The University Journal: A Doctoral Association of Eastern Africa (DAEA) Publication
Vol 8(2 Sup. Issue II)

ISSN: 2519-0997 (Print)

Social Engineering, Multidimensional Harm, and Sustained Safe Continuance Among Kenyan Social Media Users: An Integrated Framework

Julius Sirma[1]*, Patrick K. Wamunyu[2], and Dalton Ndirangu[3]
[1,2,3] USIU-Africa, Nairobi, Kenya
*Email: julius.sirma@gmail.com**

***Corresponding author**

Cite: Sirma, J., Wamunyu, P.K., & Ndirangu, D. (2026). Social Engineering, Multidimensional Harm, and Sustained Safe Continuance among Kenyan Social Media Users: An Integrated Framework. *The University Journal, 8*(2 Sup. Issue II), 49-57.

Abstract

Social engineering has become the dominant mode of attack against social media users, yet its consequences are typically studied one harm at a time and rarely connected to the protective behaviors through which users sustain safe participation. This study developed and empirically validated an integrated framework linking motivational engagement, exposure to social engineering, four distinct harms, protective capability, and sustained safe continuance among Kenyan social media users. A convergent parallel mixed-methods design combined a cross-sectional survey of 304 users with ten focus-group discussions. Survey data were analyzed through structural equation modelling, and focus-group data through reflexive thematic analysis. Of fifteen hypothesized paths, eleven were supported. Social motivations, but not technological factors, significantly drove engagement, which in turn increased exposure. Exposure produced all four harms, binding most strongly to financial harm, then psychological harm, digital wellbeing, and platform trust. Protective behavior formed an ordered progression from foundational hygiene through proactive defense and advanced protection to recovery and help-seeking. Sustained safe continuance was supported by foundational hygiene rather than advanced protection. The qualitative findings corroborated these patterns and explained the limited contribution of help-seeking by reference to the inadequacy of formal support. The framework advances a differentiated, developmental, and contextually grounded account of how users experience and resist social engineering in a mobile-money environment.

Keywords: Social engineering, Digital wellbeing, Financial harm, Cyber resilience, Sustained safe continuance

Introduction

Social media has become central to communication, commerce, and social life in Kenya, where platforms such as WhatsApp, TikTok, and YouTube mediate everyday interaction for a predominantly young and mobile-connected population. The same platforms have become the principal arena for social engineering, the manipulation of users, rather than of systems, into disclosing information or transferring value. Because social engineering exploits ordinary social trust rather than technical vulnerability, every engaged user is a potential target, and the intensity of participation that makes social media valuable is precisely what exposes users to deception.

The Kenyan context sharpens the stakes. The deep integration of social media with mobile-money services means that the boundary between social interaction and financial transaction is substantially dissolved, so that a deception succeeding in the social channel can convert directly and immediately into financial loss. This convergence, together with the economic precarity of many young users, positions financial harm as a potentially dominant consequence of social engineering in the region.

Existing research, however, has three limitations. First, the consequences of social engineering, financial, psychological, wellbeing-related, and trust-related, have largely been examined in isolation, so their relative magnitudes remain unestablished within a single population. Second, protective and resilience behaviors have been studied as parallel practices rather than as a structured, developmental capability. Third, the link between protective behavior, residual harm, and whether users continue to participate safely has rarely been modelled. Whereas a prior synthesis of the literature characterized the human, technological, and institutional factors that render Kenyan users vulnerable to social engineering (Sirma et al., 2026), the present study advances from that descriptive account to an empirical one, developing and validating an integrated framework and estimating the relationships among its constructs directly. It addresses the gaps identified above through three objectives: to examine how motivations drive the engagement that exposes users to social engineering and the differentiated harm that follows; to establish whether protective behaviors form an ordered progression of capability; and to develop and validate an integrated framework connecting these to sustained safe continuance.

Theoretical and Empirical Background

The framework integrates three complementary traditions. The uses and gratifications and technology acceptance tradition explain why users engage with social media, with social and relational motivations shown consistently to predict intensive use, often outweighing purely technological factors (Al-Azawei, 2018; Tarhini et al., 2021). Protection motivation and threat-avoidance theory explain how users appraise threats and adopt protective responses, and locates the exposure that engagement produces within ordinary social cognition rather than ignorance (Frauenstein, 2019; Yeng et al., 2022). The resilience and continuance tradition explains how users maintain participation under conditions of risk, defining continuance as conditioned on the ongoing management of that risk rather than on initial adoption (Lin et al., 2014; Goliath et al., 2026).

Empirically, higher engagement has been linked to greater exposure to deceptive content, particularly where engagement is socially motivated (Saleem et al., 2024; Liu & Li, 2025), and the advance of artificial intelligence has made that exposure more sophisticated (Kaur et al., 2024; Pushpavalli et al., 2026). The financial dimension of harm has been documented in studies of investment and romance fraud and of scam susceptibility under economic stress (Abubakari & Oseh-Ovarah, 2025; Burruss et al., 2026; Suet Ching et al., 2026), while protective behavior has been shown to depend on awareness and to develop cumulatively from basic to advanced practice (Hasan et al., 2025; Ngo et al., 2026; Naeem & Mushibwe, 2025).

Prior frameworks have, however, tended to address a single segment of this pathway. Technology-acceptance and uses-and-gratifications models explain why users engage but do not extend to the security consequences of engagement; protection motivation and threat-avoidance models explain the formation of a single protective response but do not model protective behavior as an ordered

capability or connect it to continuance; and human-factors accounts theorize susceptibility without estimating the differentiated harm it produces. The present framework departs from these approaches in three respects: it integrates the segments that prior frameworks addressed separately, so that the output of each stage is the input of the next; it differentiates the harm that exposure produces, estimating the relative magnitudes of four distinct consequences within one model; and it models protective behavior as a developmental progression connected to sustained safe continuance. The framework's contribution lies not in the novelty of any single construct but in the integration, differentiation, and connection that its structure achieves, and in its empirical demonstration within a Kenyan mobile-money setting.

Methodology

The study adopted a pragmatist philosophy and a convergent parallel mixed-methods design, in which quantitative and qualitative strands were conducted concurrently and integrated at the interpretation stage. The design was appropriate because the framework expressed its constructs as latent variables and its propositions as directional relationships suited to structural analysis, while the lived experience underlying those constructs required qualitative inquiry to interpret and corroborate.

The quantitative strand surveyed 304 social media users. The sample was predominantly female (226; 74.3%) and young, with 278 respondents (91.4%) aged 18–24 years; most held or were pursuing an undergraduate degree (170; 55.9%) or a college diploma (99; 32.6%). Platform use, which was multi-select, was dominated by WhatsApp (86.2%), TikTok (58.9%), and YouTube (44.1%). The individual social media user was the unit of analysis. Because no enumerable sampling frame exists for this population, non-probability (purposive and convenience) sampling was used, and the achieved sample satisfied the requirements of structural equation modelling, exceeding both the minimum of 200 cases and a ratio of ten cases per indicator.

The qualitative strand comprised ten focus-group discussions of approximately ten participants each (about 100 participants in total), drawn purposively from two populations of active users: church congregations and university students. These populations were selected to capture a range of ages, occupations, and social contexts within the target population. Discussion continued across the ten groups until no substantially new themes emerged, indicating thematic saturation.

Quantitative data were collected through a structured questionnaire of fifty-one items rated on a five-point Likert scale, operationalizing the framework's thirteen constructs. Items were derived from the validated construct definitions established in the literature review, supporting content validity. Qualitative data were collected through a focus-group discussion guide addressing participants' encounters with social engineering, its consequences, and their protective and help-seeking responses. Focus-group data were documented as aggregated thematic records rather than verbatim transcripts.

Survey data were analyzed using a reproducible computational pipeline that executed descriptive analysis, assessment of factorability through the Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test, exploratory factor analysis with oblique (Promax) rotation, reliability and validity assessment (Cronbach's alpha, composite reliability, and average variance extracted), and covariance-based structural equation modelling. The model was estimated separately for each objective, and fit was assessed using the comparative fit index (CFI), Tucker–Lewis index (TLI), and root-mean-square error of approximation (RMSEA), for which values of 0.08 and below

indicate acceptable fit. Covariance-based estimation was selected over partial least squares because the aim was confirmatory rather than predictive. Discriminant validity was assessed using the Fornell–Larcker criterion. The reproducibility of the pipeline ensured that every reported statistic could be traced to the data and code that produced it, a discipline that strengthened the credibility of the findings and would facilitate replication.

Qualitative data were analyzed through reflexive thematic analysis, proceeding through familiarization, coding, theme construction, review, and definition, with the trustworthiness of the resulting themes established with reference to credibility, transferability, dependability, and confirmability. The overall KMO was 0.872 (Bartlett's $\chi^2 = 7097.2$, $p < .001$), confirming the data's suitability for factor analysis.

Results and Discussion

Measurement and Model Fit

The measurement models demonstrated acceptable to good properties. Nine of thirteen constructs met the 0.70 threshold for internal-consistency reliability, and most met the thresholds for composite reliability and average variance extracted. Three constructs: threat exposure, financial harm, and digital wellbeing recorded reliability and convergent validity below convention; this limitation is reported transparently and is revisited below. The structural models for Objectives Two and Three demonstrated good fit (CFI = .929 and .923; RMSEA = .066 and .057, respectively), while the Objective One model showed weaker fit (CFI = .749; RMSEA = .091), reflecting the three weaker constructs.

Engagement, Exposure, and Multidimensional Harm

Social motivations exerted a strong and significant effect on behavioral engagement ($\beta = 0.633$, $p < .001$), whereas technological factors did not ($\beta = 0.120$, $p = .123$). In a population for whom mobile access is effectively universal, it was the social pull of the platforms, rather than the mere means of access, that drove the engagement leading to exposure. Behavioral engagement in turn significantly increased threat exposure ($\beta = 0.534$, $p < .001$), confirming that exposure arose from ordinary, motivated participation.

Threat exposure produced all four harms, but in markedly different magnitudes. The strongest effect was on financial harm ($\beta = 0.738$, $p < .001$), followed by psychological harm ($\beta = 0.565$, $p < .001$), digital wellbeing ($\beta = 0.406$, $p < .001$), and platform trust ($\beta = 0.359$, $p < .001$). The dominance of financial harm is consistent with the integration of social media and mobile money in Kenya, which allows a successful deception to convert directly into monetary loss, and with the economic precarity that heightens susceptibility to financially framed deception (Burruss et al., 2026; Abubakari & Oseh-Ovarah, 2025). The descending order of the four harms is itself a substantive finding: the burden of social engineering falls most heavily on users' finances, then their psychological state, and least on their trust in the platforms.

The practical magnitude of these relationships reinforced their significance. Effect sizes (Cohen's f^2) were large for the paths from social motivations to engagement ($f^2 = 0.669$), from engagement to exposure ($f^2 = 0.399$), and from exposure to psychological harm ($f^2 = 0.469$), and were largest of all for the path from exposure to financial harm ($f^2 = 1.196$). The exposure-to-financial-harm

relationship therefore stands out not only as statistically significant but as practically the most consequential in the model, accounting for approximately 55% of the variance in financial harm.

Group comparisons revealed a gendered asymmetry between encounter and consequence. Men reported significantly higher threat exposure than women (M = 3.92 vs. 3.72, p = .017), whereas women reported significantly higher psychological harm (M = 3.29 vs. 2.87, p = .002, Cohen's d = −0.45). Exposure did not translate into psychological harm uniformly; the affective burden of the same threat environment fell more heavily on female users, a consideration borne in mind in a sample that was itself predominantly female.

The qualitative strand corroborated this pathway. Participants across groups described encountering deception routinely because they used social media intensively for communication and business; they identified financial deception, fraudulent investment schemes, fake job offers, and impersonation-enabled money requests, as the most consequential and hardest-felt harm, connecting their susceptibility to economic pressure and youth unemployment; and they reported anxiety, embarrassment, and diminished confidence following victimization. The recurrent linkage participants drew between financial susceptibility and economic precarity illuminated why the exposure-to-financial-harm relationship was the strongest in the model, situating a statistical magnitude within the lived economic reality of the respondents.

The complete set of structural path estimates and hypothesis decisions is presented in Table 1.

Table 1

Structural path estimates and hypothesis outcomes (N = 304)

Path	β	p	Decision
Technological factors → Engagement	0.120	.123	Not supported
Social motivations → Engagement	0.633	< .001	Supported
Engagement → Threat exposure	0.534	< .001	Supported
Threat exposure → Financial harm	0.738	< .001	Supported
Threat exposure → Psychological harm	0.565	< .001	Supported
Threat exposure → Digital wellbeing	0.406	< .001	Supported
Threat exposure → Platform trust	0.359	< .001	Supported
Foundational hygiene → Proactive defense	0.531	< .001	Supported
Foundational hygiene → Advanced protection	0.284	< .001	Supported
Proactive defense → Advanced protection	0.666	< .001	Supported
Advanced protection → Recovery & help-seeking	0.786	< .001	Supported
Foundational hygiene → Sustained safe continuance	0.239	.022	Supported
Advanced protection → Sustained safe continuance	−0.117	.480	Not supported
Recovery & help-seeking → Sustained safe continuance	0.278	.051	Not supported
Psychological harm → Sustained safe continuance	−0.143	.057	Not supported

Note. β = standardized path coefficient. Paths with p = .051 and p = .057 approached but did not reach the .05 threshold.

The Structure of Protective Capability

Protective behavior formed an ordered progression of capability. Foundational hygiene significantly enabled proactive defense (β = 0.531, p < .001) and advanced protection (β = 0.284, p < .001); proactive defiance enabled advanced protection (β = 0.666, p < .001); and advanced protection enabled recovery and help-seeking (β = 0.786, p < .001). The strengthening of these coefficients toward the apex of the hierarchy indicates that the most demanding behaviors were the most tightly dependent on the capability beneath them, supporting a developmental rather than parallel account of protective behavior (Hasan et al., 2025; Naeem & Mushibwe, 2025). The qualitative theme of a graduated protective repertoire, from basic vigilance through preventive practice to post-incident recovery, mirrored this structure, and participants described acquiring these competencies cumulatively and informally.

Protective Capability, Harm, and Sustained Safe Continuance

Among the predictors of sustained safe continuance, foundational security hygiene was the only significant one (β = 0.239, p = .022). The positive effect of recovery and help-seeking (β = 0.278, p = .051) and the negative effect of psychological harm (β = −0.143, p = .057) approached but did not reach significance, and advanced protection had no significant effect (β = −0.117, p = .480). Safe continuance was thus sustained by the consistent practice of routine security fundamentals rather than by occasional advanced counter-measures, a finding consistent with the self-regulation view of continuance, in which persistence depends on the ongoing management of routine risk (Lin et al., 2014).

The limited contribution of recovery and help-seeking was illuminated by the qualitative evidence. Participants relied overwhelmingly on friends and family for assistance while finding formal support, platform reporting and institutional channels alike, complicated, slow, and ineffective. Where the formal support ecosystem on which effective recovery depends is inadequate, the willingness to seek help cannot reliably translate into restored safety and sustained participation. This mixed-methods integration converts a quantitative null result into a substantive insight about the inadequacy of formal support, an insight neither strand could have yielded alone.

Theoretical Implications

The findings extend each of the three traditions the framework integrates. By modelling four harms as distinct outcomes of a common exposure and estimating their relative magnitudes, the study advances beyond the aggregated or single-outcome treatments that have characterized prior research, furnishing a differentiated theory of social engineering harm in which the several consequences of victimization are ordered rather than equivalent. By establishing that protective behaviors form an ordered progression, the study contributes a developmental theory of protective behavior, extending the coping-appraisal logic of protection motivation theory to show that response efficacy accumulates hierarchically. And by demonstrating that safe continuance is supported by foundational hygiene while the contribution of help-seeking is conditioned by the support ecosystem, the study extends resilience theory beyond the individual to the institutional environment in which individual resilience operates. The maintenance of safe participation is, on

this account, a joint product of the user's capability and the responsiveness of the systems to which the user can turn.

Conclusions

This study developed and validated an integrated framework of social engineering exposure, multidimensional harm, protective capability, and sustained safe continuance among Kenyan social media users. It established that social rather than technological motivation drives the engagement that exposes users to social engineering; that exposure produces a differentiated profile of harm dominated by financial loss; that protective behavior develops as an ordered progression of capability; and that safe continuance is sustained by foundational security hygiene rather than by advanced protection, with the contribution of help-seeking constrained by the inadequacy of formal support. The framework's contribution lies in integrating segments that prior work addressed separately, differentiating the harms of social engineering rather than aggregating them, and connecting protective capability to sustained safe continuance.

Practically, the findings carry several implications. Because exposure bound most strongly to financial harm, the protection of users' financial interests warrants particular priority in mobile-money environments; measures that interrupt the pathway from a suspicious contact to a financial transfer address the harm most tightly coupled to exposure. Because protective capability develops cumulatively, security education should be sequenced developmentally, establishing foundational hygiene before advancing to sophisticated counter-measures rather than delivering disparate practices uniformly. Because safe continuance was sustained by foundational hygiene, interventions aimed at ordinary users may reasonably concentrate on embedding consistent security fundamentals, a priority that is encouraging in a setting of uneven cybersecurity resourcing because it lies within the reach of ordinary users. And because users defaulted to informal networks when formal channels proved unresponsive, the formal support ecosystem requires substantial strengthening, through simpler and faster reporting, timely assistance to victims, and greater public awareness of available support, so that users' help-seeking can translate into restored safety.

The study has limitations. Three constructs were measured below conventional thresholds and warrant re-specification; the sample was predominantly young and female, delimiting generalizability to the male social media users; and the cross-sectional design precludes causal confirmation. Future research should refine the weaker constructs, employ longitudinal designs, examine more demographically diverse populations, distinguish informal from formal support, and investigate how AI-enabled deception alters the pathways modelled here. Notwithstanding these limitations, the study offers an integrated, empirically grounded, and contextually situated account of how Kenyan social media users experience, resist, and sustain safe participation in the face of social engineering.

References

Abubakari, Y., & Oseh-Ovarah, V. (2025). The gamification of online romance fraud through offenders' cards. *Digital Threats: Research and Practice, 6*(4). https://doi.org/10.1145/3773289

Al-Azawei, A. (2018). Predicting the adoption of social media: An integrated model and empirical study on Facebook usage. *Interdisciplinary Journal of Information, Knowledge, and Management, 13,* 233–252. https://doi.org/10.28945/4106

Burruss, G., Weber, C., Fisk, N., & Giovannetti, F. (2026). Pandemic and predation: An analysis of economic stressors and online scam susceptibility during COVID-19. *Journal of Crime and Justice, 49*(3), 291–310. https://doi.org/10.1080/0735648X.2025.2464015

Frauenstein, E. D. (2019). An investigation into students' responses to various phishing emails and other phishing-related behaviors. *Communications in Computer and Information Science, 973,* 44–59. https://doi.org/10.1007/978-3-030-11407-7_4

Goliath, S., Tsibolane, P., & Snyman, D. (2026). Exploring the cybersecurity–resilience gap: An analysis of student attitudes and behaviors in higher education. *Communications in Computer and Information Science, 2661,* 185–200. https://doi.org/10.1007/978-3-032-09660-9_19

Hasan, M. M., Rahman, K. M., Nath, A., Fuad, M. N., Inthakab, K. S. M., & Emee, A. F. (2025). Understanding user adoption of cybersecurity practices through awareness and perception for enhancing network security in Bangladesh. *2025 IEEE 2nd International Conference on Computing, Applications and Systems (COMPAS 2025).* https://doi.org/10.1109/COMPAS67506.2025.11381666

Kaur, S., Nancy, & Kumar, P. (2024). Detection of fake URLs using deep LSTM architecture over social media. *Lecture Notes in Networks and Systems, 1048,* 30–41. https://doi.org/10.1007/978-3-031-64650-8_3

Lin, H., Fan, W., & Chau, P. Y. K. (2014). Determinants of users' continuance of social networking sites: A self-regulation perspective. *Information and Management, 51*(5), 595–603. https://doi.org/10.1016/j.im.2014.03.010

Liu, X., & Li, Y. (2025). Exploring the impact of persuasion techniques on social media users' response behavior to phishing attacks. *Lecture Notes in Business Information Processing, 551,* 233–247. https://doi.org/10.1007/978-3-031-94190-0_20

Naeem, N.-I.-K., & Mushibwe, C. P. (2025). Navigating digital worlds: A scoping review of skills and strategies for enhancing digital resilience among higher education students on social media platforms. *Discover Education, 4*(1). https://doi.org/10.1007/s44217-025-00432-7

Ngo, F. T., Agarwal, A., & Holman, K. (2026). Cyber hygiene and cyber victimization among limited English proficiency (LEP) internet users: A mixed-method study. *Victims and Offenders, 21*(5), 969–991. https://doi.org/10.1080/15564886.2024.2329765

Pushpavalli, R., Rengaramanujam, P., Berna, I. E., Suseela, D., Dilli, B. M., & Vignesh, R. S. (2026). Detection of cyber threats on social media using optimized sentiment-aware deep learning models. *International Journal of Engineering Trends and Technology, 74*(3), 336–345. https://doi.org/10.14445/22315381/IJETT-V74I3P123

Saleem, M., Kumar, R., Chawla, C., & Singh, M. (2024). Understanding the human factors in the psychology of cyber threats. In *Human impact on security and privacy: Network and human security, social media, and devices* (pp. 39–58). https://doi.org/10.4018/979-8-3693-9235-5.ch003

Sirma, J., Wamunyu, P. K., & Ndirangu, D. (2026). Factors contributing to vulnerability towards social engineering attacks among Kenyan social media users. *Kabarak Journal of Research & Innovation, 16*(2), 335–350. https://doi.org/10.58216/kjri.v16i2.676

Suet Ching, C., Sinnappan, S., & Periyayya, T. (2026). The relationship between risk factors and susceptibility to investment scams among Malaysian young adults. *Journal of Financial Crime, 33*(1–2), 37–52. https://doi.org/10.1108/JFC-07-2025-0228

Tarhini, A., Alalwan, A. A., Cao, D., & Al-Qirim, N. (2021). Integrating emotional attachment, resource sharing, communication and collaboration into UTAUT2 to examine students' behavioral intention to adopt social media networks in education. *International Journal of Technology Enhanced Learning, 13*(1), 1–23. https://doi.org/10.1504/IJTEL.2021.111588

Yeng, P. K., Fauzi, M. A., Yang, B., & Nimbe, P. (2022). Investigation into phishing risk behavior among healthcare staff. *Information, 13*(8), Article 392. https://doi.org/10.3390/info13080392