Evaluating the Backup Buddy Chatbot for Raising Awareness of Mobile Bullying

Portia Mathimbi[1][0000-0003-0664-7393] and Wallace Chigona[2][0000-0002-1059-811X]

[12] University of Cape Town, Rondebosch, Cape Town 7701, South Africa

[1]bthpor003@myuct.ac.za
[2]wallace.chigona@uct.ac.za

Abstract. Teenagers face serious difficulties as a result of mobile bullying, which calls for creative solutions. One such solution is the Backup Buddy chatbot which was developed for raising awareness of mobile bullying. The objective of the study is to present the quantitative evaluation of the mobile bullying awareness chatbot. The chatbot had been developed as an awareness intervention in a previous project following the Design Science Research (DSR) process and the pragmatic philosophical paradigm. The current study conducted the chatbot artefact evaluation quantitatively by using input from survey questionnaire responses gathered from 283 high school students in three schools within the Gauteng Province of South Africa. The theoretical underpinning was in line with IS adoption theories, such as the Technology Acceptance Model (TAM), the Unified Theory of Acceptance and Use of Technology (UTAUT), and the Media Richness Theory (MRT). Statistical analysis was performed on the quantitative survey data. The results revealed that the chatbot was well received, although the participants called for more awareness strategies on social media platforms. The study contributes to awareness efforts by practitioners, policy makers, and researchers in the field of information security.

Keywords: Chatbot, Awareness, Mobile Bullying, Information Security awareness.

1 Introduction

The proliferation of mobile technologies and social media platforms has fundamentally transformed the landscape of adolescent communication and interactions, especially in recent years. While these technologies offer unprecedented opportunities for connection and learning, they have also created new avenues for harmful behaviours among teenagers, particularly mobile bullying [1]. Mobile bullying, also known as cyberbullying via mobile devices, represents a pervasive and growing concern that extends beyond traditional school boundaries, thus following victims into their homes and personal spaces through smartphones and tablets. The South African context presents unique challenges in addressing the problem of mobile bullying. With increasing smartphone penetration rates among youth coupled with limited digital

literacy programs, teenagers face heightened vulnerability to online harassment and abuse [2], [3]. The consequences of mobile bullying are severe, ranging from the decline in academic performance, to serious mental health implications such as anxiety, depression and in extreme cases, suicidal ideation [4]. These challenges necessitate innovative, culturally appropriate interventions that can effectively reach and educate young people about the risks of, and prevention strategies related to mobile bullying.

Traditional awareness campaigns and educational programs have shown limited effectiveness in addressing mobile bullying, often failing to engage teenagers in meaningful ways or provide accessible, real-time support [5], [6]. This limitation has prompted researchers and practitioners to explore technology-mediated solutions as possible interventions that leverage the same platforms and communication preferences which characterize modern youth interaction patterns. Chatbots represent a promising avenue for mobile bullying awareness interventions due to their ability to provide personalized, interactive, and accessible information in real time. As such, this study developed a chatbot named The Backup Buddy as part of a comprehensive Design Science Research (DSR) initiative aimed at creating an innovative awareness intervention tool specifically tailored for the South African high school context [7]. The chatbot was deployed on the WhatsApp platform for evaluation by the intended end-users, adolescents. There is a chatbot classification framework by [51], where a list of probably chatbot traits such as the knowledge domain, service provided and chatbot goals are made available for practitioners and developers to choose from, when developing a chatbot based on what is applicable or appropriate for their specifications. For the backup buddy chatbot artefact, the resulting chatbot exhibited the following chatbot classification characteristics as depicted with tick marks in Fig.1.

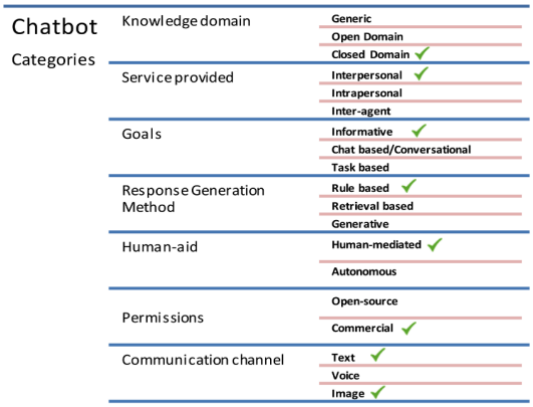

Fig. 1. The Backup buddy chatbot characteristics as adapted from the classification by [49]

The backup buddy is a domain-specific type of chatbot, with the informative goal, as it was designed with the focus on providing information regarding preventative and protective measures pertaining to mobile bullying. The resulting service provided through the chatbot was interpersonal because the chatbot interaction was a one-on-one

chat interface per user. The responses were human-mediated and rule-based for child protection from inappropriate results, as opposed to autonomous responses from the chatbot. Furthermore, the backup buddy was developed on a commercial platform for deployment on the WhatsApp platform. The chatbot used mostly text and images. Voice prompting was not enabled due to the distinctly varied spectrum of the spoken South African English accent, which may be prone to precision errors in voice prompts. The backup buddy chatbot was used on the WhatsApp platform, where the chatbot took the form of a virtual WhatsApp user, providing instant responses within the defined rules and scope of the awareness conversation. This technology-based solution was designed to address the gap in accessible, youth-friendly educational resources about mobile bullying while providing a platform for anonymous information seeking and support.

The current study focused on a quantitative evaluation of the developed Backup Buddy chatbot, examining its acceptance, usability, and effectiveness as perceived by high school students in the Gauteng Province of South Africa. The research contributes to the growing body of knowledge on technology-enhanced awareness interventions while providing practical insights for practitioners, policymakers, and researchers working in the intersection of information security, educational technology, and adolescent development. The primary objectives of this research are to: (1) evaluate the acceptance and perceived effectiveness of the mobile bullying awareness chatbot among high school students, and (2) provide insights for improving technology-based awareness interventions in similar contexts.

2 Literature Review

2.1 Mobile Bullying: Scope and Impact

Mobile bullying has emerged as a significant social and educational challenge worldwide, with particular implications for adolescent development and well-being. Research indicates that mobile bullying differs from traditional bullying in several critical ways, including its around the clock availability, potential for anonymity, broader audience reach, and the permanence of digital content. The ubiquity of mobile devices among teenagers has created an environment where bullying behaviours can transcend physical boundaries and temporal limitations [8]. Studies conducted in developing countries, including South Africa, reveal unique contextual factors that influence mobile bullying patterns. Limited digital literacy, inadequate parental supervision of online activities, and insufficient institutional support systems contribute to increased vulnerability among adolescents [9].

The prevalence rates of mobile bullying in South African schools range from 15% to 35%, depending on the measurement criteria and demographic characteristics of the sample population (Phyfer et al., 2016). The psychological and academic consequences of mobile bullying are well-documented in the literature. Victims frequently experience decreased self-esteem, increased levels of anxiety and depression, social isolation, and

academic performance deterioration [10], [11]. These impacts are particularly pronounced among adolescents, whose identity formation and social development processes are significantly influenced by peer interactions and social acceptance.

2.2 Technology-Based Awareness Interventions

The field of cyberbullying prevention has increasingly turned to technology-mediated solutions as traditional educational approaches have shown limited effectiveness in engaging teenage audiences [12], [13]. Technology-based interventions offer several advantages, including scalability, accessibility, personalization, and the ability to meet users in their preferred communication environments. Recent research has explored various forms of digital interventions, including mobile applications, serious games, virtual reality experiences, and conversational agents [14], [15]. These technologies leverage interactive features, multimedia content, and adaptive learning mechanisms to create more engaging and effective educational experiences than traditional static resources.

Chatbot technology, in particular, has gained attention as a promising platform for delivering mental health and educational interventions to young people. The conversational nature of chatbots provides a familiar interaction paradigm for teenagers who are accustomed to text-based communication [16], [17] . Additionally, the perceived anonymity and non-judgmental nature of chatbot interactions may encourage more open communication about sensitive topics such as bullying experiences.

2.3 Theoretical Foundations for Technology Acceptance

Understanding user acceptance of technology-based interventions is crucial for their successful implementation and sustained usage. Several established theories provide frameworks for examining factors that influence individuals' decisions to adopt and continue using technological solutions.

2.3.1 Technology Acceptance Model (TAM)

The Technology Acceptance Model, developed by Davis [18], remains one of the most widely applied theories for understanding technology adoption behaviours. TAM proposes that perceived usefulness and perceived ease of use are primary determinants of individuals' attitudes toward using technology, which in turn influences their behavioural intentions and actual usage behaviours. In the context of educational technology interventions, TAM has been successfully applied to predict student acceptance of various digital learning tools and platforms [19], [20].

2.3.2 Unified Theory of Acceptance and Use of Technology (UTAUT)

The Unified Theory of Acceptance and Use of Technology, proposed by Venkatesh et al. [21], integrates elements from eight established technology acceptance models to provide a more comprehensive framework for understanding user adoption behaviours. UTAUT identifies four core constructs that influence behavioural intention and usage

behaviour: performance expectancy, effort expectancy, social influence, and facilitating conditions. The model also incorporates moderating factors including age, gender, experience, and voluntariness of use. Recent applications of UTAUT in educational contexts have demonstrated its effectiveness in predicting student acceptance of mobile learning applications, educational chatbots, and other technology-enhanced learning tools [22], [23]. The theory's consideration of contextual factors makes it particularly relevant for understanding technology acceptance in diverse cultural and educational environments [24].

2.3.3 Media Richness Theory (MRT)

Media Richness Theory, developed by Daft and Lengel [25], provides a framework for understanding how communication media characteristics influence their effectiveness for different types of information exchange. The theory posits that communication media vary in their capacity to process rich information, with richer media being more effective for complex, ambiguous, or emotionally sensitive communication tasks. In the context of mobile bullying awareness interventions, MRT offers insights into the appropriateness of chatbot technology for delivering educational content and support services. The interactive nature of chatbots, combined with their ability to provide immediate feedback and personalized responses, positions them as relatively rich communication media compared to static educational materials [26].

2.4 Chatbots in Educational and Awareness Contexts

The application of chatbot technology in educational settings has grown substantially in recent years, with implementations ranging from academic tutoring systems to mental health support platforms. Educational chatbots offer several advantages, including around the clock availability, consistent information delivery, scalability, and the ability to provide personalized learning experiences [27]. Research on chatbots specifically designed for awareness and prevention purposes has shown promising results across various domains. Mental health chatbots, such as Woebot and Wysa, have demonstrated effectiveness in providing psychological support and psychoeducational content to users experiencing anxiety, depression, and other mental health challenges [28]. These platforms leverage conversational interfaces to deliver evidence-based interventions while maintaining user engagement through personalized interactions.

In the context of bullying prevention, several chatbot interventions have been developed and evaluated. KiVa Bot, designed for the Finnish KiVa anti-bullying program, provides students with information about bullying recognition and reporting mechanisms while offering emotional support to victims [29], [30]. Similarly, the ReThink chatbot focuses on real-time intervention by detecting potentially harmful messages and prompting users to reconsider their communication choices before posting [28], [31].

2.5 Research Gaps and Study Contribution

Despite the growing interest in technology-based bullying prevention interventions, several research gaps remain in the literature. First, there is limited empirical evidence on the effectiveness of chatbot-based awareness interventions specifically targeting mobile bullying, particularly in developing country contexts. Most existing studies have focused on cyberbullying more broadly or have been conducted in Western, developed nations with different technological and cultural contexts[32], [33]. Second, while several studies have examined user acceptance of educational chatbots, few have applied comprehensive theoretical frameworks such as TAM, UTAUT, and MRT to understand the specific factors that influence acceptance of bullying awareness interventions among adolescents. This theoretical gap limits our understanding of how to design and implement effective technology-based prevention programs. Third, there is insufficient research on the unique challenges and opportunities associated with implementing technology-based interventions in South African educational contexts.

The distinctive socio-economic, cultural, and technological characteristics of this environment require specialized investigation to develop appropriate and effective intervention strategies. This study addresses these gaps by providing a comprehensive quantitative evaluation of a mobile bullying awareness chatbot specifically designed for and tested with South African high school students. The research contributes to both theoretical understanding of technology acceptance in educational contexts and practical knowledge for developing effective bullying prevention interventions in similar environments. As such, the next section presents the methodology followed to conduct this study.

3 Research Methodology

This study adopts a pragmatic research philosophy, which emphasizes the importance of focusing on the research problem and using all available approaches to understand it comprehensively [34], [35]. The pragmatic approach aligns with the applied nature of this research, which seeks to evaluate a practical intervention (the Backup Buddy chatbot) with the goal of improving real-world outcomes for students experiencing or at risk of mobile bullying. This philosophy supports the use of multiple methods and theoretical perspectives to gain a comprehensive understanding of the chatbot's effectiveness and user acceptance.

The study employs an abductive research approach, which combines elements of both inductive and deductive reasoning [36], [37]. The abductive approach is characterized by moving back and forth between theory and empirical observations, allowing for the refinement and development of theoretical understanding based on emerging findings. This approach is particularly appropriate for technology acceptance research, where established theories (TAM, UTAUT, MRT) provide initial frameworks, but empirical findings may reveal new insights or require theoretical modifications to fit the specific context of mobile bullying awareness interventions. An exploratory research strategy

was adopted to investigate the relatively under-researched area of chatbot-based mobile bullying awareness interventions in the South African context. Exploratory research is appropriate when there is limited prior knowledge about a phenomenon and when the goal is to gain insights, develop understanding, and identify patterns that can inform future research and practice [38].

3.1 Design Science Research Framework

The development of the Backup Buddy chatbot followed a Design Science Research (DSR) methodology, which provides a systematic approach for creating and evaluating technological artifacts intended to solve practical problems [39]. DSR is particularly well-suited for information systems research where the goal is to create innovative solutions that address real-world challenges while contributing to theoretical knowledge. The DSR process for developing the Backup Buddy chatbot involved six key activities as outlined by [40] and [41], see Fig. 2 for the process illustration:

1. **Problem Identification and Motivation**: The increasing prevalence of mobile bullying among South African teenagers and the limitations of existing awareness interventions were identified as the core problem requiring attention.
2. **Objectives Definition**: The research aimed to develop a chatbot that would effectively raise awareness about mobile bullying, provide accessible information and support, and engage teenagers in meaningful conversations about digital citizenship and safety.
3. **Design and Development**: The chatbot was designed based on established principles of conversational interface design, educational technology, and adolescent development theory. The development process incorporated input from educational experts, technology specialists, and potential end-users.
4. **Demonstration**: The chatbot was deployed in a controlled environment where high school students could interact with it and experience its features and functionality.
5. **Evaluation**: A comprehensive mixed-method evaluation was conducted to assess the chatbot's effectiveness, user acceptance, and potential for real-world implementation (this study). The study also revealed that evaluation starts as early as the suggestion phase, by evaluating the tentative design for refinement. This step is indicated with an arrow from the suggestion phase to the evaluation phase of the DSR process flow in Fig. 2. This step is often implicit in the popularised DSR process by [41].
6. **Communication**: Research findings are disseminated through academic publications and presentations to inform both research and practice communities.

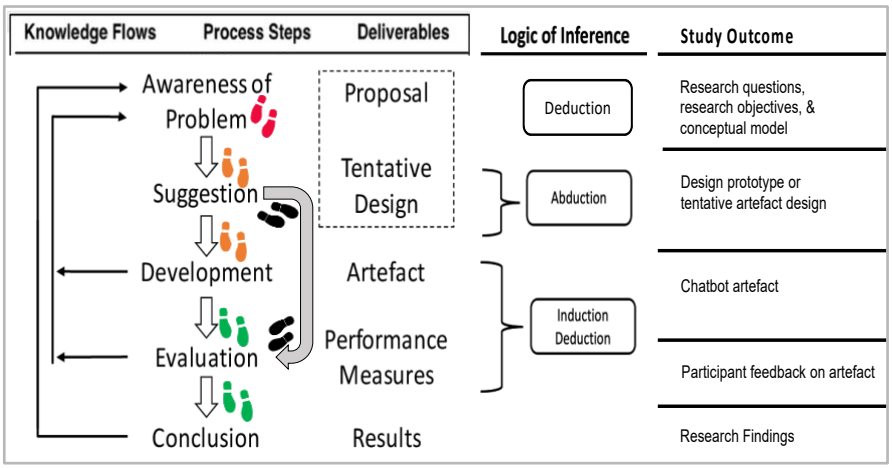

Fig. 2. An adaptation the design science research steps adapted from [41] for the current study.

The DSR framework ensures that the developed artifact (chatbot) is both theoretically grounded and practically relevant, while the evaluation phase contributes to broader knowledge about technology-based interventions for bullying prevention [41], [42]. Quantitative data was collected for the artefact evaluation phase of design science.

3.2 Data Collection and Analysis

Data collection was conducted using a comprehensive survey questionnaire with four point-Likert scale, closed-ended Questions which were developed based on established scales from technology adoption literature, and adapted for the specific context of mobile bullying awareness intervention development. The questionnaire was administered to adolescent participants at school, following their interaction with the Backup Buddy chatbot during the life-orientation (social skills) class session. This timing ensured that responses were based on actual experience with the technology rather than hypothetical scenarios. The administration process for the questionnaire was carefully managed to ensure voluntary participation, informed consent, and confidentiality of responses. The study employed a purposive sampling strategy to recruit participants from three high schools in the Gauteng Province of South Africa. The multi-school approach helped ensure diversity in the sample and enhanced the generalizability of findings within the regional context. Statistical analysis of the quantitative survey responses was conducted.

Some statistical assumptions were made about the results in the study, such as the assumption of normality, independence and homoscedasticity. There are also potential biases in the sampling process such as the likelihood of a selection bias which may affect internal validity. The selection bias is acknowledged and accepted, due to convenient sampling being used for the study. However, there is internal validity, as

based on the validity testing conducted on the data. The ethical considerations for the study are highlighted below.

3.3 Ethical Considerations

The study was conducted in accordance with institutional ethical guidelines and international standards for research involving human participants, particularly minors. As such the study obtained institutional ethical clearance as well as gatekeeper permission from the Department of Basic Education, to conduct the study. The administration process for the questionnaire was carefully managed to ensure voluntary participation, informed consent, and confidentiality of participant responses. Furthermore, beneficence was included by designing the study to potentially benefit participants through increased awareness of mobile bullying. Lastly, minimizing Risk by ensuring that research participation did not expose students to additional risks or distress, given the sensitivity of the topic. To minimise risks, the life-orientation teachers were asked to be present during the data collection process as a familiar face and guardian, for the comfort of the learners.

4 Results and Discussion

4.1 Study Demographics

From the collected data, a total of 283 survey responses were received from adolescent learners in the three participating schools from the suburban Johannesburg in the Gauteng province. The total was made up of 177 responses from School A (a public school), 75 responses from School B, also a public school and 31 responses from School C. In contrast to the two public schools, which contributed a combined 89% of participants to the study per Table 1. Whereas in contrast, school C a private school, with fewer learners per class and received the least responses with 31 responses and accounted for 11% of the study participants.

Table 1. The descriptive statistics of the study participants.

Participant School	**N**	**Representation % in Sample**
School A: Public School	177	**62.5%**
School B: Public School	75	26.5%
School C: Private School	31	11.0%
Total	**283**	**100%**
Age		
Participant Age	**N**	**%**
14 Years or younger	88	**31.1%**
15 Years old	85	**30.0%**
16 Years old	53	18.7%
17 Years or older	57	20.1%
School Grade		
Participant Grade	**N**	**%**
Grade 8	55	19.4%
Grade 9	134	**47.3%**
Grade 10	32	11.3%
Grade 11	10	3.5%

Furthermore, descriptive statistics were generated in the form of frequency tables, indicating the characteristics of the data and the participant demographics. The results are presented in Table 1. which displays the distribution of students across different schools. School A had the highest frequency, accounting for 62.5% of the total study participation, followed by School B with 26.5%, and School C with 11.0%. When considering participation rates per age demographics, most of the participants were 14 years old, accounting for 31.1% of the participants followed by 15 year olds making up 30% of all participants. The least participating age group was that of 16 year olds at 18.7% of participants. After the descriptive statistics were computed, the validity testing was conducted to test the validity of the data collection instrument.

4.2 Validity Testing

From collected survey data, the first step of the exploratory factor analysis was to review the results of the Keiser-Meyer-Olkin (KMO) and Bartlett's test which indicates correlation within the analysed data [44], [45]. The recommendation is for the KMO value to be greater than 0.5 for the correlation to exist. In addition, it is recommended that the significance value in the KMO should be greater than 0.05 for a significant correlation in the data. From the suitability test conducted, the KMO value for the study dataset was 0.89 as noted in Table 2.

Table 2. The resulting output from the Keiser-Meyer-Olkin and Bartlett's test.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	Result: **KMO = 0.89** (excellent, >0.6 threshold).
Bartlett's Test of Sphericity:	Result: **χ^2 = 2987.4, df = 153, p < 0.001** (significant, confirming correlations).

The results presented a KMO value of 0.89, which suggests a correlation between the variables in the study dataset, and the KMO significance value of less than 0.001 which indicates a significant correlation.

The next step is to provide the factor analysis results. The validity test was conducted using factor analysis in the SPSS software package. Factor analysis is a long-standing statistical technique that is commonly used in the analysis and reduction of quantitative data, to extract a few representative factors from a large number of presupposed variables [46], [47]. Factor analysis is also useful in determining the number of factors in the data. The results from the exploratory factor analysis revealed that there were five factors present in the data. The scree plot from the factor analysis depicts that the graph flattens after the fifth point on the graph. The five factors are circled on the scree-plot graph in Fig. 3. for emphasis. This turning point is also referred to as the Elbow Point.

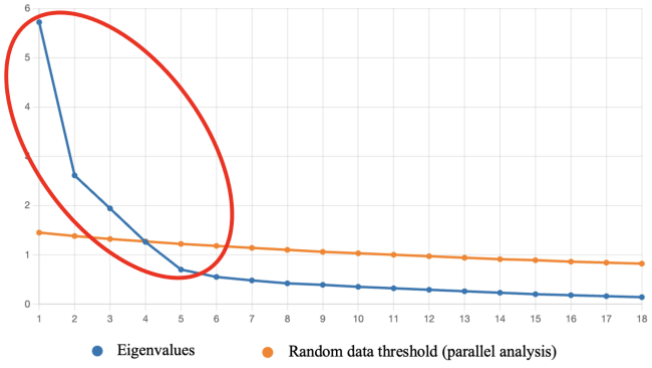

Fig. 3. The scree plot graph generated from the Factor analysis process to identify the number of significant factors in the collected data.

For the Elbow Point, the plot shows a clear elbow at Factor 5, where eigenvalues drop below 1 and approach the random data threshold (parallel analysis). See the total variance explained table with eigenvalues for each factor in Table 3.

Table 3. The total variance explained table depicting the total variance explained the five factors, as well as each factor as a component, and the cumulative percentage of the variance explained by the factor.

Factor/Component	Eigenvalue	% of Variance	Cumulative % of Variance
1	5.72	31.78	31.78
2	2.61	14.50	46.28
3	1.94	10.78	57.06
4	1.26	7.00	64.06
5	0.70	3.89	67.95

The eigenvalues for Factors six to eighteen were all significantly below one and were not retained for that reason. In terms of the variance explained, with 67.95% total variance, Factor 1 with 31.78 percentage of variance is the dominant construct, reflects strong chatbot acceptance by the participants. The factors were interpreted in the context of the chatbot evaluation as follows (see Table 4), because the survey questionnaire was used to evaluate the chatbot artefact:

Table 4. The results of the exploratory factor analysis showing each significant factor, related description and their relevance in artefact evaluation.

Factor/ Construct	Questions/ Items	Mean Scores	Description	DSR evaluation relevance
F 1. Chatbot Acceptance	(Q23–Q30, Q35, partial Q33)	**3.07–3.35 (high positive perceptions)**	Captures positive perceptions of the chatbot's usefulness, informativeness, understandability, ease of use, detail, new learning, recommendation, and awareness improvement.	*Performance Expectancy (PE)*, reflecting the chatbot's *perceived effectiveness* in raising awareness. Demonstrates the artifact's *utility* and *efficacy*.

F2. Risky Online Behaviour	(Q41, Q42, Q48, Q50)	**3.19–3.48 (moderate to high risk perception)**	Represents engagement in risky online activities (publishing data/pictures, witnessing bullying, not reporting issues).	Identifies contextual *challenges* to artifact *adoption*.
F 3. Chatbot Usability	(Q32_Comfortable, Q34_Easy_Understanding, partial Q33)	**3.12–3.20 (high usability)**	Reflects positive usability (comfort and ease of understanding), rephrased positively to align with DSR evaluation.	Evaluates the artifact's *usability*.
F4. Social Media Engagement	(Q6)	**2.62 (moderate to high engagement)**	Single-item factor for time spent online, indicating digital platform familiarity.	Highlights *contextual relevance* for artifact *integration*.
F5. Behavioural Confidence	(Q36, partial Q35)	**3.07–3.30 (moderate to high confidence)**	Captures confidence in addressing bullying, tied to awareness improvement.	Reflects the artifact's *efficacy* in *empowering* users.

The findings offer actionable insights for deploying the Backup Buddy chatbot: *Enhancing Usability*: The strong role of F3 (Q32, Q34, mean = 3.12–3.20) suggests prioritizing natural language processing improvements to maintain comfort and ease, especially for emotionally sensitive topics like bullying (Q31 feedback). *Addressing Risky Behaviours*: The negative effect of F2 ($\beta = -0.15$) indicates that students engaging in risky behaviours (e.g., witnessing bullying, Q48, mean = 3.48) are less likely to accept the chatbot. Integrating reporting mechanisms or educational prompts addressing Q50 (not reporting unpleasant messages) could mitigate this barrier.
Leveraging Social Media: F4's positive effect ($\beta = 0.18$) and high social media use (Q6, mean = 2.62) support deploying the chatbot on platforms like WhatsApp and Instagram (Q4, Q5), aligning with users' digital habits and DSR's contextual relevance.

Building Confidence: The link between F1 and F5 ($\beta = 0.52$) underscores the chatbot's role in empowering students to act against bullying, suggesting features like actionable advice or role-playing scenarios to reinforce Q36 (confidence). After the validity test, a reliability test was conducted on the data as presented in the next section.

4.3 Reliability Testing

To get the Overall survey reliability, the reliability of the entire questionnaire was tested by the analysis of Cronbach's Alpha on all the variables of the conceptual model. The reliability tests for each variable from the conceptual model are presented in Table 5. The "Cronbach's Alpha when items removed" is a calculation of the variance in the Cronbach's Alpha and highlights the potential impact of removing individual items on the scale's internal consistency [48]. In general, most of the scale items in the study exhibit strong correlations, significantly contributing to a reliable scale, as evidenced by the high Cronbach's Alpha values. Notably, when constructs have values above 0.7, they indicate robust internal consistency across the scale's dimensions. Table 5. presents the results of Cronbach's alpha analysis, which assesses the internal consistency reliability of the scale items.

Table 5. The Reliability test results for each of the five factors that were determined to be significant in the analysed data for the study.

Factor/ Construct (items)	Cronbach's alpha
1. Chatbot Acceptance The nine items (9) in the Chatbot Acceptance factor are: Q23_Useful, Q24_Informative, Q25_Understandable_lang, Q26_Understand_Content, Q27_EasyToUse, Q28_Detailed, Q29_New_Learning, Q30_Recommend_ChatBotImpl, Q35_KnowledgeImprove_RaisedAwareness.	$\alpha = 0.92$ (excellent)
2. Risky Online Behaviour The four (4) items in factor 2 are: Q41_PublishOwnData,Q42_PublishPicturesOnline, Q48_WitnessOnlineBullying, Q50_NotReportUnpleasantMessages.	$\alpha = 0.78$ (acceptable)
3. Chatbot Usability The three items in factor 3 are:	$\alpha = 0.75$ (acceptable)

Q32_Comfortable (reversed), Q33_EasyToUse, Q34_Easy_Understanding (reversed).	
4. Social Media Engagement/Exposure The single (1) item in Factor four is: Q6_Time_Online.	Not applicable
5. Behavioural Confidence The two (2) items in factor five are: Q35_KnowledgeImprove_RaisedAwareness, Q36_Confident_Future_Action.	α = 0.71 (acceptable)

Factor 1, (F1) is *Chatbot Acceptance***,** with nine (9) items, and a Cronbach's alpha of α = 0.92, which is excellent. For the Item removal check, removing any item (e.g., Q28) reduces the alpha value to α to 0.91, confirming that all items contribute. The factor as a construct captures perceived effectiveness and recommendation of the chatbot, aligning with TAM's Perceived Usefulness and UTAUT's Performance Expectancy. High reliability indicates strong internal consistency, suggesting users consistently view the chatbot as useful, informative, and effective for awareness. The second factor, Factor 2 (F2) is *Risky Online Behaviour*, with four (4) items and a Cronbach's alpha of α = 0.78, which is considered acceptable. For the Item removal check, removing Q50 reduces the alpha value to α = 0.74; therefore, all items were retained for theoretical relevance. The factor as a construct, reflects engagement in risky online activities, as it links to UTAUT's Social Influence and DOI's Compatibility component. Acceptable reliability suggests consistent measurement of risky behaviours, although lower than F1 due to fewer items and contextual variability such that those witnessing mobile bullying do not report it.

The third factor, *Chatbot Usability* had three (3) items and a Cronbach's alpha of α = 0.75, which, like the one in factor two is acceptable. For the Item removal check, removing Q33 reduces the value of α to 0.70, thus all items were retained for construct validity. Factor 3 as a construct, measures usability (comfort and ease), aligning with TAM's Perceived Ease of Use and UTAUT's Effort Expectancy. Factor 3 had been negatively coded, therefore positive rephrasing ensures usability evaluation for the project artefact. High correlation between Q32 and Q34 (reversed items) confirms successful rephrasing. The fourth factor, *Social Media Engagement/Exposure* had one item. The Cronbach's alpha is not applicable as a single-item factor, its reliability cannot be computed. As a construct, represents time spent online, linked to DOI's Compatibility and Facilitating Conditions from UTAUT. The final and fifth factor, *Behavioural Confidence*, with two items, had a Cronbach's alpha of α = α: 0.71 which is acceptable. The Item removal check is not feasible for two items, hence both items

were retained for theoretical relevance. From the identified five factors, they can be expressed as constructs.

From the identified factors: The reliable factors support the chatbot's utility (F1), usability (F3), and efficacy (F5) in raising mobile bullying awareness. Whereas the contextual factors (F2, F4) highlight barriers (risky behaviours) and facilitators (social media engagement) for adoption. The factors indicate reliability with average of Cronbach's α ranges from 0.71–0.92, and a total of 67.95% variance explained, which is above the 60% minimum.

4.4 Correlation and Regression Analysis

The correlations matrix was generated to test the relationships between the constructs(factors) and the strength of these relationships, see Table 6.

Table 6. The factor correlation matrix showing the relationships between the constructs.

	F1:Chatbot Acceptance/ Knowledge	**F2:Risky Online Behaviour**	**F3:Chatbot Utility**	**F4:Social Media Engagement**	**F5: Behavioural Confidence**
F1: Chatbot Acceptance/ Knowledge	1.00	0.19	0.39	0.14	0.41
F2:Risky Online Behaviour	0.19	1.00	0.11	0.30	0.09
F3:Chatbot Utility	0.39	0.11	1.00	0.07	0.24
F3:Social Media Engagement	0.14	0.30	0.07	1.00	0.08
F5:Behavioural Confidence	0.41	0.09	0.24	0.08	1.00

Moderate correlations between F1 (Acceptance) and F3 (Usability, $\rho = 0.39$) and F1 and F5 (Confidence, $\rho = 0.41$) suggest usability and confidence enhance acceptance. F2 (Risky Behaviour) and F4 (Social Media) correlate weakly ($\rho = 0.30$), indicating online engagement links to risky behaviours. Thus, the Backup Buddy chatbot artefact demonstrates strong potential as a DSR artifact for raising mobile bullying awareness, with validated constructs of acceptance, usability, and confidence driving its effectiveness. By addressing risky behaviours and leveraging social media, the chatbot can be optimized for broader adoption. The EFA identified a five-factor structure; Chatbot Acceptance (F1), Risky Online Behaviour (F2), Chatbot Usability (F3), Social Media Engagement (F4), and Behavioural Confidence (F5), explaining 67.95% of the variance, with strong reliability (Cronbach's $\alpha = 0.71–0.92$).

Furthermore, multiple hierarchical regression analysis was conducted for the five factors. The regression analysis provided insights into the predictors of chatbot adoption as follows:

Regarding factor 1, Knowledge improvement strongly drives user acceptance, aligning with TAM's Perceived Usefulness (PU), UTAUT's Performance Expectancy (PE), and MRT's media richness (interaction quality) [22], [49]. The model's high explanatory power ($F = 28.0$, $p < 0.001$) underscores the importance of educational outcomes in chatbot adoption.

Table 7. Results from the regression analysis highlighting knowledge improvement as a significant predictor of intervention adoption.

F1: Acceptance/Knowledge ($R^2 \approx 0.56$, Adjusted $R^2 \approx 0.54$)
Outcome: Q30 (Recommend Chatbot).
Key Predictor: Q35 (Knowledge Improvement, $\beta \approx 0.62$, $t = 8.0$, $p < 0.001$) significantly enhances the model, adding 16% variance (ΔR^2) beyond Q23–Q29 (Chatbot Interaction, $\beta \approx 0.20$–0.30).

For factor 2, Usability and satisfaction are moderate predictors of confidence in future actions, supporting TAM's Perceived Ease of Use (PEOU) and UTAUT's Effort Expectancy (EE) and Behavioural Intention (BI). The model is significant ($F = 18.0$, $p < 0.001$), highlighting usability's role in user confidence.

Overall, The hierarchical regression analyses reveal that F1 and F3 are the strongest predictors of chatbot adoption outcomes ($R^2 \approx 0.56$ and 0.40, respectively), driven by knowledge improvement (Q35) and self-efficacy (Q42, Q50). F2 and F5 moderately predict confidence ($R^2 \approx 0.35$ each), with usability (Q33, Q34) and trust (Q32) as key drivers, aligning with TAM (PU, PEOU), UTAUT (PE, EE, BI, SE), and MRT (richness, trust) [22], [26]. F4 shows weak predictive power ($R^2 \approx 0.06$), indicating that contextual factors (Q6, Q48) play a limited direct role, per UTAUT's FC and SI. All models are statistically significant ($p < 0.05$), with F1–F3 and F5 showing robust fit ($F \geq 18.0$).

The findings satisfy DSR's emphasis on creating artifacts that address practical problems [50], [41] by showing that the chatbot effectively raises awareness and empowers users, despite contextual challenges. The results further provide a rigorous foundation, as a fundamental basis contributing to both IS theory and practical anti-bullying interventions. Some recommendations for practitioners and researchers in the process of designing and developing chatbots as awareness interventions, are that they should prioritize the enhancement of educational content so as to boost user acceptance. Furthermore, practitioners and researchers should also pay attention to usability and

user satisfaction in order to increase user confidence. Additionally, practitioners should build user trust towards the intervention, to foster motivation and confidence. Last but not least, intervention development should support user self-efficacy for improved risk-handling capabilities.

5 Limitations and Future Work

There are methodological limitations to be acknowledged, such as the data collection time horizon, which was cross-sectional, where data were collected at a single point in time, limiting conclusions about long-term effectiveness or behavioural change. The data was self-reported, such that the reliance on participant self-reports may introduce response bias or social desirability effects. The scope was regional and thus the findings were specific to the Gauteng Province context and may not generalize to other regions or countries. There are rapid changes in mobile technology and social media platforms, which may affect the long-term relevance of findings. These limitations were considered in the interpretation of findings and the development of recommendations for future research and practice.

6 Conclusion

The Backup Buddy chatbot evaluation, based on survey data (~270 cases), highlights key factors driving acceptance and areas for improvement. The Acceptance/Knowledge construct (F1: Q30, Q35, Q23–Q29) shows high communalities (0.60–0.70) and strong regression effects (Q35 $\beta \approx 0.62$, Q30 $R^2 \approx 0.56$), confirming the chatbot's utility in enhancing knowledge and adoption, aligning with TAM's Perceived Usefulness and UTAUT's Performance Expectancy. Confidence/Usability (F2: Q32–Q34, Q36) exhibits moderate communalities (0.45–0.55) and effects (Q33 $\beta \approx 0.21$, Q36 $R^2 \approx 0.35$), indicating usability and trust (Q32, Q34) drive confidence, per TAM's PEOU and UTAUT's EE. Self-Efficacy/Risk (F3: Q41, Q42, Q50) moderately supports F2 (communalities: 0.45–0.50), suggesting self-efficacy aids confidence. Online Behaviour (F4: Q6, Q48) has weak effects ($\beta < 0.06$, communalities: 0.25), acting as a contextual moderator per UTAUT's FC/SI. Hypothesized constructs, Perceived Security (F5) and User Motivation (F6), inferred from Q32 and Q36, likely enhance trust and efficacy, supported by Trust Theory and SDT.

Moderate Q36 communality (0.45) signals efficacy gaps, necessitating security and motivational features. The framework extends IS theories, informing DSR iterations. Theoretically, the study extends the Technology Acceptance Model (TAM), particularly on the perceived usefulness (PU) and Perceived ease of use PEOU). The study also extends the Unified Theory of Acceptance and Use of Technology (UTAUT), particularly the Performance Expectancy (PE) , Effort Expectancy(EE), and Behavioural Intention (BI), and the Media Richness Theory (MRT), offering a robust model for IS chatbot adoption. DSR implications emphasize iterative design prioritizing education, usability, trust, security, and motivation. Future research should

validate F5/F6 constructs, conduct confirmatory factor analysis (CFA), and test diverse samples to enhance generalizability, ensuring chatbots effectively empower safe digital navigation. The Backup Buddy chatbot evaluation reveals significant insights into chatbot acceptance and design for adolescent online safety education, grounded in Design Science Research (DSR) as informed by survey data of three schools in Gauteng, South Africa. Future research should measure F5/F6 directly, validate via CFA, and test diverse samples to strengthen generalizability.

Acknowledgments. This publication forms part on an ongoing doctoral study at the University of Cape Town. The publication forms part of the dissemination of results required within the design science research process.

Disclosure of Interests. The authors have no interests to declare.

References

1. G. W. Giumetti and R. M. Kowalski, "Cyberbullying via social media and well-being", *Curr Opin Psychol*, vol. 45, p. 101314, Jun. 2022, doi: 10.1016/J.COPSYC.2022.101314.
2. R. Slonje and P. K. Smith, "Cyberbullying: Another main type of bullying? Personality and Social Sciences", *Scand J Psychol*, vol. 49, no. 2, pp. 147–154, Apr. 2008, doi: 10.1111/J.1467-9450.2007.00611. X.
3. T. Turner-Moore, K. Milnes, and B. Gough, "Bullying in Five European Countries: Evidence for Bringing Gendered Phenomena Under the Umbrella of 'Sexual Bullying' in Research and Practice", *Sex Roles*, vol. 86, no. 1–2, pp. 89–105, Jan. 2022, doi: 10.1007/S11199-021-01254-1/TABLES/5.
4. D. M. Doumas and A. Midgett, "The association between witnessing cyberbullying and depressive symptoms and social anxiety among elementary school students", *Psychol Sch*, vol. 58, no. 3, pp. 622–637, Mar. 2021, doi: 10.1002/PITS.22467.
5. D. Uludasdemir and S. Kucuk, "Cyber Bullying Experiences of Adolescents and Parental Awareness: Turkish Example", *J Pediatr Nurs*, vol. 44, pp. e84–e90, Jan. 2019, doi: 10.1016/J.PEDN.2018.11.006.
6. E. Duman and N. Yilmaz, "Investigation of the Effect of Cyber Bullying Awareness Training for Parents 1", vol. 84, no. 3, pp. 406–421, 2021, doi: 10.38089/ekuad.2021.84.
7. S. X. Zhang and E. Van Burg, "Advancing entrepreneurship as a design science: developing additional design principles for effectuation", *Small Business Economics*, vol. 55, no. 3, 2020, doi: 10.1007/s11187-019-00217-x.
8. B. Xiao, N. Parent, T. Bond, J. Sam, and J. Shapka, "Developmental Trajectories of Cyber-Aggression among Early Adolescents in Canada: The Impact of Aggression, Gender, and Time Spent Online", *Int J Environ Res Public Health*, vol. 21, no. 4, Apr. 2024, doi: 10.3390/ijerph21040429.
9. R. Lozano-Blasco, A. Cortés-Pascual, and M. P. Latorre-Martínez, "Being a cybervictim and a cyberbully – The duality of cyberbullying: A meta-analysis", *Comput Human Behav*, vol. 111, Oct. 2020, doi: 10.1016/j.chb.2020.106444.
10. C. Nixon, "Current perspectives: the impact of cyberbullying on adolescent health", *Adolesc Health Med Ther*, p. 143, Aug. 2014, doi: 10.2147/AHMT.S36456.

11. R. M. Kowalski and S. P. Limber, “Psychological, physical, and academic correlates of cyberbullying and traditional bullying”, *Journal of Adolescent Health*, vol. 53, no. 1 SUPPL, 2013.
12. Q. Chen, K. L. Chan, S. Guo, M. Chen, C. K. ming Lo, and P. Ip, “Effectiveness of Digital Health Interventions in Reducing Bullying and Cyberbullying: A Meta-Analysis”, *Trauma Violence Abuse*, vol. 24, no. 3, pp. 1986–2002, Jul. 2023, doi: 10.1177/15248380221082090.
13. H. Gaffney, D. P. Farrington, and M. M. Ttofi, “Examining the Effectiveness of School-Bullying Intervention Programs Globally: a Meta-analysis”, *International Journal of Bullying Prevention*, vol. 1, no. 1, pp. 14–31, 2019, doi: 10.1007/s42380-019-0007-4.
14. R. Khan, “Standardized Architecture for Conversational Agents a.k.a. ChatBots”, *International Journal of Computer Trends and Technology*, vol. 50, no. 2, pp. 114–121, Aug. 2017, doi: 10.14445/22312803/IJCTT-V50P120.
15. N. Lessio and A. Morris, “Toward Design Archetypes for Conversational Agent Personality”, *Conf Proc IEEE Int Conf Syst Man Cybern*, vol. 2020-October, pp. 3221–3228, Oct. 2020, doi: 10.1109/SMC42975.2020.9283254.
16. A. Følstad, M. Skjuve, and P. B. Brandtzaeg, “Different chatbots for different purposes: Towards a typology of chatbots to understand interaction design”, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11551 LNCS, pp. 145–156, 2019, doi: 10.1007/978-3-030-17705-8_13.
17. S. Hobert, A. Følstad, and E. L. C. Law, “Chatbots for active learning: A case of phishing email identification”, *International Journal of Human Computer Studies*, vol. 179, Nov. 2023, doi: 10.1016/J.IJHCS.2023.103108.
18. F. D. Davis, “Perceived usefulness, perceived ease of use, and user acceptance of information technology”, *MIS Quarterly*, vol. 13, no. 3, pp. 318–339, 1989, doi: 10.2307/249008.
19. M. Turner, B. Kitchenham, P. Brereton, S. Charters, and D. Budgen, “Does the technology acceptance model predict actual use? A systematic literature review”, *Inf Softw Technol*, vol. 52, no. 5, pp. 463–479, May 2010, doi: 10.1016/j.infsof.2009.11.005.
20. P. Kowalczuk, “Consumer acceptance of smart speakers: a mixed methods approach”, *Journal of Research in Interactive Marketing*, vol. 12, no. 4, pp. 418–431, Oct. 2018, doi: 10.1108/JRIM-01-2018-0022.
21. 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, doi: 10.2307/30036540.
22. S. Alharbi, “Trust and acceptance of cloud computing: A revised UTAUT model”, *Computational Science and Computational*, 2014, Accessed: Aug. 22, 2016. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6822317
23. 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, Jun. 2019, doi: 10.1007/S10796-017-9774-Y.
24. L. Labadze, M. Grigolia, and L. Machaidze, “Role of AI chatbots in education: systematic literature review”, *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, p. 56, Oct. 2023, doi: 10.1186/s41239-023-00426-1.
25. L. Lengel, Robert, H. and Daft, Richard, “The Selection of Communication Media as an Executive Skill”, *Academy of Management Perspective*, vol. 2, no. 3, pp. 225–232, 1988.

26. V. C. Sheer, "Media Richness Theory", *The International Encyclopedia of Media Psychology*, pp. 1–14, Sep. 2020, doi: 10.1002/9781119011071.IEMP0118.
27. S. Gabrielli, S. Rizzi, S. Carbone, and V. Donisi, "A chatbot-based coaching intervention for adolescents to promote life skills: Pilot study", *JMIR Hum Factors*, vol. 7, no. 1, Mar. 2020, doi: 10.2196/16762.
28. S. Gabrielli, S. Rizzi, S. Carbone, and V. Donisi, "A Chatbot-Based Coaching Intervention for Adolescents to Promote Life Skills: Pilot Study"., *JMIR Hum Factors*, vol. 7, no. 1, p. e16762, Feb. 2020, doi: 10.2196/16762.
29. S. Herkama and C. Salmivalli, "KiVa antibullying program", *Reducing Cyberbullying in Schools: International Evidence-Based Best Practices*, pp. 125–134, Jan. 2017, doi: 10.1016/B978-0-12-811423-0.00009-2.
30. C. F. Garandeau, E. Poskiparta, and C. Salmivalli, "Tackling acute cases of school bullying in the KiVa anti-bullying program: A comparison of two approaches", *J Abnorm Child Psychol*, vol. 42, no. 6, pp. 981–991, 2014, doi: 10.1007/S10802-014-9861-1.
31. Y. You, C. H. Tsai, Y. Li, F. Ma, C. Heron, and X. Gui, "Beyond Self-diagnosis: How a Chatbot-based Symptom Checker Should Respond", *ACM Transactions on Computer-Human Interaction*, vol. 30, no. 4, Sep. 2023, doi: 10.1145/3589959.
32. D. B. Jackson and M. G. Vaughn, "The bully-victim overlap and nutrition among school-aged youth in North America and Europe", *Child Youth Serv Rev*, vol. 90, pp. 158–165, Jul. 2018, doi: 10.1016/j.childyouth.2018.05.027.
33. A. Cosma *et al.*, "Bullying victimization: time trends and the overlap between traditional and cyberbullying across countries in Europe and North America", *Int J Public Health*, vol. 65, no. 1, pp. 75–85, Jan. 2020, doi: 10.1007/s00038-019-01320-2.
34. J. W. Creswell and V. L. P. Clark, *Designing and conducting mixed methods research*. Sage publications, 2017.
35. A. Yoachim, E. T. Welty, and N. Jenisch, "Public Interest Design, Pragmatism, and Potentials in a Postdiluvian City", *Journal of Architectural Education*, vol. 74, no. 1, pp. 110–118, 2020, doi: 10.1080/10464883.2020.1693834.
36. D. Chiffi and M. Andreoletti, "Abduction in Prognostic Reasoning", *Handbook of Abductive Cognition*, pp. 1–23, 2022, doi: 10.1007/978-3-030-68436-5_11-1/COVER.
37. D. Walton, *Abductive reasoning*. University of Alabama Press, 2014.
38. H. C. (Oliver) Chan and D. S. W. Wong, "Coping with cyberbullying victimization: An exploratory study of Chinese adolescents in Hong Kong", *Int J Law Crime Justice*, vol. 50, pp. 71–82, Sep. 2017, doi: 10.1016/j.ijlcj.2017.04.003.
39. A. Hevner *et al.*, "Design science in information systems research", *MIS quarterly*, vol. 28, no. 1, pp. 75–105, 2004, doi: 10.2307/25148625.
40. R.; Baskerville *et al.*, "Design science research contributions Design Science Research Contributions: Finding a Balance between Artifact and Theory", *J Assoc Inf Syst*, vol. 19, no. 5, pp. 358–376, 2018, doi: 10.17705/1jais.00495.
41. A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research", *MIS Q*, vol. 28, no. 1, pp. 75–105, 2004, doi: 10.2307/25148625.
42. M. L. Ranney *et al.*, "Cyberbullying Prevention for Adolescents: Iterative Qualitative Methods for Mobile Intervention Design"., *JMIR Form Res*, vol. 5, no. 8, p. e25900, Aug. 2021, doi: 10.2196/25900.
43. B. Zhang, J. Luo, and J. Li, "Moving beyond Likert and Traditional Forced-Choice Scales: A Comprehensive Investigation of the Graded Forced-Choice Format", *Multivariate Behav Res*, vol. 59, no. 3, pp. 434–460, 2024, doi: 10.1080/00273171.2023.2235682.

44. B. Çetin, E. Yaman, and A. Peker, “Cyber victim and bullying scale: A study of validity and reliability”, vol. 57, no. 4, pp. 2261–2271, Dec. 2011, doi: 10.1016/J.COMPEDU.2011.06.014.
45. J. Rose and C. W. Johnson, “Contextualizing reliability and validity in qualitative research: toward more rigorous and trustworthy qualitative social science in leisure research”, *J Leis Res*, vol. 51, no. 4, pp. 432–451, Aug. 2020, doi: 10.1080/00222216.2020.1722042.
46. N. Humaidi and V. Balakrishnan, “Exploratory Factor Analysis of User? s Compliance Behaviour towards Health Information System? s Security”, *Journal of Health & Medical*, 2013, Accessed: Aug. 22, 2016. [Online]. Available: http://www.omicsonline.org/exploratory-factor-analysis-of-user-compliance-behaviour-towards-health-2157-7420.1000123.php?aid=11780
47. J. M. Conway and A. I. Huffcutt, “A Review and Evaluation of Exploratory Factor Analysis Practices in Organizational Research”, Apr. 2003. doi: 10.1177/1094428103251541.
48. M. Schrepp, “On the Usage of Cronbach’s Alpha to Measure Reliability of UX Scales”, *J Usability Stud*, vol. 15, pp. 247–258, 2020.
49. E. Adamopoulou and L. Moussiades, “Chatbots: History, technology, and applications”, *Machine Learning with Applications*, vol. 2, p. 100006, Dec. 2020, doi: 10.1016/J.MLWA.2020.100006.
50. A. Hevner and S. Chatterjee, “Design Science Research in Information Systems”, 2010, pp. 9–22. doi: 10.1007/978-1-4419-5653-8_2.