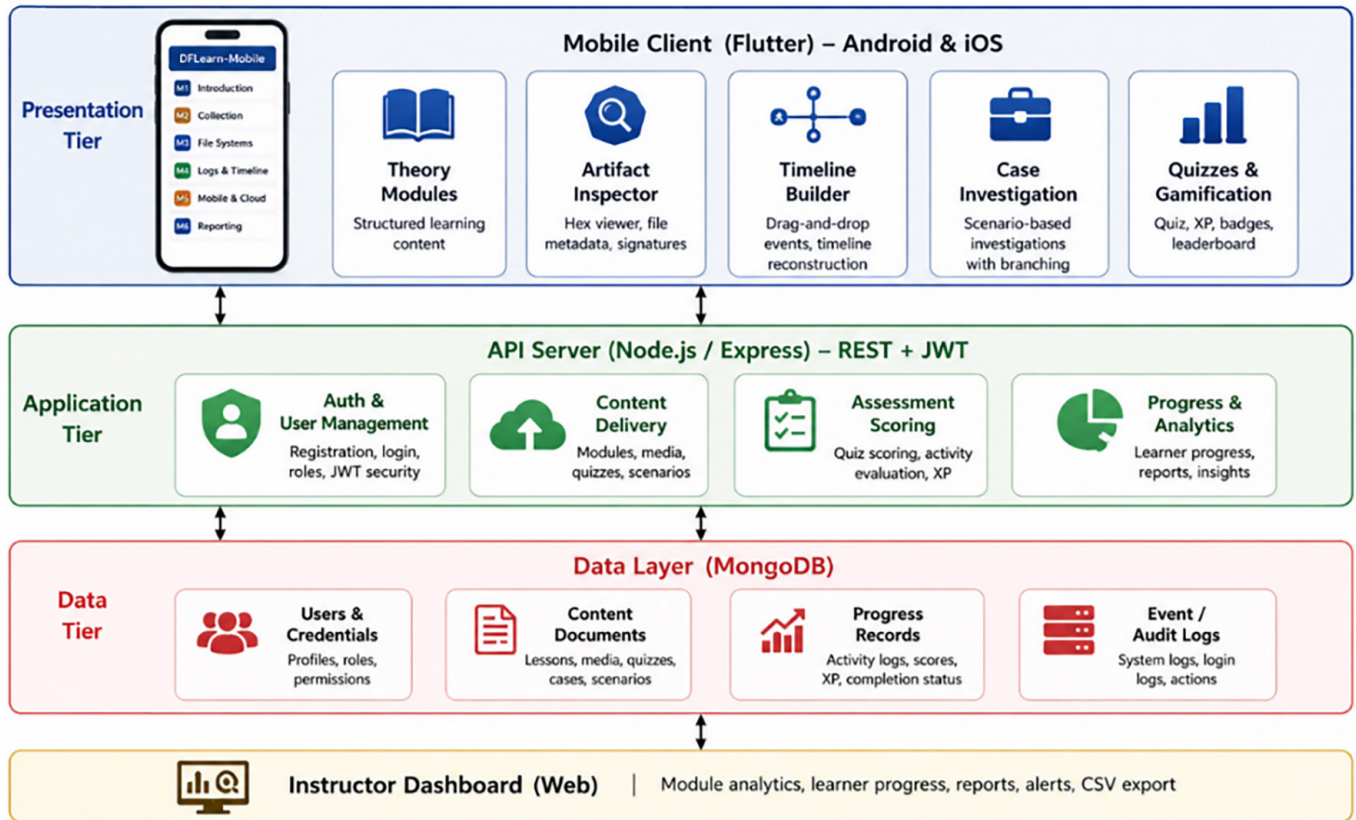


Fig. 1. DFlearn-mobile system architecture (three-tier)

3.3 Learning module design

The curriculum is organized into six modules aligned with the NIST process model and adapted for mobile delivery, as summarized in Table 2.

Table 2. Learning modules, interactive activities, and estimated completion times

Module	Topic	Interactive Activity	Est. Time
M1	Introduction to Digital Forensics	Concept-mapping drag-and-drop	10 min
M2	Identification & Collection	Chain-of-custody form completion simulation	15 min
M3	Examination — File Systems & Signatures	Hex/magic-number inspection mini-tool	20 min
M4	Analysis — Logs & Timeline	Drag-to-order timeline reconstruction	20 min
M5	Mobile & Cloud Artefacts	Artefact-source matching game	15 min
M6	Reporting & Chain of Custody	Guided report-drafting scaffold	20 min

Each module closes with a 5–10 item formative quiz and unlocks a corresponding case-investigation challenge in the gamified track. The mapping of curriculum modules to the NIST process model is shown in Figure 2.

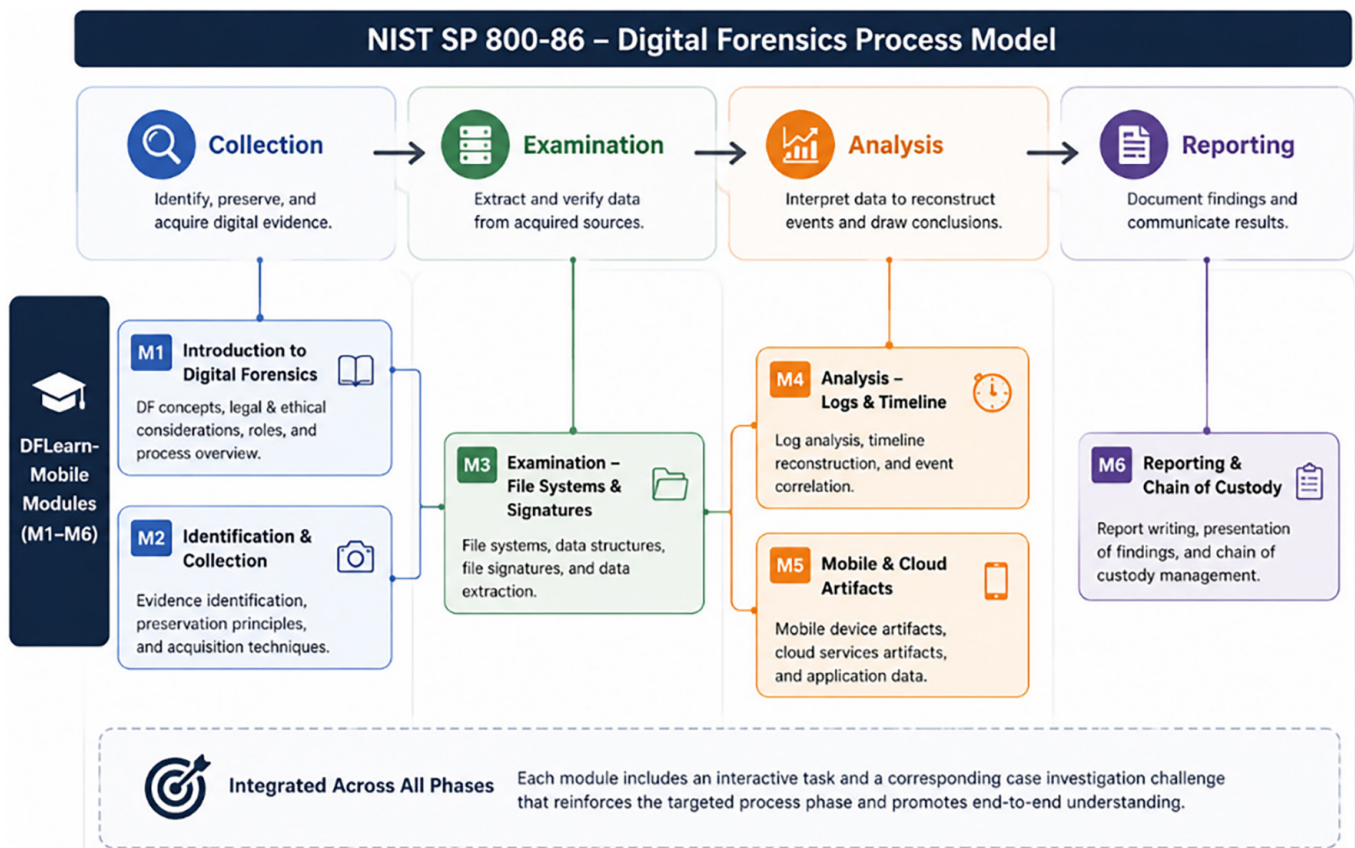


Fig. 2. Mapping of DFLearn-Mobile curriculum modules (M1–M6) to the NIST DF process model

3.4 Interactive elements

Four interaction modalities are implemented:

1. Artefact Inspector: A built-in viewer for sample artefacts (hex, EXIF, log lines), allowing learners to “tap to reveal” metadata fields.
2. Timeline Builder: Drag-and-drop event reconstruction, scored against a reference timeline.
3. Case Investigations: Branching scenario narratives with decision points, scored on procedural correctness (not just final answer).
4. Knowledge Quizzes: Spaced-repetition multiple-choice and short-answer items, with explanatory feedback.

Representative interface screens and the in-module learning flow are illustrated in Figure 3.

Beyond the entry-level UI, Figure 3c–e illustrates the learning flow within a module. Each lesson is delivered as a sequence of micro-cards (≤ 200 words) with a short body, an optional illustration, and a highlighted key term (panel c). When the learner reaches the end of the card sequence, a formative quiz item is presented; selected answers receive immediate explanatory feedback, including the underlying rationale (panel d). On successful module completion, the learner is shown an aggregate summary—quiz score, time on task, XP earned, and any newly unlocked badge—that closes the module loop and prompts the next module (panel e). This loop design is intended to consolidate retrieval practice and reinforce mastery before progressing.

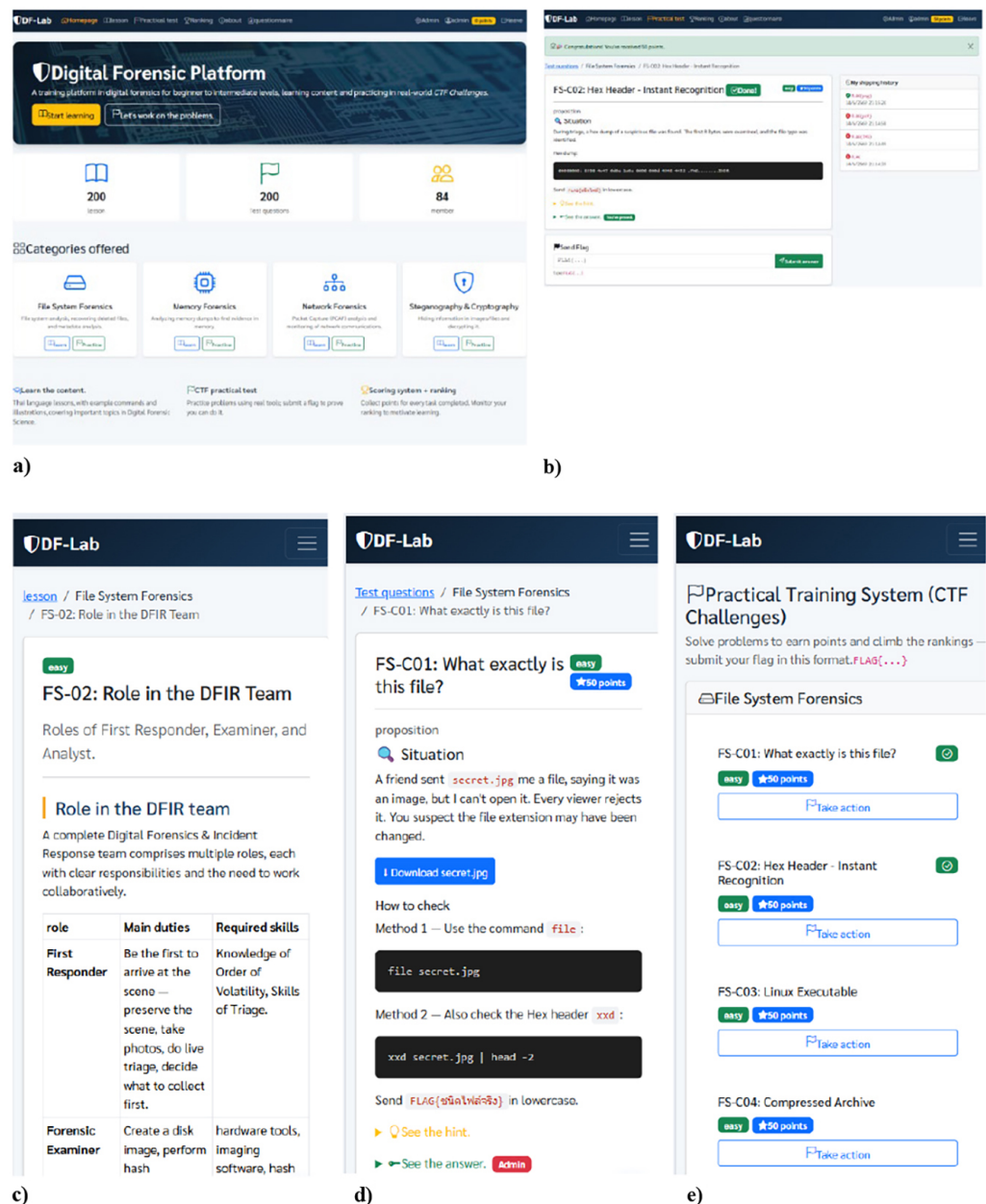


Fig. 3. DFLearn-mobile interface and learning flow (screenshots captured from the operational application): (a) module home; (b) artefact inspector and timeline builder; (c) lesson card with theory text and key-term highlight; (d) formative quiz item with immediate explanatory feedback; (e) module-completion screen showing XP earned and badge unlocked

Note: Figures 3–4 are screenshots captured from the operational DFLearn-Mobile application.

Figure 4a–c presents three of the platform’s interactive task interfaces drawn from different modules. The Process Order Game in M1 (panel a) asks the learner to drag eight scenario cards into the correct NIST phase column, with a running score and adaptive hints. The Chain-of-Custody Simulator in M2 (panel b) guides the learner through a 12-field CoC form using a clue panel populated from the scenario; fields turn green as they are completed correctly, and adaptive hints appear if a field is left blank beyond a short threshold. The Artefact Source Matcher in M5 (panel c) is a drag-line matching activity that pairs ten common mobile/cloud artefacts with their underlying data sources. All three tasks share a common scoring contract—partial credit, error-tolerant feedback, and a single “Submit” affordance—to keep the interaction predictable across modules.

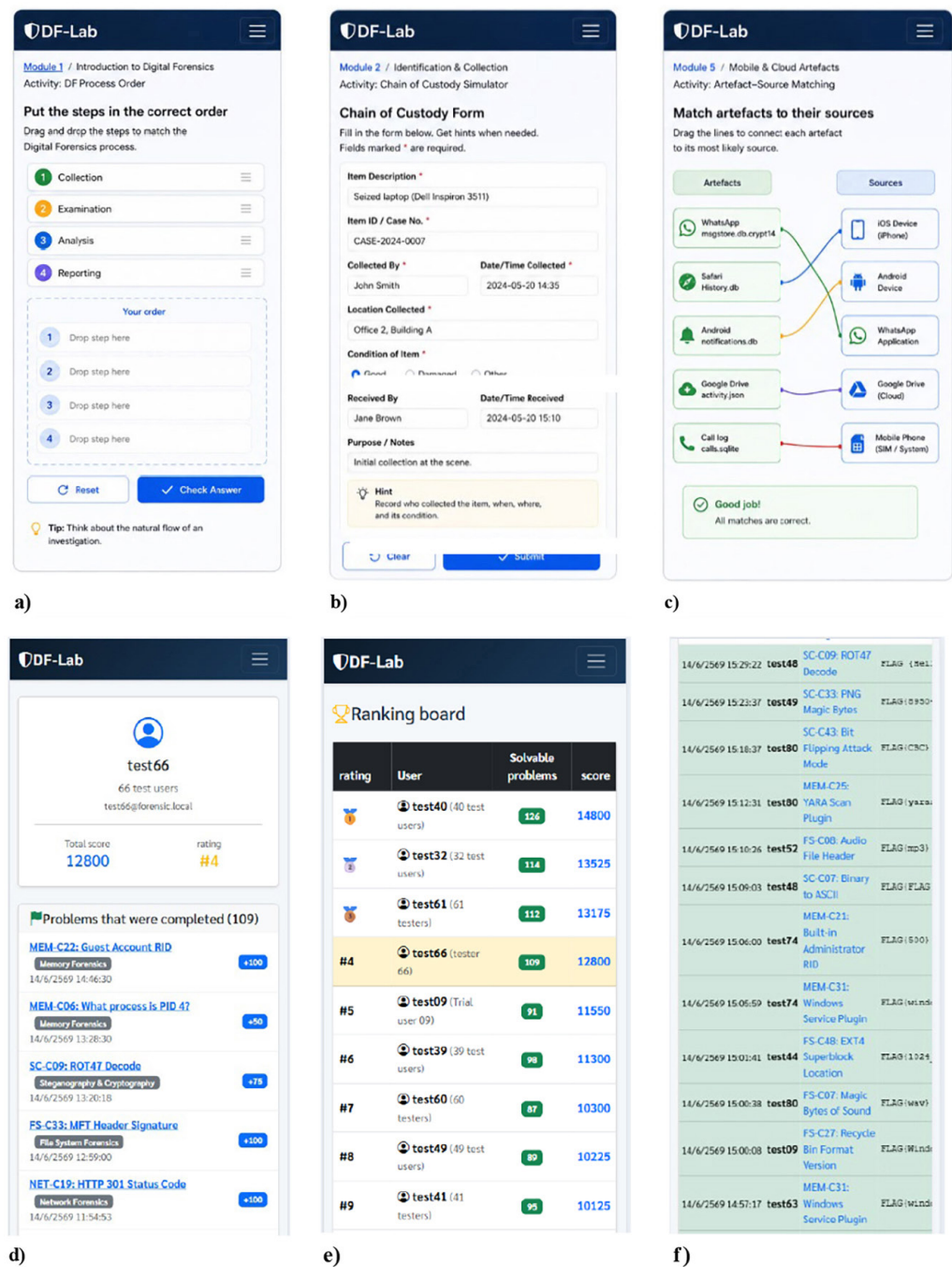


Fig. 4. Interactive tasks and gamification layer (screenshots captured from the operational application): (a) drag-and-drop process-order game (M1); (b) chain-of-custody form simulator with adaptive hints (M2); (c) drag-line artefact–source matching game (M5); (d) learner profile showing XP, badges, and per-module progress; (e) opt-in anonymized class leaderboard; (f) capstone branching-scenario case investigation

3.5 Gamification layer

A lightweight gamification layer awards experience points (XP), badges (e.g., “First Responder,” “Timeline Sleuth”), and a non-public class leaderboard. Gamification follows self-determination theory principles [22]—emphasizing competence and autonomy rather than purely extrinsic reward—and can be disabled per user. Figure 4d–f illustrates the three surfaces that constitute this layer. The profile screen

(panel d) summarizes the learner's current level, XP toward the next level, badges earned, and per-module completion bars—supporting self-regulated learning by making progress visible. The class leaderboard (panel e) is opt-in and shows anonymized display names, with the learner's own row highlighted in context; this is intended to provide social comparison without exposing identity, and learners may opt out at any time. Finally, the capstone case investigation (panel f) presents a branching scenario in which the learner makes procedural choices (e.g., whether to pull the power cable, photograph the screen, or close an application first); the running “procedural score” rewards correct DF practice rather than only the final answer, encouraging deliberate reasoning about evidence integrity.

3.6 Accessibility and localization

The interface supports Thai and English; font scaling, high-contrast mode, and screen-reader labels are implemented to align with WCAG 2.1 AA where feasible on mobile.

3.7 Administrator dashboard and usage analytics

DFLearn-Mobile records pseudonymized event logs (lesson views, quiz attempts, interactive-task submissions, badge unlocks, session start/stop) and aggregates them into a web-based administrator dashboard accessible to instructors after role-based login. Five read-only views cover cohort engagement (daily/weekly active users and enrolment), module progress (completion rates, time on task, drop-off points), assessment performance (item difficulty and discrimination), interactive-task analytics (procedural scores and most-frequent error patterns), and engagement and retention (session counts, badge distribution, retention curve, opt-in leaderboard). All views are exportable to CSV for offline analysis; personally identifying information is restricted to the secure user database and is never exposed in the dashboard. These analytics support instructor monitoring, formative feedback, iterative refinement of module content, and the post-hoc evaluation reported in Section 5.5.

4 METHODOLOGY

This study is framed as an educational intervention study that evaluates learning outcomes, usability, and learner acceptance associated with DFLearn-Mobile in an authentic undergraduate setting. A quasi-experimental quantitative design was adopted, combining pre- and post-test measurement of DF knowledge with standardized usability (SUS), TAM, and platform analytics.

4.1 Research questions

The study addresses three research questions:

1. RQ1: Are DF knowledge scores significantly higher after learners use DFLearn-Mobile than before use?
2. RQ2: How do undergraduate users perceive the usability of DFLearn-Mobile, as measured by the System usability scale?
3. RQ3: What are users' perceptions of usefulness, ease of use, attitude, and behavioral intention to use, as measured by an adapted TAM instrument?

4.2 Participants

Participants were undergraduate students enrolled in cybersecurity-related courses at the Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi (RMUTSB), across multiple campuses. Recruitment used voluntary purposive sampling, with the following inclusion criteria: (i) currently enrolled undergraduate, (ii) owns and is able to use an Android or iOS smartphone, and (iii) provides written informed consent. Participants received no financial compensation. A total of 92 students were initially enrolled; 12 were excluded from analysis for failing to complete all six modules ($n = 8$) or failing to submit post-test responses ($n = 4$), yielding the final analytic sample of $N = 80$. The target sample was $n \approx 60\text{--}120$ for the quantitative components. Baseline records for the 12 excluded participants were not retained in the analysis dataset, so baseline comparisons between included and excluded participants could not be performed; complete-case inclusion may therefore have produced upwardly selected estimates of knowledge gain, engagement, usability, and acceptance, and the results should not be read as representative of all 92 initially enrolled students (see Section 6.6). Demographics—gender, age, year of study, programmed, prior DF exposure, and smartphone OS—were collected at baseline and are reported in Section 5.1. The study was conducted under institutional ethical clearance granted by the Research Ethics Committee of Rajamangala University of Technology Suvarnabhumi (approval reference: RMUTSB-REC-2025-CS-01). All participants provided written informed consent prior to enrolment. Participation was voluntary with no penalty for withdrawal, and all platform data were pseudonymized before analysis. The class leaderboard was active during the study period; participants were informed of this at the time of consent and could opt out without penalty. No personally identifying information was retained beyond the study period.

4.3 Experimental flow

The intervention followed a four-week timeline, summarized in Figure 5.



Fig. 5. Evaluation procedure for DFLearn-Mobile (pre- and post-test, SUS, TAM, and platform analytics)

- **Week 1: Recruitment, consent, baseline.** Eligible students were briefed in class and through institutional channels. Those who consented completed a baseline questionnaire and the DF knowledge pre-test (10 items) via an online form.
- **Weeks 2–3: Intervention.** Participants used DFLearn-Mobile on their personal smartphones at their own pace. The minimum usage threshold for inclusion in the analysis was completion of all six modules (M1–M6), corresponding to approximately 100 minutes of platform time. Platform analytics (module completion, time on task, quiz scores) were logged server-side and used to verify adherence.
- **Week 4: Post-test and questionnaires.** Immediately after the intervention, participants completed (i) the DF knowledge post-test (same item domain as pre-test, items reordered to mitigate practice effects; because identical items were used, a residual practice effect cannot be fully excluded and is treated as a limitation), (ii) the System usability scale, and (iii) the adapted TAM questionnaire.

4.4 Instruments

- **DF Knowledge Test (10 items):** Multiple-choice items covering the six modules (M1: Introduction; M2: Identification & Collection; M3: File Systems & Signatures; M4: Logs & Timeline; M5: Mobile & Cloud Artefacts; M6: Reporting & Chain of Custody). Items were developed by the research team grounded in NIST SP 800-86 [1] and Carrier [6] and reviewed by two external DF instructors using the Item Objective Congruence (IOC) index, with items retained at $\text{IOC} \geq 0.50$.
- **System usability scale (SUS, 10 items):** The standard instrument [20], administered in Thai and English. The Thai translation was back-translated and pilot-tested for equivalence.
- **Adapted TAM Questionnaire:** An 18-item pool (Perceived Usefulness and Perceived Ease of Use, five items each; Attitude Toward Use and Behavioral Intention to Use, four items each) was adapted from Davis [18] and Venkatesh et al. [19], translated into Thai and independently back-translated to English, and reviewed by three-to-five subject experts using the Item-Objective Congruence (IOC) index; items with $\text{IOC} \geq 0.50$ were retained, yielding a 15-item instrument across four constructs—Perceived Usefulness (four items), Perceived Ease of Use (four items), Attitude Toward Use (four items), and behavioral Intention to Use (three items). A 5-point Likert scale (1 = strongly disagree, 5 = strongly agree) was used, and the full item list is provided in the Supplementary TAM Appendix. Item-level responses were aggregated to construct means for the present analysis, so construct-level internal-consistency (Cronbach's α) could not be recomputed here (see Section 6.6).

4.5 Statistical analysis

Distributional normality of interval-scale outcomes was examined with the Shapiro–Wilk test and variance homogeneity with Levene's test; where normality was rejected, the corresponding non-parametric procedure was used. For RQ1, the Wilcoxon signed-rank test served as the primary pre/post inferential test, with the matched-pairs rank-biserial correlation and Hake's average normalized gain $\langle g \rangle$ —calculated as the mean of individual learner gains $g_i = (\text{post}_i - \text{pre}_i) / (10 - \text{pre}_i)$ —as effect-size indices; exploratory subgroup comparisons used the Kruskal–Wallis H-test (ϵ^2) and an independent-samples t-test (Cohen's d) with a Mann–Whitney U

sensitivity analysis. For RQ2, the overall SUS score (Brooke's scoring procedure) was compared against the canonical benchmark of 68 [20] with a one-sample t-test. For RQ3, because the TAM construct distributions were non-normal, Spearman's rho correlations were reported among PU, PEOU, ATU, and BI. All analyses were performed in Python (pandas, SciPy, pingouin, statsmodels) and re-verified with an independent reimplementaion; significance was set a priori at $\alpha = 0.05$ (two-tailed), with Bonferroni correction applied to families of follow-up comparisons. Full computational details—exact and tie-corrected Wilcoxon p-value procedures, tie handling, z statistics, and confidence-interval conventions—are provided in the supplementary statistical appendix.

5 RESULTS

5.1 Participant demographics

A total of $N = 80$ undergraduate students from the Faculty of Science and Technology, RMUTSB, met the inclusion criteria and completed the full intervention (pre-test, platform use, post-test, SUS, and TAM). Demographic characteristics are summarized in Table 3.

Table 3. Participant demographics ($N = 80$)

Variable	Category	<i>n</i>	%
Sex	Male	45	56.3
	Female	31	38.8
	Other/prefer not to say	4	5.0
Age (years)	$M = 21.18, SD = 1.37$, range 19–24	–	–
Year of study	Year 2	28	35.0
	Year 3	22	27.5
	Year 4	30	37.5
Smartphone OS	Android	49	61.3
	iOS	25	31.3
	Other	6	7.5
Prior DF exposure	None	62	77.5
	Some (workshop/course)	18	22.5

Participants were predominantly male (56.3%) and Android users (61.3%). The mean age was 21.18 years ($SD = 1.37$). Most learners reported no prior exposure to digital forensics (77.5%).

5.2 Knowledge improvement (RQ1)

Shapiro-Wilk testing indicated non-normal distributions for the pre-test, post-test, and paired gain scores, including the gain-score distribution relevant to the paired-samples t-test assumption. Therefore, following the final statistical-analysis plan,

the Wilcoxon signed-rank test was used as the primary inferential test for the pre/post comparison. Descriptive, normality, and inferential results are summarized in Table 4.

Table 4. Pre- and post-test DF knowledge scores: descriptive, normality, and inferential statistics ($N = 80$)

Panel A—Descriptive Statistics and Normality							
Measure	N	Mean	SD	Median	Min-Max	Shapiro-Wilk W	p
Pre-test (0–10)	80	3.06	1.06	3.00	1–6	0.859	<.001
Post-test (0–10)	80	6.99	1.37	7.00	4–10	0.929	<.001
Gain (Post – Pre)	80	3.92	0.61	4.00	3–5	0.769	<.001
Panel B—Inferential Statistics							
Comparison	Test		Statistic	df	p	Effect Size	95% CI
Post vs. Pre	Wilcoxon signed-rank		V = 3,240	–	<.001	rank-biserial r = 1.00	gain: [3.79, 4.06]
Post vs. Pre	Hake normalised gain		<g> = 0.586	–	–	mean of individual gains; medium band	–
Post by Year of study	Kruskal-Wallis H-test		H = 2.16	2	.340	ε² = 0.002	–
Post by Prior DF exposure	Independent-samples t-test (Levene p = .381)		t = –7.38	78	<.001	Cohen’s d = 1.98	d: [1.37, 2.59]
Post by Prior DF exposure	Mann–Whitney U sensitivity test		U = 109	–	<.001	rank-biserial = 0.805	–

The Wilcoxon signed-rank test indicated that post-test DF knowledge scores were significantly higher than pre-test scores ($V = 3,240$, where V is the sum of positive ranks, $W+$; $z = 7.77$, $p < .001$, matched-pairs rank-biserial $r = 1.00$, 95% CI for the raw gain [3.79, 4.06]). The rank-biserial effect is maximized because all participants showed a positive raw gain; given the short 10-item instrument and the narrow gain-score variance, this value should be interpreted mainly as evidence that gains were consistently positive rather than as a conventional learning-effect magnitude. Hake's average normalized gain, calculated as the mean of individual learner normalized gains, was $\langle g \rangle = 0.586$ (medium band) and provides the more interpretable index of learning change. The exploratory comparison of post-test scores by year of study was not significant (Kruskal–Wallis $H(2) = 2.16$, $p = .340$, $\epsilon^2 = 0.002$). Learners with some prior DF exposure ($n = 18$) had higher post-test scores ($M = 8.61$, $SD = 1.14$) than those with no prior exposure ($n = 62$, $M = 6.52$, $SD = 1.04$). Levene's test supported variance homogeneity ($F = 0.78$, $p = .381$), and the independent-samples t-test was significant, $t(78) = -7.38$, $p < .001$, Cohen's $d = 1.98$. A Mann–Whitney U sensitivity analysis supported the same conclusion ($U = 109$, $z = -5.33$, $p < .001$; rank-biserial correlation = 0.805). This strong prior-exposure association should be controlled in future causal comparisons; it does not by itself establish that the platform caused the within-participant increase.

5.3 Usability—SUS (RQ2)

The overall SUS score, computed using Brooke's standard procedure, was $M = 61.03$ ($SD = 12.76$, median = 60.00, range 35.0–97.5; Shapiro-Wilk $W = 0.977$, $p = .161$). This places the platform in the OK/Marginal range relative to commonly

used SUS interpretation benchmarks, on which 68 is treated as the average score and adjective ratings of “Good” and “Excellent” correspond to approximately 73 and 80.3, respectively [34]. A one-sample t-test against the commonly used benchmark of 68 was significant in the negative direction ($t(79) = -4.89$, $p < .001$), confirming that mean SUS was reliably below this benchmark. For descriptive completeness, 11 of 80 learners (13.8%; verified against the dataset) returned $SUS \geq 73$ (“Good” or higher); this subgroup figure is reported descriptively and does not offset the overall result, which places mean usability in the OK/Marginal range below the benchmark. The SUS finding is discussed alongside the TAM ease-of-use score and the future-work agenda in Section 6.

5.4 Technology acceptance—TAM (RQ3)

Descriptive statistics and inter-construct Spearman correlations for the TAM constructs are presented in Table 5. The TAM instrument used here is a multi-item Likert questionnaire (1 = strongly disagree, 5 = strongly agree); the present dataset records construct-level means per learner (each construct mean is computed across its 3–4 underlying items by the survey platform).

Table 5. TAM constructs: descriptive statistics and inter-construct Spearman correlations ($N = 80$)

Panel A—Descriptive Statistics and Normality						
Construct	k Items	Mean	SD	Median	Min-Max	Shapiro-Wilk p
Perceived Usefulness (PU)	4	3.95	0.36	4.00	3.25–5.00	.002
Perceived Ease of Use (PEoU)	4	3.85	0.37	3.75	3.25–4.75	.001
Attitude Toward Use (ATU)	4	3.95	0.41	4.00	3.00–5.00	<.001
Behavioural Intention to Use (BI)	3	3.90	0.36	4.00	3.33–5.00	<.001
Panel B—Inter-Construct Spearman Correlation Matrix						
	PU	PEoU	ATU	BI		
Perceived Usefulness (PU)	—					
Perceived Ease of Use (PEoU)	0.394***	—				
Attitude Toward Use (ATU)	0.329**	0.408***	—			
Behavioural Intention to Use (BI)	0.458***	0.356**	0.191	—		

Notes: Spearman correlations are reported because TAM construct distributions were non-normal. *** $p < .001$; ** $p < .01$ (two-tailed). All scores were measured on a 5-point Likert scale.

Five of the six pairwise Spearman correlations were statistically significant and positive, broadly compatible with several relationships expected under the TAM [18], [19]; these are bivariate associations and do not constitute a full measurement-model test. The strongest association was Perceived Usefulness $\leftrightarrow$ behavioral Intention ($\rho = 0.458$, $p < .001$), in line with the expectation that perceived usefulness relates to behavioral intention. Perceived ease of use correlated with attitude ($\rho = 0.408$, $p < .001$), PU ($\rho = 0.394$, $p < .001$), and BI ($\rho = 0.356$, $p = .001$), and attitude correlated with PU ($\rho = 0.329$, $p = .003$). The ATU-BI association was positive but not

statistically significant after reanalysis ($\rho = 0.191$, $p = .090$). The TAM construct means cluster between 3.85 and 3.95 on the 5-point scale—i.e., the platform is rated as mildly positive but not strongly endorsed, which is internally consistent with the OK/Marginal SUS finding in Section 5.3 and points to specific usability and content-fit areas for the next development iteration.

5.5 Platform analytics

Figure 6 reports the platform analytics extracted from the administrator dashboard described in Section 3.7. Across the $N = 80$ cohort, median lesson-card completion was 138.0 of the platform's ~200 individual lesson cards (range 72–200; ~69%). Here, “lesson-card completion” counts the number of individual learning activities completed across the platform—and is distinct from module completion, which every participant in the analytic sample achieved per the inclusion criteria (Sections 4.2–4.3). Mean cumulative platform score was 7,327.50 XP ($SD = 2,567.45$, range 1,950–14,800), mean challenges solved was 63.49 ($SD = 21.33$), mean total attempts was 190.64 ($SD = 26.23$), and mean per-learner interactive-task accuracy was 32.96% ($SD = 9.52$). The dashboard's interactive-task analytics view showed the lowest mean accuracy in the file-systems module (Module 3; 21.4%), suggesting that hex/magic-signature recognition was comparatively difficult for learners new to forensic artefacts; this pattern informed a content revision after the pilot.

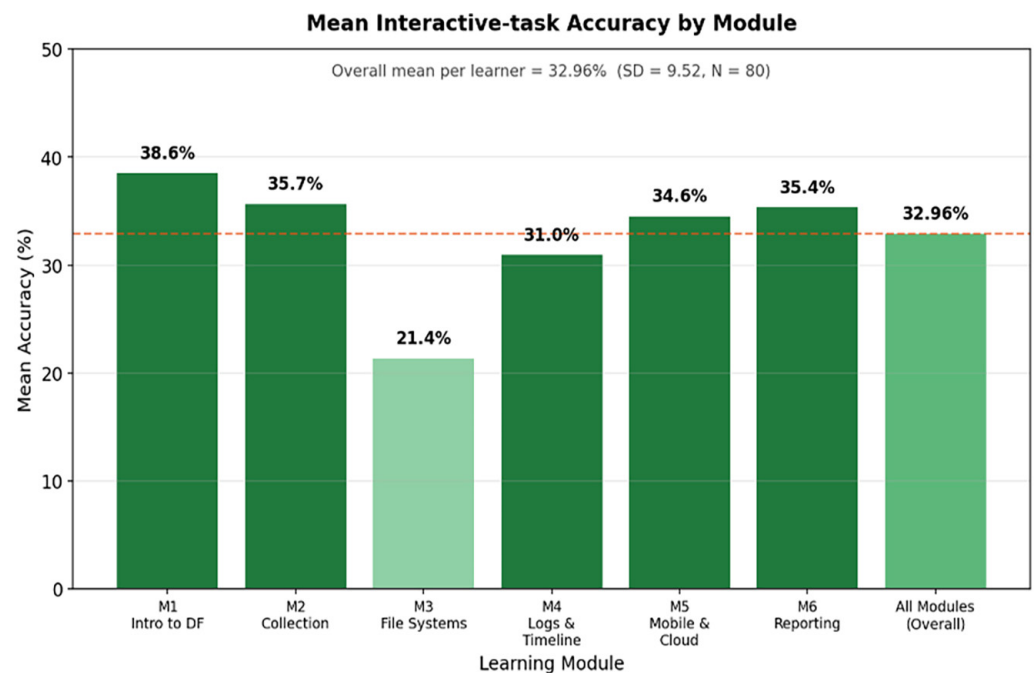


Fig. 6. Mean interactive-task accuracy (%) across the six learning modules

Notes: Unit of analysis = learner; per-learner accuracy = correct interactive-task submissions ÷ total submissions, averaged over $N = 80$ learners.

6 DISCUSSION

The Section 5 findings addresses the three identified gaps (domain, pedagogical, empirical) and the three research questions addressed by the pilot. We organized the

discussion around six themes that emerged when the pre/post, usability, acceptance, and platform-analytics results were read together.

6.1 Mobile reach and usability

Android dominated the cohort (61.3%), and platform analytics confirmed broad reach: median lesson-card completion was 138 of ~200 cards (~69%), with cumulative scores spanning 1,950–14,800 XP—supporting the design goal of mobile-first access in a resource-constrained, multi-campus setting where desktop DF labs and licensed tools [6], [16], [17] would have constrained participation. The usability result qualifies this optimistic reading: the OK/Marginal SUS outcome indicates a platform usable for a controlled pilot but not yet polished for the cohort as a whole. Three interface areas are likely contributors: (1) error-prone drag-and-drop interactions on small touch-screens, particularly in the Timeline Builder; (2) unclear progress indication between branching case steps; and (3) insufficiently prominent feedback panels after incorrect quiz responses. These hypotheses are consistent with the low interactive-task accuracy, particularly in Module 3, and will be the primary targets of the next UI iteration. For descriptive completeness, 13.8% of learners (11 of 80) returned $SUS \geq 73$; this subgroup figure is reported descriptively and does not offset the overall result.

6.2 Observed knowledge-score change

The pre/post comparison confirmed significantly higher post-test scores, and every participant improved by 3–5 points. Because the short 10-item test, reused/reordered items, and narrow gain-score variance maximize the rank-biserial statistic, the average normalized gain ($\langle g \rangle = 0.586$, medium band) is the more interpretable indicator. Year of study did not differentiate post-test performance ($H(2) = 2.16$, $p = .340$), but prior DF exposure did ($M = 8.61$ vs. 6.52 ; Cohen's $d = 1.98$), and the Mann–Whitney sensitivity analysis reached the same substantive conclusion. These findings indicate a promising pilot gain associated with platform use, not causal proof of effectiveness; a controlled comparison that balances or adjusts for prior DF exposure is required to isolate the platform-specific contribution.

6.3 Acceptance is mildly positive, driven by usefulness

Technology acceptance model construct means clustered between 3.85 and 3.95 on the 5-point scale—i.e., mildly positive but not strongly endorsed. The strongest association in the Spearman matrix was Perceived Usefulness $\leftrightarrow$ behavioral Intention ($\rho = 0.458$, $p < .001$), broadly compatible with the direction of the Davis [18] and UTAUT [19] formulations, though the present analysis reports bivariate associations rather than a measurement-model test. Perceived ease of use correlated with attitude ($\rho = 0.408$, $p < .001$) and PU ($\rho = 0.394$, $p < .001$), while the ATU-BI association was weak and not statistically significant after reanalysis ($\rho = 0.191$, $p = .090$). The acceptance picture is internally consistent with the OK/Marginal SUS finding: learners see the value of a mobile DF tool, but the friction of using it on real devices currently dampens enthusiasm. Self-determination theory [22] would predict that closing the usability gap (competence) and adding more agency-supporting customization (autonomy) will move both SUS and TAM constructs upward together.

6.4 Potential value of scenario-based tasks

Digital forensics is a procedural craft. The *order* of actions counts as much as the technical answer. The capstone branching scenario and the chain-of-custody simulator scaffold that reasoning by giving credit for procedural steps, not just final answers. The pre/post gain anchored in the NIST process model [1], together with the platform-analytics view of where learners struggle (the low-accuracy cluster in Module 3 file-system signature recognition), suggests that scenario-based tasks aligned with NIST phases may support the development of procedural DF competency beyond recall-focused activities. This design aligns with, and conceptually extends, earlier VR work on collection [23] across the full Collection–Examination–Analysis–Reporting cycle. This is a design interpretation rather than an experimentally isolated effect: the pilot did not separate scenario-based tasks, micro-learning, adaptive hints, or gamification, so their independent contributions to the observed score change cannot be established.

6.5 Micro-learning design and multi-campus replicability

Each module breaks into cards of five minutes or less. The platform is designed to support use in repeated short sessions rather than long blocks; session-frequency and duration were not analyzed in this pilot, so this remains a design intention to be verified in future work. Micro-learning [28] is therefore more than a UI preference—it is a structural choice that keeps technical content reachable in the gaps of a student's day [10], [11]. The Flutter/Node.js/MongoDB stack, bilingual content, offline mode, and the administrator dashboard (Section 3.7) together make the platform—and the analytic toolkit reported in Tables 2–4—portable to other Thai public universities with multi-campus, multi-faculty structures. The dashboard's per-section filter and CSV export are intended specifically to make the same analyses reported here repeatable at any partner institution. Beyond the Thai institutional context, three design principles extracted from this pilot have broader applicability: (1) aligning module content to an internationally recognized process model (NIST SP 800-86) provides a content scaffold that transfers to any institution teaching the same standard; (2) the four-instrument evaluation toolkit (pre/post knowledge test + SUS + TAM + platform analytics) is replicable with minimal adaptation at any university with mobile-capable students; and (3) the offline-capable, bilingual architecture reduces deployment barriers in low-connectivity, multilingual settings common across Southeast Asia, South Asia, and sub-Saharan Africa. Future multi-institution studies should test whether the learning gains and acceptance patterns observed here generalize beyond a single Thai university.

6.6 Limitations

Eight limitations qualify the present findings. First, the absence of a matched control group and random assignment means that the added value of DFLearn-Mobile cannot be isolated from maturation, repeated testing, or other contextual influences; the strong prior-DF-exposure association ($d = 1.98$) makes covariate control or stratification particularly important. Second, the knowledge test is short (10 items), which produces narrow gain variance and makes the rank-biserial effect appear maximized; the average normalized gain $\langle g \rangle$ should be treated as the

more interpretable index. Third, the post-test reused the same items in a reordered sequence, so a residual practice effect cannot be excluded; parallel forms are needed in future studies. Fourth, the SUS sits in the OK/Marginal band, indicating that the platform was usable enough for a controlled pilot but not yet polished enough for broader deployment without interface refinement. Fifth, the intervention lasted only four weeks, so novelty effects and short-term engagement may have influenced the results. Sixth, single-institution purposive recruitment limits generalizability to other institutions and learner populations. Seventh, the analysis was restricted to intervention completers (80 of 92 enrolled; 8 excluded for incomplete module completion and 4 for missing post-test responses); baseline records for the 12 excluded participants were not available in the retained analysis dataset, so a formal included-versus-excluded comparison could not be performed. Complete-case inclusion may therefore have produced an upwardly selected estimate of knowledge gain, engagement, usability, and acceptance, so the findings should not be read as representative of all 92 initially enrolled students. Eighth, item-level TAM responses were aggregated to construct means, so construct-level internal-consistency estimates (Cronbach's α) could not be recomputed; the reported inter-construct correlations are bivariate associations rather than a measurement-model validation.

6.7 Future work

Planned extensions include (i) a randomized between-subjects design with a matched control group receiving conventional instruction to support causal claims. (ii) a targeted UI/UX revision explicitly aimed at lifting SUS from the OK band into Above-Average (≥ 68), tracked via paired SUS measurement before and after the revision. (iii) a longer-form knowledge test (≥ 25 items) to improve measurement precision, reduce ceiling effects, and provide more stable effect-size estimates. (iv) augmented-reality artefact inspection for selected modules. (v) integration with a lightweight cyber-range for live evidence acquisition exercises. (vi) federated multi-institution evaluation across additional Thai and Southeast Asian universities. (vii) a longitudinal follow-up study measuring DF skill retention three to six months after the intervention.

7 CONCLUSION

This paper introduced *DFLearn-Mobile*, a mobile app for undergraduate DF education. The design rests on four learning theories—constructivist [25], experiential [26], situated [27], and microlearning [28]—and a six-module curriculum aligned with the NIST DF process model [1]. The app is built in Flutter on a Node.js back-end with MongoDB, and runs on Android and iOS.

We evaluated the app with $N = 80$ undergraduates at a Thai multi-campus university. DF knowledge scores were higher after platform use (Hake's average normalized gain, calculated across individual learners, $\langle g \rangle = 0.586$, medium band; Wilcoxon $V = 3,240$, $p < .001$). Prior DF exposure was strongly associated with post-test performance ($d = 1.98$), with the same conclusion supported by a Mann–Whitney sensitivity analysis ($U = 109$, $p < .001$). The SUS score ($M = 61.03$, $SD = 12.76$) sits in the OK/Marginal band—significantly below the 68 above-average benchmark—indicating that the platform was usable for this controlled pilot but requires focused interface refinement before broader deployment. TAM construct means clustered

between 3.85 and 3.95 (mildly positive acceptance), with PU–BI emerging as the strongest Spearman association ($\rho = 0.458$). Platform analytics showed broad cross-cohort engagement (median completion of 138 of ~200 lesson cards; mean interactive-task accuracy 32.96%) and flagged Module 3 file-system signature recognition as a recurring difficulty point.

The contributions span three categories. The design contribution is a mobile-first pedagogical framework that integrates constructivist, experiential, situated, and microlearning principles with gamification across the full NIST DF process model. The technical contribution is a working cross-platform implementation (Flutter/Node.js/MongoDB) with scenario-based NIST-aligned activities and an instrumented administrator dashboard. The empirical contribution is preliminary pilot evidence—not causal proof—from $N = 80$ undergraduates, accompanied by an honest diagnostic baseline showing where the platform succeeds (knowledge gain, broad engagement) and where it requires further work (usability, accuracy on complex tasks). *DFLearn-Mobile* offers a practical reference design—and a diagnostic baseline—for mobile DF teaching in resource-constrained, multi-campus higher education across Southeast Asia and similar contexts.

8 ACKNOWLEDGEMENTS

The authors thank the Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi, for institutional support.

9 REFERENCES

- [1] K. Kent, S. Chevalier, T. Grance, and H. Dang, *Guide to Integrating Forensic Techniques into Incident Response*. NIST Special Publication 800–86, National Institute of Standards and Technology, Gaithersburg, MD, USA, 2006. <https://doi.org/10.6028/NIST.SP.800-86>
- [2] E. Casey, *Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet*, 3rd ed. Waltham, MA, USA: Academic Press, 2011.
- [3] S. L. Garfinkel, “Digital forensics research: The next 10 years,” *Digital Investigation*, vol. 7, pp. 64–73, 2010. <https://doi.org/10.1016/j.diin.2010.05.009>
- [4] G. Horsman, “Tool testing and reliability issues in the field of digital forensics,” *Digital Investigation*, vol. 28, pp. 163–175, 2019.
- [5] A. Lang, M. Bashir, R. Campbell, and L. DeStefano, “Developing a new digital forensics curriculum,” *Digital Investigation*, vol. 11, pp. 76–84, 2014. <https://doi.org/10.1016/j.diin.2014.05.008>
- [6] B. Carrier, *File System Forensic Analysis*. Upper Saddle River, NJ, USA: Addison-Wesley, 2005.
- [7] I. Palmer, E. Wood, S. Nagy, G. Garcia, M. Bashir, and R. Campbell, “Digital forensics education: A multidisciplinary curriculum model,” in *Digital Forensics and Cyber Crime (ICDF2C 2015)*, LNICST, vol. 157, Springer, Charm, 2015, pp. 3–15. https://doi.org/10.1007/978-3-319-25512-5_1
- [8] H. Crompton and D. Burke, “The use of mobile learning in higher education: A systematic review,” *Computers & Education*, vol. 123, pp. 53–64, 2018. <https://doi.org/10.1016/j.compedu.2018.04.007>
- [9] Y. M. Huang, Y. T. Lin, and S. C. Cheng, “Effectiveness of a mobile plant learning system in a science curriculum in Taiwanese elementary education,” *Computers & Education*, vol. 54, no. 1, pp. 47–58, 2010. <https://doi.org/10.1016/j.compedu.2009.07.006>

- [10] W. H. Wu, Y. C. J. Wu, C. Y. Chen, H. Y. Kao, C. H. Lin, and S. H. Huang, "Review of trends from mobile learning studies: A meta-analysis," *Computers & Education*, vol. 59, no. 2, pp. 817–827, 2012. <https://doi.org/10.1016/j.compedu.2012.03.016>
- [11] Y. T. Sung, K. E. Chang, and T. C. Liu, "The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis," *Computers & Education*, vol. 94, pp. 252–275, 2016. <https://doi.org/10.1016/j.compedu.2015.11.008>
- [12] A. Tsohou, M. Karyda, and S. Kokolakis, "Analyzing the role of cognitive and cultural biases in the internalization of information security policies: Recommendations for information security awareness programs," *Computers & Security*, vol. 52, pp. 128–141, 2015. <https://doi.org/10.1016/j.cose.2015.04.006>
- [13] L. McDaniel, E. Talvi, and B. Hay, "Capture the flag as cyber security introduction," in *49th Hawaii Int. Conf. on System Sciences (HICSS)*, Koloa, HI, USA, 2016, pp. 5479–5486. <https://doi.org/10.1109/HICSS.2016.677>
- [14] F. Giannakas, G. Kambourakis, and S. Gritzalis, "CyberAware: A mobile game-based app for cybersecurity education and awareness," in *Int. Conf. on Interactive Mobile Communication Technologies and Learning (IMCL)*, Thessaloniki, Greece, 2015, pp. 54–58. <https://doi.org/10.1109/IMCTL.2015.7359553>
- [15] S. Karagiannis, E. Maragkos-Belmpas, and E. Magkos, "An analysis and evaluation of open source capture the flag platforms as cybersecurity e-learning tools," in *Information Security Education, Information Security in Action (WISE 2020)*, IFIP AICT, vol. 579, Springer, Charm, 2020, pp. 61–77. https://doi.org/10.1007/978-3-030-59291-2_5
- [16] T. Chothia and C. Novakovic, "An offline capture the flag-style virtual machine and an assessment of its value for cybersecurity education," in *USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE)*, Washington, DC, USA, 2015.
- [17] J. Vykopal, R. Ošlejšek, P. Čeleda, M. Vizváry, and D. Tovarňák, "KYPO cyber range: Design and use cases," in *12th Int. Conf. on Software Technologies (ICSOT)*, Madrid, Spain, 2017, pp. 310–321. <https://doi.org/10.5220/0006428203100321>
- [18] 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. <https://doi.org/10.2307/249008>
- [19] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS Quarterly*, vol. 27, no. 3, pp. 425–478, 2003. <https://doi.org/10.2307/30036540>
- [20] J. Brooke, "SUS: A 'quick and dirty' usability scale," in *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland, Eds., London, UK: Taylor & Francis, 1996, pp. 189–194.
- [21] D. L. Kirkpatrick and J. D. Kirkpatrick, *Evaluating Training Programs: The Four Levels*, 3rd ed. San Francisco, CA, USA: Berrett-Koehler, 2006.
- [22] E. L. Deci and R. M. Ryan, *Intrinsic Motivation and Self-Determination in Human Behavior*. New York, NY, USA: Plenum Press, 1985. <https://doi.org/10.1007/978-1-4899-2271-7>
- [23] C. Hassenfeldt, J. Jacques, and I. Baggili, "Exploring the learning efficacy of digital forensics concepts and bagging & tagging of digital devices in immersive virtual reality," *Forensic Science International: Digital Investigation*, vol. 33, p. 301011, 2020. <https://doi.org/10.1016/j.fsidi.2020.301011>
- [24] M. M. Yamin, B. Katt, and V. Gkioulos, "Cyber ranges and security testbeds: Scenarios, functions, tools and architecture," *Computers & Security*, vol. 88, p. 101636, 2020. <https://doi.org/10.1016/j.cose.2019.101636>
- [25] J. S. Bruner, *Toward a Theory of Instruction*. Cambridge, MA, USA: Harvard University Press, 1966.

- [26] D. A. Kolb, *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1984.
- [27] J. Lave and E. Wenger, *Situated Learning: Legitimate Peripheral Participation*. Cambridge, UK: Cambridge University Press, 1991. <https://doi.org/10.1017/CBO9780511815355>
- [28] T. Hug, M. Lindner, and P. A. Bruck, Eds., *Microlearning: Emerging Concepts, Practices and Technologies after E-Learning*. Innsbruck, Austria: Innsbruck University Press, 2005.
- [29] M. Sailer and L. Homner, "The gamification of learning: A meta-analysis," *Educational Psychology Review*, vol. 32, no. 1, pp. 77–112, 2020. <https://doi.org/10.1007/s10648-019-09498-w>
- [30] J. Krath, L. Schürmann, and H. F. O. von Korflesch, "Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning," *Computers in Human Behavior*, vol. 125, p. 106963, 2021. <https://doi.org/10.1016/j.chb.2021.106963>
- [31] S. Hart, A. Margheri, F. Paci, and V. Sassone, "Riskio: A serious game for cyber security awareness and education," *Computers & Security*, vol. 95, p. 101827, 2020. <https://doi.org/10.1016/j.cose.2020.101827>
- [32] A. D. Samala, S. A. Papadakis, and S. Rawas, "Global insights into mobile learning in higher education: A PRISMA-guided bibliometric analysis from 2007 to 2023," *International Journal of Educational Reform*, 2025. <https://doi.org/10.1177/10567879251341869>
- [33] S. Papadakis and T. Karakose, "Gamification and student achievement: Potential benefits, limitations, and effective use in educational environments," *Educational Process: International Journal*, vol. 19, Art. no. e2025529, 2025. <https://doi.org/10.22521/edupij.2025.19.529>
- [34] A. Bangor, P. T. Kortum, and J. A. Miller, "Determining what individual SUS scores mean: Adding an adjective rating scale," *Journal of Usability Studies*, vol. 4, no. 3, pp. 114–123, 2009.
- [35] F. Giannakas, G. Kambourakis, A. Papasalouros, and S. Gritzalis, "Security education and awareness for K-6 going mobile," *International Journal of Interactive Mobile Technologies (ijIM)*, vol. 10, no. 2, pp. 41–48, 2016. <https://doi.org/10.3991/ijim.v10i2.5473>
- [36] S. R. Sobral, "Mobile learning in higher education: A bibliometric review," *International Journal of Interactive Mobile Technologies (ijIM)*, vol. 14, no. 11, pp. 153–170, 2020. <https://doi.org/10.3991/ijim.v14i11.13973>
- [37] H. F. El-Sofany and N. El-Haggar, "The effectiveness of using mobile learning techniques to improve learning outcomes in higher education," *International Journal of Interactive Mobile Technologies (ijIM)*, vol. 14, no. 8, pp. 4–18, 2020. <https://doi.org/10.3991/ijim.v14i08.13125>
- [38] A. Ajlouni, A. AlOmary, M. Al-Smadi, M. Al-Abdallat, and F. Abu Awwad, "Factors influencing student satisfaction with LMS mobile application: A TAM-based SEM analysis," *International Journal of Interactive Mobile Technologies (ijIM)*, vol. 18, no. 19, pp. 106–128, 2024. <https://doi.org/10.3991/ijim.v18i19.50237>

10 SUPPLEMENTARY APPENDICES

10.1 Appendix A. Adapted technology acceptance model (TAM) instrument

Adaptation, translation, and item screening. An 18-item pool was adapted from Davis (1989) [18] and Venkatesh et al. (2003) [19] to the DFLearn-Mobile context. Items were prepared bilingually (English wording below); the Thai wording was produced by translation with independent back-translation to check equivalence. The pool was reviewed by three-to-five subject experts using the Item-Objective

Congruence (IOC) index, and items with $\text{IOC} \geq 0.50$ were retained. This screening yielded the final 15-item analysis instrument reported in the manuscript: Perceived Usefulness (4), Perceived Ease of Use (4), Attitude Toward Use (4), and Behavioural Intention to Use (3). All items used a 5-point Likert scale (1 = strongly disagree ... 5 = strongly agree). Item-level responses were aggregated to construct means for analysis; construct-level Cronbach's α could not be recomputed for the retained dataset (noted as a limitation, Section 6.6).

A. Final 15-item analysis instrument (retained after IOC screening)

Perceived Usefulness (PU)—4 items

PU1. Using DFLearn-Mobile improves my understanding of digital forensics concepts.

PU2. Using DFLearn-Mobile enhances my ability to identify and analyse digital evidence.

PU3. Using DFLearn-Mobile makes it easier for me to learn digital forensics.

PU4. Using DFLearn-Mobile increases my efficiency when studying digital forensics topics.

Perceived Ease of Use (PEoU)—4 items

PEoU1. Learning to use DFLearn-Mobile was easy for me.

PEoU2. I find it easy to get DFLearn-Mobile to do what I want it to do.

PEoU3. My interaction with DFLearn-Mobile is clear and understandable.

PEoU4. I find DFLearn-Mobile easy to use.

Attitude Toward Use (ATU)—4 items

ATU1. Using DFLearn-Mobile is a good idea.

ATU2. I enjoy using DFLearn-Mobile.

ATU3. I like the idea of learning digital forensics through a mobile application.

ATU4. Using DFLearn-Mobile makes learning more interesting.

Behavioural Intention to Use (BI)—3 items

BI1. I intend to use DFLearn-Mobile in the future for my studies.

BI2. I would recommend DFLearn-Mobile to other students.

BI3. I plan to use DFLearn-Mobile for review and self-study.

B. Items from the original pool not retained after IOC screening (3)

The following three items were part of the adapted pool but were not retained in the final analysis instrument:

PU5 (Perceived Usefulness). Overall, DFLearn-Mobile is useful for my learning.

PEoU5 (Perceived Ease of Use). It is easy for me to become skilful at using DFLearn-Mobile.

BI4 (Behavioural Intention to Use). Given the opportunity, I would continue using DFLearn-Mobile.

10.2 Scoring and validity notes

Per construct: mean and SD across the retained items. Content validity via expert IOC review (retain $\text{IOC} \geq 0.50$). Because the construct distributions were non-normal in the analysed sample, inter-construct associations were reported with Spearman's rho (not Pearson). The reported correlations are bivariate associations and do not constitute a full measurement-model test.

11 AUTHORS

Krommavut Nongnuch is a Lecturer at the Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi, Thailand. His research interests include digital forensics education, mobile learning, and applied cybersecurity (E-mail: krommavut.n@rmutsb.ac.th).

Saowalak Leelawongsarote is a Lecturer at the Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi, Thailand, with interests in educational technology and instructional design (E-mail: saowalak.l@rmutsb.ac.th).

Tawan Khunarsa is a Lecturer at the Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi, Thailand, with interests in information systems and mobile application development (E-mail: tawan.k@rmutsb.ac.th).

Suthin Kasetratanachai is a Lecturer at the Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi, Thailand, with interests in network security and cybersecurity education (E-mail: suthin.k@rmutsb.ac.th).