Journal of Innovative Construction and Petrochemical Technologies, Vol.3, No.2 (December, 2026), 537-557

P-ISSN: xxxx- xxxx | E-ISSN: 3007-9861 | DOI: https://doi.org/10.71285/icpt.v3i2.81

Journal of Innovative Construction and Petrochemical Technologies
https://icpt.ouk.kz/index.php/icpt/index



Research Article

Check for updates

Cybersecurity Awareness, Confidence Calibration, and Simulated Phishing Detection Performance Among Undergraduate Students : A Cross-Sectional Study

Aseel Hamoud Hamza[1]*
College of Women's Sciences, Department of Computer Science, University of Babylon, Babylon, Iraq
aseel.hamod@uobabylon.edu.iq

Suhad Al-Shoukry[2]
Department of Software and Artificial Intelligence Engineering Techniques Najaf Polytechnic College, Al-Furat Al-Awsat Technical University, Najaf, Iraq
inj.suhadaalzhra2010@atu.edu.iq

Zalili Binti Musa[3]
Faculty of Computing, University Malaysia Pahang Al-Sultan Abdullah, Pekan 26600, Malaysia
zalili@umpsa.edu.my

Abstract: The aim of this study was to examine how well students in a large public university could recognize phishing messages from non-phishing messages. It also looked at whether there was any relationship between students' reported confidence in recognizing phishing messages, their actual performance on the message recognition task, and their overall knowledge about cybersecurity. This research will help us understand what we can do to increase students' ability to identify phishing messages, which are often used by hackers to steal personal information. Objective: The aim of this research paper was to assess cybersecurity awareness, students' perceived levels of confidence regarding their ability to protect themselves online, student-reported cybersecurity behaviors (protective practices), student performance on a simulated phishing detection task, and the relationship between students' perceived confidence level and actual performance on the simulated phishing detection task (confidence calibration) for undergraduate students at the University of Al-Qadisiyah, Iraq. The research design for this study was a cross - sectional study that included 100 third-year undergraduate students representing various academic disciplines. Information was obtained via a structured survey that assessed the student's level of awareness regarding cyber security; confidence in performing tasks related to protecting themselves online; and actual behaviors used by the students to protect themselves online. In addition to the survey, each participant completed a classification task consisting of ten simulated emails (five legitimate emails and five simulated phishing emails). Aggregated study records were then analyzed using descriptive statistics, classification performance measures, and exploratory confidence-calibration metrics. Results: Mean cybersecurity awareness scores were 3.78 on a 5-point scale, and mean general confidence scores were 3.77. Overall accuracy for 1,000 classification emails was 73.0%. Phishing sensitivity was 67.8%, specificity was 78.2%, and F1-score was 71.5%. Mean decision-level confidence was 63.75%, whereas observed accuracy was 73.0%. Expected calibration error was 16.25 percentage points below recorded confidence coding. Conclusion: Self-reported cybersecurity awareness among students was relatively high; however, their ability to correctly identify simulated phishing messages was low. These results suggest that there is a need for student-focused practical detection exercises, confidence assessments and specific types of cybersecurity education. Due to the confidence-scale coding limitations, the calibration findings should be interpreted with caution.

Keywords: Cybersecurity Awareness; Phishing Detection; Confidence Calibration; University Students; Simulated Phishing; Cybersecurity Education.

Copyright © Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

1. Introduction

As technology is being increasingly integrated within higher education, students have been exposed to more cyber security threats, most notably phishing attacks. Phishing attacks take advantage of the human factor of deception in communication to elicit sensitive information or promote unsafe behaviors. that the role of human factors, environmental influences, and cyber security awareness in determining how susceptible an individual will be to a phishing attack [1], [2]. While cyber security awareness is necessary to protect oneself digitally, self-reported awareness does not always translate into being able to identify phishing attempts. Students can possess knowledge about the concepts of cyber security but still experience difficulty identifying fraudulent emails and/or communications from those that are legitimate. Performance based assessment, therefore, is crucial in assessing the ability of a student to detect fraudulent emails and/or communications beyond what they believe they are prepared to do [3], [4].

Confidence calibration angle to view the relation between subjective confidence of people and the objective validity of their decisions. When we look at confidence in conjunction with detection performance we can see if there are gaps between how well people think they are doing and what they have actually demonstrated as far as their cybersecurity knowledge and skills go [5], [6]. Although there is a significant amount of research being done on phishing awareness and human behaviors related to security, more empirical data needs to be gathered that combines awareness about cybersecurity, people's self-assessed levels of confidence in their own abilities, the types of protective measures people use, and how well people do in simulations where they must identify phishing attempts, all within a group of individuals from one university [7], [8]. This cross-sectional study will assess these four areas among 100 third-year students of Al-Qadisiyah, Iraq, with particular emphasis on assessing students' ability to detect simulated phishing emails and calibrating students' confidence in their decision-making processes for identifying phishing emails, so that we can create evidence-based educational materials for improving cybersecurity.

Human Factors and Phishing Susceptibility

As well as technology, phishing is a very large security issue for organizations due to the fact that phishing's success relies heavily upon the user's cognitive abilities, behavioral responses, and decision-making processes. Desolda et al. [1] the literature on individual awareness, knowledge, and psychological characteristics relating to users' vulnerability to phishing attacks. Desolda et al.'s results suggest that technological aspects of vulnerability to phishing attacks do not explain the full extent of users' vulnerabilities to phishing attacks. Similarly, Zhuo et al. [2] developed a human-centered framework to understand phishing susceptibility across multiple levels of users' interaction. The human-centered framework developed by Zhuo et al. emphasizes the role of individual and contextual variables when determining how users will react to deceptive communication. Naqvi et al. [9] phishing mitigation techniques and concluded that relying solely on technology-based defenses is insufficient and therefore suggested that organizations should integrate human capacity into their security defense strategy.

The literature also discusses the evolving role of generative AI in phishing and human vulnerability, although the present simulation was not designed to isolate AI-generated content.

More recently, [7] studied how phishing research has developed over time using bibliometric analysis, which identified human-related vulnerabilities as a significant area of research. All of these studies together studying user behavior and awareness to understand the overall readiness of users to protect themselves from phishing-based cyber threats. Additional empirical studies have explored the factors related to phishing vulnerability in various samples [10]. These studies have provided further evidence that there are differences in the relationships between phishing susceptibility and different types of predictors; specifically, some are universal while others depend upon the context. However, simply identifying what makes individuals susceptible to phishing attacks does not guarantee that individuals will be able to tell the difference between malicious communications and legitimate communications. Therefore, it may be beneficial to integrate users' subjective assessments of their cybersecurity skills with users' performance-based assessments of their cybersecurity skills. In addition to reviewing previous research, examined both individual and contextual factors that influence phishing susceptibility among a group of participants from Portugal. The results indicated that the relationship between perceived ability to detect phishing scams and online habits is complicated and depends upon the context,

indicating that people should not assume that they are safe from phishing scams simply because they have experience with technology in general.

The scope of the types of messages that users can receive through a generative-AI supported phishing attack has expanded as a result of recent research on the subject. [11] studied the phishing attacks conducted using generative AI and found that the impact of the human factor remains relevant; therefore, it will be necessary to examine various forms of deception when evaluating these types of attacks in the future.

Cybersecurity Awareness and Protective Practices Among University Students

Cybersecurity awareness includes how people perceive digital risks; their ability to identify suspicious messages and take actions to protect themselves from cyberattacks. Awareness and the actual practice of cyber safety are two separate concepts. Ismail et al. [12] studied information security awareness regarding phishing among IT students, and found that education can help prepare them to avoid becoming victims of online attacks. Gwenhure [13] studied university students' practices for avoiding email phishing depend upon the Health Beliefing Model, finding that both psychological and behavioral variables are important for explaining why people engage in protective behaviors. Shahbazi et al. [14] studied students' awareness levels about cyber security, along with their performance in detecting AI-based phishing attempts, showing an increasing focus on developing human-centered assessments of cybersecurity capabilities as well as technological solutions to address security issues. Additionally, Mossano and Volkamer [15] reviewed common misconceptions about phishing, indicating that users may have limited or incorrect knowledge about deceptive communications.

The studies mentioned above show how significant understanding and perceiving cybersecurity is. On the other hand, the real that a student reports they are aware do not meant they are competent in applying what they know. A student might know about the basic rules for being secure online, yet they could have trouble using the rules when trying to figure out if an unfamiliar email is legitimate or a trick. There have been other similar studies that used simulations as analyzed the effects of various types of training in different environments [8], [16], [17]. Although the findings of these studies help establish the validity of the current study's measurement tool, they do not validate it. Protecting yourself from scams online goes beyond just knowing when an email looks suspicious. It also includes taking action independently to confirm whether a request made in an email was legitimate or not; protecting your login information; downloading attachments correctly; and reporting potential scams. So, if we look at both knowing about the threats to online security as well as taking actions to protect yourself from them, we will get a better idea of how prepared students are for dealing with security issues. Mungo [18] studied self-paced online education on phishing awareness for retail employees. This is a good example of applied training being conducted in a workplace environment rather than at a university. Although the results of such studies can inform instructional design, we cannot assume that the results would be applicable to undergraduate students.

Simulated Phishing Detection and Performance-Based Assessment

Performance-based phishing assessments is the way to assess an individual's capability to differentiate between genuine emails and fraudulent emails. While questionnaires are used to assess what individuals believe they know, performance-based assessments can help determine how well individuals make decisions about whether or not to open an email based upon its content. For example, Gallo et al. [3] studied user behavior while reading phishing emails and concluded that assessing users' behaviors when dealing with deceptive communications is as important as measuring users' perceptions of knowledge. This finding supports the idea that assessing users' behaviors can be a useful tool for understanding why individuals fall victim to phishing attacks. Additionally, Singh et al. [5] studied cognitive factors associated with learning and differentiation in anti-phishing training programs. The authors found that training users to differentiate between legitimate and fraudulent messages was more effective at reducing phishing susceptibility than simply increasing users' overall levels of suspicion. Finally, the characteristics of the messages themselves can affect users' ability to detect them. Tian et al. [19] evaluated how different influence strategies were used in phishing attacks across various organization types. The results refer the success rate of various influence techniques varied significantly across organizational types.

Phishing-detection evaluations need to assess several performance metrics. Sensitivity

assesses what percentage of phishing emails are correctly identified by the system, while specificity assesses what percentage of legitimate emails are correctly identified. Precision and F1-score are two other performance metrics that help evaluate how well phishing classifications are made. Assessing all four performance metrics will give a better picture of the performance of the system than just using overall accuracy as a metric. This is especially true if there are different risks associated with incorrectly identifying a phishing email as legitimate versus incorrectly identifying a legitimate email as a phishing email. Alsaqer et al. [16] developed a customizable platform that allows users to create simulations of phishing attacks and analyze user behaviors. Their work demonstrates how simulated communications can be used for studying aim ; however, they evaluated their simulated platform using simulated data rather than observing actual user behavior in a real-world setting.

Self-Reported Confidence and Confidence Calibration

Confidence refers to a user's subject assessments of the accuracy of his/her evaluation. Confidence can affect the way users respond to phishing messages; it can determine if they will trust the message, verify the sender, or view the message as suspicious. There is a clear difference between users' self-reported overall confidence in their ability to detect phishing, and users' decision-specific confidence that relates directly to the classification they make. Self-reported overall confidence is a user's general perception of his/her ability to identify phishing messages. On the other hand, decision-specific confidence is a user's confidence in the classification he/she makes for each message. Calibration refers to the relationship between a user's subjective confidence and the actual accuracy of his/her evaluations. A well-calibrated user has confidence levels that accurately reflect the percentage of correct evaluations he/she makes. Conversely, a poorly-calibrated user consistently has confidence levels that are higher or lower than the actual accuracy of his/her evaluations. Research on misconceptions and cognitive aspects of phishing provide a reason to consider subjective evaluations along with performance. Mossano and Volkamer [5] studied misconceptions that may impact how users perceive threats posed by phishing. Singh et al. [15] studied the cognitive processes that users undergo to learn how to distinguish between malicious and legitimate communications.

However, researchers' findings related to awareness of cyber security issues or perception of ability cannot be used as direct evidence for calibrating confidence. To calibrate confidence, it is necessary to assess the degree of certainty a person expresses when making a decision and then compare this assessment to whether the decision was correct or incorrect. Research has also investigated the use of self-paced training programs to educate employees about phishing attacks in the work place [18] however, large scale experimental studies have shown that learning outcomes depend upon the context of the experiment [20].The distinction is particularly important for university students as there does not appear to be a consistent relationship between what they believe is their level of knowledge about cyber security threats and how well they perform in simulated phishing experiments. Therefore, integrating decision-level confidence measures with objective classification results could provide another exploratory view of how prepared students are for cyber security threats. Researchers have recently looked at adapting phishing education to meet the needs of the learner. Kuem et al. [21] conducted three experiments to investigate the relationships between performance facilitators and message-specific coping responses. This study demonstrates the importance of feedback, incentives, and task context in determining training performance. Malloy et al. [22] studied a cognitive large language model approach for choosing anti-phishing examples and creating feedback. They reported positive learning outcomes in their experimental study. Although these intervention studies suggest future areas of investigation, they do not establish that any of the interventions are effective in the current sample.

Cybersecurity Training and Educational Interventions

Cybersecurity education is seen as a key part of preventing phishing. The success of any educational program will depend upon the content, delivery method, how relevant it is to the person receiving the education, and what form of evaluation is used. Hillman et al. [23] studied how organizations use phishing awareness training on a large scale. They showed how critical it is to evaluate the effects of the training based on what happens when employees go back to work. Singh et al [5] looked at the mental processes involved in anti-phishing training. They found that learning and getting feedback are very helpful for employees to be able to tell the difference

between real emails and fake ones. Yin et al. [24] ran a study where they changed when employees received feedback after a fake email simulation. They found that the way feedback is designed makes a big difference in how well employees learn. Morić et al. [8] ran a similar study at a Croatian university. They used both simulations of phishing emails and educational interventions. They found that the best way to help employees develop resistance to phishing is to make sure the education relates to their daily lives. On an institutional level, Merritt et al. [25], in guidelines from the National Institute of Standards and Technology, said that cybersecurity learning programs need to focus on producing measurable results and changing employee behavior instead of just giving employees information.

The findings across the three articles imply that effective cybersecurity education includes use of realistic phishing scenarios, a process for providing structured feedback to learners, and methods to assess demonstrated learning. While associations between prior training and performance in detecting attacks in observational studies do not demonstrate a causal relationship, there may be other differences between those who have received prior training and those who have not. Prümmer et al. [17] reviewed common approaches to cybersecurity education and Prümmer et al. [26] conducted a meta-analysis of the effects of cybersecurity education on end users. Both of these publications provide insight into how to evaluate educational outcomes and emphasize that performance changes are not synonymous with increased awareness. Gonzalez-Jimenez et al. [20] also examined cybersecurity learning through a large-scale field experiment with small and medium-sized enterprises. The results from this study can help inform the development of future field experiments, however, the experimental design was different than that of the university simulation described above.

Research Gap and Contribution of the Present Study

The literature reviewed has demonstrated that there has been significant development in our understanding of phishing vulnerability, human elements related to cybersecurity awareness, educationally focused programs and interventions. A number of studies that were systematic reviews have identified multiple factors that contribute to a person's vulnerability to phishing attacks both individually and contextually. Many additional studies have examined how individuals react to phishing attempts, incorrect beliefs about phishing, educational methods for improving awareness of cybersecurity issues and the effectiveness of performance-based learning. While the studies reviewed in this literature provide valuable insights into these areas, they also differ from one another in terms of population, methods used to measure variables and major findings. Thus, additional contextually-specific research should be conducted to evaluate the level of cybersecurity awareness among university students, their perception of their own ability to identify phishing emails, their utilization of protective behaviors, their performance in identifying simulated phishing emails, and whether they are able to accurately estimate their level of confidence in making decisions about whether an email is legitimate or malicious.

The current research provides a solution to the research gap created by the lack of investigation into the four constructs identified in the previous section. To provide evidence for the validity of these constructs in the context of higher education in Iraq, the study uses a sample of 100 third-year students from the University of Al-Qadisiyah. This study has several contributions. It assesses both student confidence level and students' knowledge/awareness of cyber threats. It also assesses student behavior related to protection of personal data and security. The study evaluates student's ability to identify whether a received email is a phishing message or a legitimate communication using five performance measures; accuracy, sensitivity, specificity, precision and F1-score. The study examines the relationship between confidence levels during the process of making decisions on emails and accuracy in classification. This provides an evaluation of student preparedness based upon their confidence levels, rather than just their reported preparedness. The study does not claim that the four constructs examined here have not been studied before. However, the study does contribute empirical evidence of the relationship between the four constructs based on an integrated assessment of the four constructs in a higher educational environment in Iraq. The results of the study may lead to the creation of curriculum for cybersecurity education that would include teaching practical recognition skills, independent verification of received information, reporting procedures, and correct interpretations of uncertainty.

2. Materials and Methods

Study Design

This is a cross-sectional observational assessment that uses a self-report questionnaire and a simulated email-classification task with third year undergraduate students. The questionnaire measures perceived awareness of cybersecurity issues, levels of general confidence and protective behaviors, while the classification task assesses students' ability to identify both phishing and legitimate email scenarios. All results described below were derived from the retained aggregate study tables. In addition to assessing the accuracy of each individual email classification, student confidence at the decision-making stage was also assessed using confidence category summaries. This research design provides for descriptive comparisons between variables, however, it cannot provide evidence for causal relationships.

Study Setting and Participants

The sample of 100 third year undergraduate students from the University of Al-Qadisiyah (Iraq), was drawn from several disciplines including: Medical and Health Sciences; Engineering and Technology; Natural and Applied Sciences; Humanities and Social Sciences; Education; Agriculture; Law / Administration / Economics; and other disciplines. Of these 100 individuals, 56 were males, 42 were females and 2 chose to remain anonymous with respect to their gender. Age categories were defined as follows: 18 - 20; 21 - 23; 24 - 26; and 27 years old and above. Data were gathered through direct contact with each individual student on-site at the University of Al-Qadisiyah during September 2026. Researchers provided the questionnaire to each individual student on campus. Each student completed the questionnaire and returned it to the researcher. The resulting aggregated data set consisted of 100 third year undergraduate students. Since the records pertaining to recruitment and the completed questionnaires are no longer available, there is no way for the researchers to verify how many students were contacted, the non-response rate and the exact dates when the data collection occurred.

Research Instrument

The original questionnaire and scenario-based assessment had 5 components that are identified by the retained summary tables. The original paper questionnaire and completed forms were not retained so we can only reconstruct the items at the construct/topic level using the summary tables. We cannot verify the exact wording of questions, response options, how the instrument was developed, pilot tested, and whether it was validated.

Section A: Demographic and Educational Characteristics

The first part of the study gathered demographic and background information about the participants, such as their age, gender, field of study (academic discipline), prior formal education in cybersecurity, prior instruction on how to recognize phishes, and how often they used electronic mail.

Section B: Cybersecurity Awareness

The assessment of student's level of knowledge regarding cyber security included evaluating the student's perception of their ability to recognize phishing emails, deceitful email sender information, suspicious email attachments, emergency requests for credentials (passwords), disclosures of passwords, and reporting processes.

The provided workbook contains the average (mean) value and the variability (standard deviation) for each individual item using a five-point scale. It was not possible to determine what the initial words used as responses to the questionnaire were because the questionnaire is no longer accessible.

Section C: General Self-Reported Confidence

Three questionnaire items measured students' self-reported ability to differentiate between legitimate communication and phishing attempts, evaluate unsolicited e-mail messages, and identify e-mail messages that appear to come from the university.

These items assessed an individual's overall perception of their own ability as opposed to how confident they were about making specific email classification choices.

Section D: Protective Cybersecurity Practices

Six questions from the questionnaire measure what respondents do when they take steps to protect themselves on-line. These include confirming who is sending messages; independently verifying suspicious requests; dealing with unexpected attachments; avoiding entering credentials on unsecure sites; reporting suspicious emails; and evaluating urgent messages.

Section E: Simulated Email-Classification Assessment

The performance-based assessment included ten simulated emails, with five of those being phishing attempts and five being legitimate emails.

Each scenario was created with an existing "correct" answer key. This means that instead of producing data on how users responded to live phishing attempts, the study produced five simulated phishing messages and five legitimate messages for each participant.

Measurement of Study Variables

The unweighted mean of all ten of the reported item means was used to create an overall awareness measure. In addition, the unweighted mean of the three reported item means was used to represent overall confidence. Protective practice measures were developed by creating individual measures of each of the six reported item means and standard deviations. It is important to note that the above averages do not represent participant composite scores, since these types of scores can only be created when individual responses are available. Accuracy performance was created based on the total number of correctly identified classifications and the total number of incorrectly identified classifications. Classification decisions (rather than student observation) served as the basis for calculating overall accuracy and class specific metrics. Confidence in the classification decision was measured with five ordered categories that were numerically coded (0%, 25%, 50%, 75%, and 100%). Descriptive labels for each of these categories can be found in section 3.7. There was no assumption made regarding the validity of the confidence-category value as a probability scale.

Simulated Phishing-Detection Procedure

There are 1,000 decisions that have been made by 100 students and these decisions are based on the ten simulated messages per student. Of these 1,000 decisions, 500 were about phishing and 500 were about legitimate messages. After evaluating each decision using its respective scenario answer key, the total aggregate count of responses for each decision were then used to develop a confusion matrix and calculate the proportion of responses that were correct for each scenario. These decisions are grouped together by participants. If there is no participant-level data available, we cannot estimate the variance in decisions among participants nor can we take into account the fact that some participants may have made multiple decisions. The way in which the study was administered, the order of presentation of scenarios, the amount of time given to complete the tasks, and whether or not there were randomizations involved in the administration of the study are not documented in the retained records and will need to be verified through the use of the original protocol.

Classification Performance Measures

Overall accuracy, sensitivity (phishing recall), specificity, precision, and phishing-class F1-score were derived from the aggregated confusion matrix, with phishing designated the positive class. Balanced accuracy was defined as the mean of sensitivity and specificity.

Phishing messages were designated as the positive class, and legitimate messages as the negative class.

True positives (TP) represented phishing messages correctly classified as phishing. True negatives (TN) represented legitimate messages correctly classified as legitimate. False positives (FP) represented legitimate messages incorrectly classified as phishing, while false negatives (FN) represented phishing messages incorrectly classified as legitimate.

The metrics were calculated as follows:

Accuracy

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Sensitivity

$$\text{Sensitivity} = TP / (TP + FN)$$

Specificity

$$\text{Specificity} = TN / (TN + FP)$$

Precision

$$\text{Precision} = TP / (TP + FP)$$

F1-score

$$F1 = 2 \times (\text{Precision} \times \text{Sensitivity}) / (\text{Precision} + \text{Sensitivity})$$

All proportions use the relevant aggregate decision counts as their denominators. They are descriptive performance measures; the 1,000 decisions must not be treated as 1,000 statistically

independent participants.

Decision-Level Confidence Assessment

The decision level of certainty was represented through the use of an ordinal scale with five levels, "Pure guess," "Low," "Moderate," "High," and "Completely confident." The order of these categories was maintained in the retained workbook with the numerical codes of 0%, 25%, 50%, 75%, and 100%. The accuracy of each category was determined through computation of the percentage of correct classifications within each category. Computation of the average descriptive numerical confidence was accomplished using a weighted average based upon the number of decisions recorded for each of the categories. Computation of the expected calibration error (ECE) was accomplished through a weighted average of the absolute differences between the observed category accuracy and the numerical category code, weighted by the number of decisions made for each category. Computation of a Brier score was also accomplished using the same coded values as above along with binary correctness. These are all coding-dependent exploratory quantities, not validated probabilistic calibration estimates: in particular, "Pure guess" cannot be assumed to represent a 0% probability of being correct. Therefore, the primary result regarding respondent confidence is the descriptive pattern of observed accuracy across the ordered certainty categories; there is no claim that respondents generated well-calibrated subjective probabilities.

Statistical Analysis

In addition to the previous report, the workbook included four previously reported correlation coefficients with their associated p-values. However, the specific correlation estimator, as well as the original response-level observations used in the prior reports, are not included in the workbook. Sample characteristics and cybersecurity exposure were presented using frequencies and percentages. Reported item means and standard deviations were used to describe levels of awareness, overall confidence, and protective practices. Confusion matrix metrics and scenario level accuracy were generated based upon the aggregated decision counts. Confidence category accuracy and exploratory calibration values that depend on how respondents were coded into categories were generated using the category counts and correct responses.

In light of this, the associations being presented here are simply a summary of unverified historical data, rather than being based upon an independent replication of a significant test. Descriptive summaries also apply to the prior training groups means. Since there was no access to the original individual level response data, it was not possible to assess whether the scales were reliable (i.e. Cronbach's alpha), whether confidence intervals could be calculated taking into account the effects of clustering, whether multivariate regression analyses could be conducted, or whether there were statistically significant differences among various subgroups of participants. No inference is drawn from the total number of reconstructed decisions and no causal relationship is inferred from the observational design used in the study. All mean academic discipline performances that contained unresolvable contradictions were removed from the primary results of the correction.

Ethical Considerations

The researcher gave paper questionnaires to undergraduate students at the university during September 2026, and then collected the completed questionnaires in person. No formal research ethics review had been conducted prior to the distribution of these questionnaires, and no formal research ethics approval had been obtained. There is nothing in the records to indicate whether each student who completed a questionnaire was provided with an information sheet, a statement of voluntariness, or a clear and formal informed consent process. The mere fact that a student completed a questionnaire does not provide sufficient evidence that a formal informed consent process was used. Prior to submitting the results of this study for publication in a journal, the author should obtain a formal decision from either the appropriate research ethics committee or institutional research authority regarding the ethics of the study that has already taken place and what type of disclosure or subsequent action is required. If such a formal decision is made, it will need to be reported in the manuscript using its actual date and reference number, if a reference number is issued. It cannot be reported as having been approved in advance. At present, only aggregated data tables are retained for the analyses presented in this paper. Individual level data, as well as the original completed questionnaires, are no longer available. Therefore, the extent of anonymity and confidentiality protections offered to participants cannot

be determined retrospectively. Also, it cannot be asserted that there were particular procedures in place for managing the original data without supporting documentation.

3. Results

The descriptive results of the retained aggregate studies will be reported. The 1,000 email evaluations were made by 100 individuals and should not be considered 1,000 individual participants. The performance of classifications is determined directly from the contingency table, while the previously reported correlations and scores for the training group will be identified as unverified estimates based on the level of the participants.

Participant characteristics and cybersecurity exposure

The sample included 100 third-year university students. Students aged 21-23 made up the greatest portion of participants (49%); 56% identified themselves as male and 42% as female. Academic fields were most common in engineering and technology (18%). Formal cybersecurity training prior to participating was reported by 28%, and 50% reported using e-mail each day or multiple times per day (Tables 1 and 2).

Table 1. Demographic and educational characteristics (N = 100)

Characteristic	Category	n (%)
Age	18–20 years	38 (38%)
	21–23 years	49 (49%)
	24–26 years	9 (9%)
	≥27 years	4 (4%)
Gender	Male	56 (56%)
	Female	42 (42%)
	Prefer not to say	2 (2%)
Discipline	Medical/health sciences	15 (15%)
	Engineering/technology	18 (18%)
	Natural/applied sciences	14 (14%)
	Humanities/social sciences	15 (15%)
	Education	14 (14%)
	Agriculture	8 (8%)
	Law/administration/economics	12 (12%)
	Other	4 (4%)

Table 2. Prior training, phishing guidance, and email-use frequency (N = 100)

Variable	Category	n (%)
Formal training	Yes / No / Not sure	28 (28%) / 67 (67%) / 5 (5%)
Phishing guidance	Yes / No / Not sure	45 (45%) / 48 (48%) / 7 (7%)
Email use	Less than weekly	8 (8%)
	1–2 days per week	16 (16%)
	3–5 days per week	26 (26%)
	Daily	31 (31%)
	Several times per day	19 (19%)

Cybersecurity awareness, perceived confidence, and protective practices

The unweighted arithmetic mean across the ten awareness item means was 3.78/5. The highest item score concerned avoiding password disclosure by email (M = 4.30, SD = 0.80), whereas reporting suspicious email had the lowest awareness-item mean (M = 3.20, SD = 1.10). Mean general confidence across three item means was 3.77/5. Among protective practices, avoiding password entry through unexpected email links (M = 4.10, SD = 0.90) scored highest and reporting suspicious messages (M = 2.80, SD = 1.20) lowest (Table 3). These reported

aggregates are not participant-level composite scores.

Table 3. Awareness (B), general confidence (C), and protective practices (D)

Item	Indicator (abbreviated)	M	SD
B1	Understanding phishing	3.90	0.90
B2	Recognizing warning signs	3.70	1.00
B3	Misleading sender names	3.60	1.00
B4	Inspecting sender addresses	3.40	1.10
B5	Urgent credential requests	4.20	0.80
B6	Unexpected attachment risks	4.00	0.90
B7	Professional-looking phishing	3.80	0.90
B8	Reporting suspicious emails	3.20	1.10
B9	Avoiding password disclosure	4.30	0.80
B10	University and personal account risk	3.70	1.00
C1	Distinguishing phishing and legitimate mail	3.90	0.90
C2	Judging unexpected emails	3.80	0.90
C3	University impersonation	3.60	1.00
D1	Checking sender address	3.30	1.10
D2	Independent verification	3.00	1.10
D3	Verifying unexpected attachments	3.80	0.90
D4	Avoiding credential entry after unexpected links	4.10	0.90
D5	Reporting suspicious emails	2.80	1.20
D6	Reviewing urgent information requests	3.70	1.00

Note. Item means and SDs are reproduced from the aggregated workbook; the arithmetic mean across awareness items is 3.78, and across general-confidence items is 3.77.

Overall simulated phishing-detection performance

Across ten scenarios, participants made 1,000 classifications, of which 730 were correct (73.0%). Among 500 phishing-message classifications, 339 were correct (sensitivity 67.8%); among 500 legitimate-message classifications, 391 were correct (specificity 78.2%). Precision was 75.67%, the phishing-class F1-score was 71.52%, and balanced accuracy was 73.0%. Sensitivity was 10.4 percentage points lower than specificity (Tables 4 and 5; Figure 1).

Table 4. Confusion matrix (phishing designated as the positive class)

Actual message	Predicted phishing	Predicted legitimate	Total
Phishing (positive)	339 (TP)	161 (FN)	500
Legitimate (negative)	109 (FP)	391 (TN)	500
Total	448	552	1,000

Table 5. Performance metrics derived from the confusion matrix

Metric	Value
Overall accuracy	73.00%
Phishing sensitivity / recall	67.80%
Specificity	78.20%
Precision	75.67%
Phishing F1-score	71.52%
Balanced accuracy	73.00%
Mean correct responses per participant	7.30 of 10

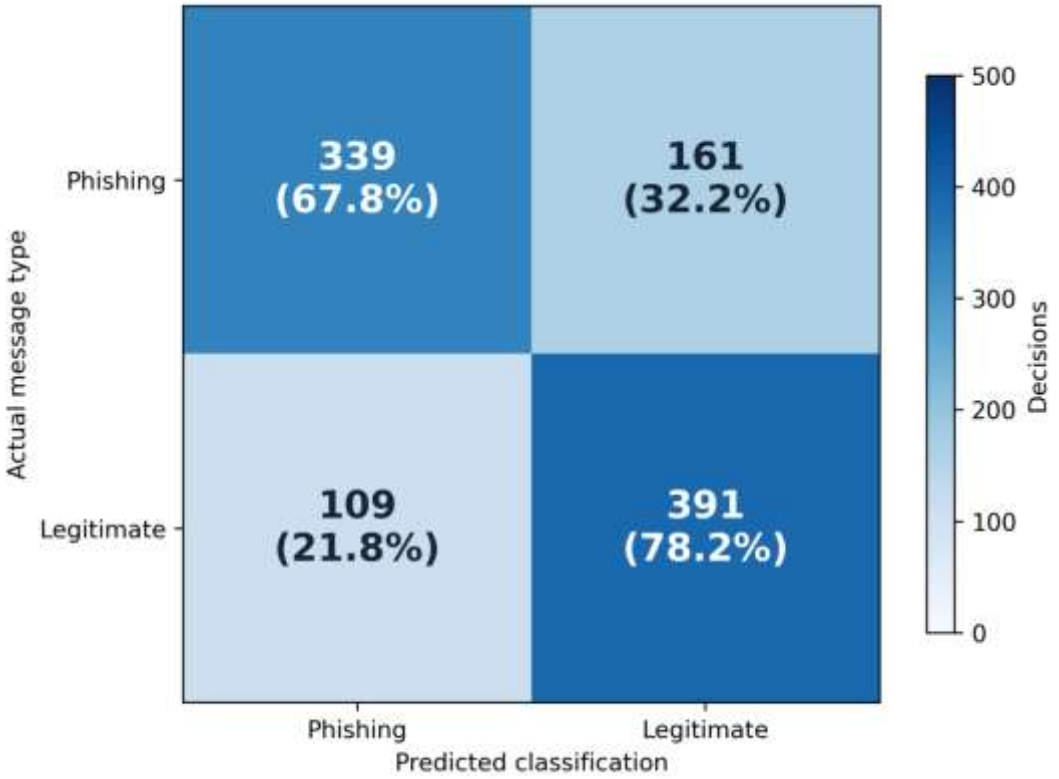

Figure 1. Confusion matrix of the simulated phishing-classification task. Percentages in cells use each actual-class total (n = 500) as denominator.

Performance across simulated email scenarios

Performance varied across the ten messages (Table 6; Figure 2). The highest correct-classification count was obtained for the phishing university-account warning (82/100), whereas the phishing financial-information request had the lowest (58/100). The five phishing scenarios summed to 339 correct judgments and the five legitimate scenarios to 391, agreeing with the aggregate confusion matrix. Message-level differences are descriptive; causal explanations of specific message features cannot be determined from these data.

Table 6. Correct classifications by simulated email scenario (100 decisions per scenario)

ID	Scenario	Answer key	Correct n (%)
E1	University account warning	Phishing	82 (82%)
E2	Lecture announcement	Legitimate	78 (78%)

E3	Library account update	Phishing	73 (73%)
E4	Book return reminder	Legitimate	75 (75%)
E5	Student prize notification	Phishing	61 (61%)
E6	Scheduled system maintenance	Legitimate	81 (81%)
E7	IT support message with attachment	Phishing	65 (65%)
E8	Student activity announcement	Legitimate	77 (77%)
E9	Financial information request	Phishing	58 (58%)
E10	Course file confirmation	Legitimate	80 (80%)

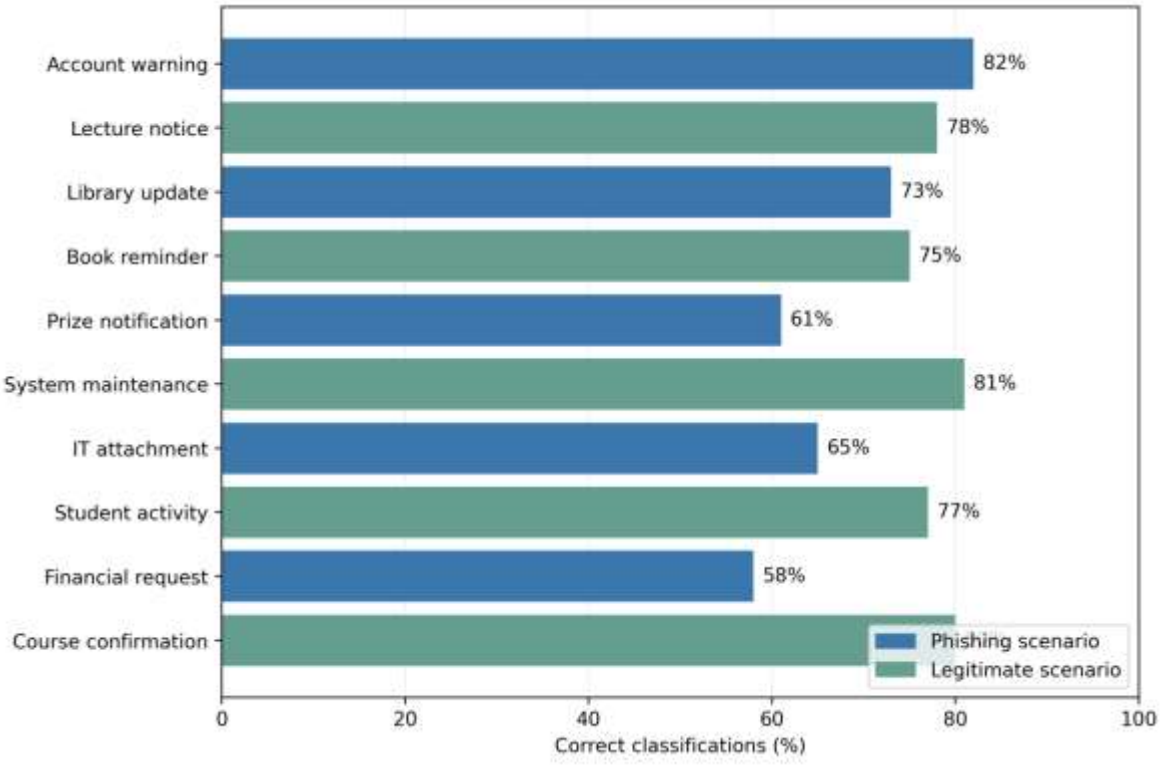

Figure 2. Percentage of correctly classified decisions for each scenario (n = 100 judgments per message).

Decision-Level Confidence and Classification Accuracy

Across 1,000 email-classification decisions, confidence was recorded in five ordered categories (Table 7; Figure 3). Observed accuracy increased from 40.0% for "Pure guess" (24/60) to 55.0% for "Low" (66/120), 70.0% for "Moderate" (175/250), 80.0% for "High" (280/350), and 84.09% for "Completely confident" (185/220). The average numerical confidence value of the weighted samples was 63.75 percent, while the actual accuracy rate of the total sample was 73.0 percent. The average coded confidence for those who correctly identified their samples was 68.36 percent, while the average coded confidence for those who incorrectly identified their samples was 51.30 percent. A notable aspect of this study is that there were 35 decisions made at the highest level of confidence that were incorrect. Using the coded numerical scale, the expected calibration error was 16.25 percentage points and the Brier score was 0.218875. This is an exploratory study: "pure guesses" were assigned a value of 0 percent, however, this does not imply that respondents believe they have zero percent chance of being correct. Therefore, the main result is a descriptive relationship between ordered confidence and classification correctness, not a validated probabilistic calibration.

Table 7. Recorded decision confidence and observed accuracy

Recorded confidence	Decisions	Correct	Observed accuracy
0% ("Pure guess")	60	24	40.00%
25% ("Low")	120	66	55.00%
50% ("Moderate")	250	175	70.00%
75% ("High")	350	280	80.00%
100% ("Completely confident")	220	185	84.09%
Total	1,000	730	73.00%

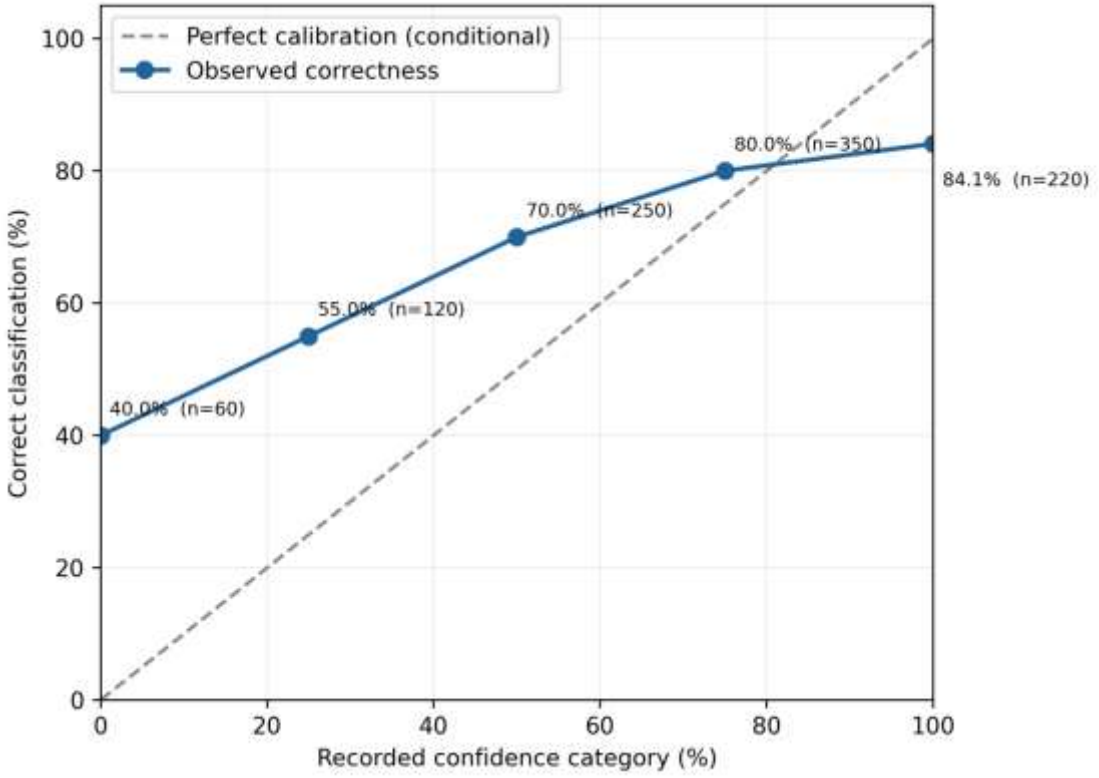

Figure 3. Observed classification accuracy across decision-level confidence categories.

The figure illustrates the proportion of correct email classifications within five ordered confidence categories. Classification accuracy increased with reported confidence; however, the categories should not be interpreted as a validated continuous probability scale. The numerical label of 0% for "Pure guess" reflects the original dataset coding rather than a confirmed subjective probability of correctness.

Previously reported bivariate associations

Table 8 shows that the source workbook retains four statistically significant relationships between study variables and outcomes. Although the original individual record data and which statistical test was used (i.e., Pearson r vs Spearman rho) are no longer available; the p-values and coefficient estimates cannot be independently calculated, and as such, only summary statistics can be reported.

Table 8. Previously summarized correlations (unverified from participant-level data)

Variable pair	Reported coefficient	Reported p
Awareness–detection accuracy	0.32	0.001
General confidence–detection accuracy	0.28	0.005

Awareness–general confidence	0.41	0.001
Protective practices–detection accuracy	0.24	0.016

Descriptive Detection Performance by Prior Formal Training

Participants who had received formal training in cybersecurity in the past had a mean detection score of 7.8/10 (SD=1.5, n=28) versus 7.1/10 (SD=1.7, n=67) for participants who did not report any training and 6.8/10 (SD=1.6, n=5) for participants who were unsure whether they had ever received training (Table 9). The weighted average of the rounded subgroup means is approximately 7.28/10, which is consistent with the direct calculation of the total score of 7.30/10 when rounded to one decimal place. These are descriptive group summaries; since the individual participant-level data were unavailable to conduct independent inferential analysis, we cannot make any conclusions regarding the causal effects of training. In addition, the academic field-specific detection scores were excluded from the comparison analysis due to inability to reconcile the weighted average of the academic field-specific detection scores (approx. 7.22/10) with the overall average of 7.30/10 using normal rounding rules.

Table 9. Reported mean performance by prior formal training

Training history	n	Mean (/10)	SD
Received formal training	28	7.8	1.5
No formal training	67	7.1	1.7
Not sure	5	6.8	1.6

4. Discussion

Principal Findings

The current research examined university-level undergraduate students' cybersecurity awareness, their self-confidence in their abilities regarding cybersecurity issues, their protective behaviors regarding cybersecurity issues, and how accurately they could identify whether a simulated email was legitimate or not. The results showed that the participants generally reported having positive views toward their cybersecurity awareness and self-confidence. Their average score for each item related to cybersecurity awareness was 3.78, and their average score for each item related to self-confidence was 3.77. However, the overall classification accuracy rate was 73.0%. Therefore, the participants made incorrect classification decisions for approximately one-fourth of all classification decisions. An additional significant finding was the disparity in participants' phishing sensitivities (67.8%) versus participants' specificity for identifying legitimate communications (78.2%). Additionally, participants' classification accuracy rates tended to increase as participants' self-confidence increased; however, participants still made incorrect judgments when they had high levels of confidence. In sum, the above-mentioned results demonstrate that it is essential to evaluate cybersecurity preparation using both self-reported measures and performance-based assessments. This multi-dimensional approach to evaluating cybersecurity preparation aligns with Desolda et al. as they state that people's cognitive processes, awareness, and behavioral characteristics are significant contributors to their susceptibility to phishing attacks. Additionally, Desolda et al. further state that it would be beneficial to examine users' responses to misleading email messages [1].

Cybersecurity Awareness and Simulated Phishing-Detection Performance

The mean cybersecurity awareness item score of 3.78 indicated that the respondents indicated a good understanding of phishing threats and common cybersecurity best practices. Although the respondents had high levels of self-reported cybersecurity awareness, they performed poorly when asked to distinguish between real and fake emails. This result relates to the differences in how we assess an individual's perceived readiness to defend against cyber attacks, compared to how well they can make decisions about whether or not to trust incoming messages. While survey-based assessments typically measure what people believe they know about cybersecurity and how prepared they feel, scenario-based assessments provide insight into how well people are able to use their knowledge of cybersecurity to evaluate different types of

communication. Desolda et al. [1] found that there are many factors involved in making someone susceptible to phishing, such as cognitive abilities and behaviors. In addition, Gallo et al. [3] emphasized the importance of assessing how users respond to emails containing different types of deception.

The present findings are consistent with the need for such complementary assessments. However, because awareness scores and classification accuracy were measured using different scales, their descriptive difference should not be seen as a direct indication that students were overconfident about what they could do. It would seem from these results that university cyber security education needs to go beyond teaching people about potential threats and provide them with hands on experience in recognizing potentially harmful messages. This finding also warrants comparison with Ribeiro et al. [10], where the relationship between perceived detection ability and phishing susceptibility was found to be more complicated. There are differences in the samples used by each study and the way in which variables were operationally defined so we cannot directly compare numbers; however, both studies indicate that using only self assessments as measures of practical defense is not recommended.

Recognition of Phishing and Legitimate Communications

Phishing sensitivity and specificity was found to have a 10.4-percentage-point difference in classification performance. In other words, participants correctly classified 67.8 percent of phishing emails as such; however, they were able to correctly classify 78.2 percent of non-phishing emails. This means that identifying fraudulent emails was more difficult than identifying non-fraudulent emails during the simulated testing process. The false-negative rate of 32.2 percent is defined as emails that were actually phishing emails but were misclassified as legitimate emails. Conversely, the false-positive rate of 21.8 percent is defined as legitimate emails that were misclassified as phishing emails. Both types of errors are critical when evaluating an individual's level of preparedness for cyber threats. On one hand, misclassifying phishing emails can result in individuals receiving fraudulent requests, and on the other hand, misclassifying legitimate emails can disrupt academic communication. However, the current classification study did not evaluate whether users clicked on links contained in the emails, provided their credentials after clicking on links, or exhibited any subsequent behaviors after being exposed to phishing attempts. Singh et al. [5] showed that how anti-phishing training is designed, specifically providing users with detailed feedback, affects whether or not users will be able to differentiate between phishing emails and legitimate emails. Singh et al.'s research supports the idea that organizations should focus on helping users improve their discrimination skills regarding electronic communication rather than simply instilling a general sense of distrust toward all forms of electronic communication. Therefore, the findings of the current study support education programs that help users learn how to better recognize malicious messages without compromising their ability to accurately identify legitimate institutional communications.

Variation in Classification Accuracy Across Email Scenarios

In examining each of the individual email conditions, we found some degree of variation in the participants' classification accuracy. One email condition, the university account warning, had the greatest success rate at classifying phishing emails correctly (82%). Conversely, one email condition, the financial information request, had the least success rate at classifying phishing emails correctly (58%). Other difficult email conditions for participants to identify as phishing include the student prize notification (61%) and the IT support message with an attachment (65%). This suggests that participants' classification ability varied depending on the type of simulated communication they were evaluating. There could be numerous explanations for this variation, such as message credibility, perceived familiarity with the sender, or the nature of the request. However, since there were no experimental manipulations of those variables in this study, it is impossible to determine precisely which variable(s) led to the differences in participants' performance. Gallo et al. [3] indicated that it would be beneficial to examine how users behave when exposed to phishing emails with varying characteristics and deceptive techniques. Our findings add descriptive support to the notion that all simulated phishing scenarios do not elicit the same level of recognition among members of a university community. Our findings also suggest that future educational exercises may benefit from identifying those categories of simulated phishing emails that require additional focus (e.g., financial requests, reward-related messages, and unexpectedly attached files).

Decision-Level Confidence and Classification Accuracy

An important aspect of the current research is to assess the relationship between how confident an individual feels when making a decision and the degree to which their decisions are correct. The number of correct decisions made consistently increased as individuals became more certain of their choices, ranging from 40.0% accuracy in the "Pure guess" category to 84.1% accuracy in the "Completely confident" category. This finding suggests there was a positive correlation between individuals' self-reported levels of confidence and the percent of times they correctly classified items. Nevertheless, individuals' reported certainty did not guarantee accuracy. For example, out of the 220 decisions that received the highest level of confidence, 35 were incorrect. This finding highlights the difference between an individuals' subjective certainty and demonstrating competency in cyber security. Confidence did provide a clue to an individuals' likelihood of being right; however, the fact that there were individuals who were wrong even though they had the highest level of confidence demonstrates that certainty alone cannot serve as evidence of an individual's reliability. Additionally, the structure of the original confidence scale also restricts formal probabilistic interpretations. Specifically, the category labeled "Pure guess" was numerically assigned a value of 0%, even though it may not necessarily represent a subjective probability of zero correctness. As such, the expected calibration error and Brier score cannot be considered definitive measures of probabilistic calibration.

The results of this study can be viewed as supporting the existence of a relationship between the ordering of the confidence categories and the accuracy of the classification. Confidence category ordering is related to the accuracy of the classification. Therefore, these results are best viewed as evidence that there is a relationship between the ordering of the confidence categories and the accuracy of the classification. Confidence categories are ordered in terms of the amount of confidence associated with each category. Confidence categories represent different levels of confidence in a given decision or judgment. Students who participate in simulated phishing exercises where they make a judgment about whether something is phishing or not, and then receive feedback regarding how well their judgment was accurate, will be encouraged to reflect on the certainty of their decision making. When combined with explanatory feedback, students can use such activities to learn from their mistakes and to understand when it would be appropriate to obtain further validation of a judgment. Singh et al. [3], demonstrated the importance of providing detailed feedback for developing discrimination skills during anti-phishing training. While Singh et al. [5] did not provide any evidence that the current study's confidence scale has been calibrated, the findings from this study do provide a compelling educational justification for developing feedback-oriented learning experiences.

Protective Cybersecurity Practices

The results of the self-reported protective behaviors indicated that there are two distinct types of protective behavior: avoiding potential risk by not entering passwords via unexpected email links (mean = 4.10) and not opening unexpected attachments prior to verification (mean = 3.80); and taking active steps to protect yourself from cyber attacks, such as independently verifying suspicious requests (mean = 3.00), and reporting suspicious emails (mean = 2.80). Therefore, the data indicates that respondents may engage in more protective avoidance behaviors than protective response behaviors. This distinction is significant since effective cybersecurity awareness involves not only recognizing potentially hazardous communications, but also responding appropriately when the legitimacy of those communications cannot be determined. Merritt et al. [25], in their guide to developing cybersecurity/privacy learning programs published by the National Institute of Standards and Technology (NIST), state that successful educational programs will develop measurable behavioral outcomes and continually evaluate program effectiveness. Our results indicate that university cybersecurity education programs might want to focus on providing students with easy-to-use mechanisms for reporting suspicious communications, developing guidelines for students to independently verify suspicious institutional communications, and educating students on what to do when they receive suspicious institutional communications. However, since we used self-reported questionnaire items to assess students' protective behaviors, these results cannot be viewed as direct evidence of how students will behave in actual cybersecurity incidents.

Implications for Cybersecurity Education in Higher Education

There are four main reasons as to how these findings can impact university level

cybersecurity education. Firstly, as noted previously, there was a positive response toward awareness of cybersecurity issues; however, there were also poor classification results. This means that although students are aware of cybersecurity issues, they do not necessarily understand them well enough to make accurate classifications. Therefore, the results support adding to theoretical instruction with practical assessment. Secondly, as noted above, students had low sensitivity to identify phishing messages. This implies that there is a need for students to develop better skills in recognizing fraudulent communication. However, it will be important to improve student's skill in identifying fraudulent communication without increasing their rejection of legitimate emails. Thirdly, as discussed earlier, there were variations in email scenarios. Therefore, it would be beneficial for educational programs to include various types of communication reflecting academic services, financial requests, rewards, attachments, and account management notifications in their training activities. Lastly, the finding regarding the correlation between confidence and correctness implies that confidence reflection could potentially provide another educational avenue for students. In addition to providing feedback on why students judged things correctly or incorrectly, confidence reflection could provide students with additional educational opportunities. Singh et al. [5] demonstrated the significance of detailed feedback to phishing discrimination. Additionally, Moric et al. [8] studied education and simulated phishing interventions in a Croatian university. Both studies support continuing to evaluate the efficacy of practical cybersecurity education in institutional settings.

These recommendations are also consistent with the behavior-oriented educational framework presented by Merritt et al. [25].

Nevertheless, the present study did not experimentally evaluate an educational intervention. Therefore, the proposed approaches should be regarded as recommendations for future implementation and evaluation rather than interventions whose effectiveness has been demonstrated by the current data.

The practical recommendations are consistent with the broader training literature. Prümmer et al. [26], reviewed training effects across different studies; Kuem et al. [21] explored how training facilitators and coping mechanisms influence outcomes. Malloy et al. [22] studied the effects of cognitive language model based feedback. Although these studies may offer potential interventions for this problem, they should not be viewed as showing that the effects identified in those studies will occur in the present study.

Strengths and Limitations

The study possessed many advantages in its methodology. It utilized two forms of measurement (questionnaire and simulated email classification) to assess cybersecurity knowledge. Both types of emails (phishing and non-phishing) allowed the researcher to compute various classification indicators such as sensitivity, specificity, precision, and F1-score. Additionally, the study provided information about the participants' level of confidence in making decisions about cybersecurity issues in an Iraqi university population, there are a few limitations that need to be addressed. the sample size was small (100) and consisted only of third year students at one university. This limits how generalizable the results can be to other universities, academic years, and populations. Second, the cross-sectional observational design do not agree for causal inferences regarding the relationship between awareness, confidence, protective practices, and detection performance. Third, there were only 10 simulated email scenarios used in the study. Therefore, those simulated scenarios may not adequately reflect the complexities, varieties, and contextual characteristics of phishing attacks that occur in actual settings. Fourth, most of the data collected was in aggregated form. As such, the researchers were unable to conduct independent participant level analysis or assess internal consistency of instruments used in the study. In addition, they were unable to verify if participants were consistent in their classification performance across multiple simulated email classification tasks. Also, since the researchers did not retain the original paper instrument nor completed questionnaires from each participant, they were unable to audit the precise phrasing of questions, original response options, and quality of participant-level data. Fifth, the confidence response categories were not developed as a valid probability based scale. Thus, any findings related to confidence should primarily be interpreted as ordinal associations with classification performance rather than definitive assessments of probabilistic calibration. Finally, due to a lack of documentation supporting

reconciliation of discipline-specific performance summaries with the overall classification performance results, comparisons based on disciplines were omitted from principal interpretation. Nevertheless, the study offers descriptive evidence regarding awareness of cybersecurity, simulated phishing detection performance, and decision-level confidence among undergraduates enrolled in an Iraqi higher education institution.

Directions for Future Research

Future research will need to study the awareness level of cyber security and the ability to detect phishing scams among a wider variety of college students from more universities. The use of multiple sites can help establish the validity of the results and allow researchers to compare the data collected at each site. It will be important for future researchers to collect and store the individual responses of each student so that they can conduct comprehensive reliability tests, use multivariable statistical techniques and use other methods to take into consideration the fact that each student may judge the same scenario multiple times. Future phishing assessments should include many types of simulated scenarios. Where it is acceptable to do so ethically, future phishing assessments should assess students' behaviors (actions taken after receiving a scam email) in addition to students' classification judgments (students' identification of a scam email). Researchers should continue to explore the relationship between students' confidence calibrations (their level of confidence in their judgments) and students' actual performance on phishing assessments. Researchers should also determine if providing students with structured feedback, scenario-based instruction and hands-on training regarding verifying and reporting phishing scams results in improved phishing detection and protective behaviors over time.

5. Conclusions and Recommendations

Conclusions

This research project studied the third year undergraduate students at the University of Al-Qadisiyah, Iraq to determine their level of cybersecurity awareness, self-assessment of general confidence, use of protective cybersecurity practices and their ability to accurately identify simulated phishing emails. To provide an extensive assessment of the students' readiness for cybersecurity issues, this research project combined the results from student responses to a survey with simulated email classification exercises and confidence ratings based upon their decisions. Students reported a high level of self-assessed cybersecurity awareness and overall confidence; with average item scores of 3.78 and 3.77 respectively on a five point scale. However, the total percentage of correct email classification decisions was 73.0%. Therefore, a significant number of student decisions regarding the simulation emails were incorrect. Also, phishing detection sensitivity (67.8%) was lower than legitimate email identification specificity (78.2%). Therefore, students had more trouble identifying fraudulent communications than they did identifying authentic communications. In regards to protective practices used by students, there was less reported use of independent verification of communication and reporting of suspicious emails than there was reported avoidance of potentially harmful links or attachments. This indicates areas where there should be more focus placed on improving students' understanding of cybersecurity in college education programs.

The relationship between confidence in decision-making and actual accuracy in classifying observations was positive; however, mistakes in classification were made at all levels of confidence including the highest confidence level. Since the recorded confidence levels are not a single, clear-cut probabilistic scale, the results suggest there is a descriptive connection between confidence and accuracy as opposed to a definitive conclusion regarding the probabilistic nature of the confidence scale. The main value of the present study is the combination of perceived cybersecurity preparedness, performance on simulated phishing detection tasks, and confidence assessment of decisions taken within a sample of students/faculty from an Iraqi university. The results indicate that self-reported awareness and confidence are not equivalent to performance-based evaluations. Thus, the present study offers institutional-specific support for incorporating hands-on phishing recognition exercises, independent confirmation processes, and feedback with respect to confidence into undergraduate/graduate level cybersecurity curricula. However, the effectiveness of such practices will need to be tested through future experimental/controlled studies.

Recommendations

Based on the study findings, the following recommendations are proposed:

Including hands-on phishing detection training in the curriculum of cyber-security education.

Universities can provide students with hands-on learning experiences through assessments using both real-world phishing and legitimate emails in addition to providing theoretical understanding of phishing. These types of assessment will allow students to practice making decisions about which email is a threat, as well as providing faculty to obtain where students are having problems identifying threats in emails.

Prioritizing challenging phishing scenarios.

Educational activities must include similar situations to those that were less accurately detected during this research, such as requests for financial data, communication regarding rewards, and unannounced attachments. Additionally, training materials must focus on the indicators used to obtaine if a message is a reliable source of information instead of using only surface level warning indicators.

Strengthening independent verification and reporting procedures.

Universities should offer specific instructions on how to independently verify suspicious requests using trusted forms of communication, providing instructions on how to report potentially fraudulent emails. These recommendations are especially pertinent since there were very low self-report rates for both verifying and reporting suspicious e-mails.

Incorporating confidence reflection and explanatory feedback.

Confidence ratings are included with feedback from simulated phishing assessments to explain why a student's classification was either right or wrong. These confidence-rating-based simulated phishing assessments can help students understand their own uncertainty, and therefore may cause them to re-evaluate their confidence when they make an incorrect decision. However, we cannot assume that these simulated phishing assessments will provide education; instead we must evaluate whether or not they do so.

Establishing regular performance-based cybersecurity assessments.

The institution should perform an evaluation of the student's ability to recognize phishing on a regular basis in addition to assessing awareness levels and the protective practices being utilized. This evaluation will assist the institution in identifying what needs to be taught first, as well as tracking improvements in the student's classification abilities throughout their studies.

Improving the methodological design of future investigations.

In the end should have bigger sample sizes, keep track of individual observation points, and utilize clearly defined probability based confidence intervals. Controlled interventions or longitudinal designs are recommended when attempting to decidee if there is an improvement in detection performance and/or protective behaviors after receiving cybersecurity training.

6.3. Final Research Implication

The results of this study indicate that the degree to which a university can prepare for cyber security threats cannot be fully measured by only the level of student awareness of such issues. Student performance on distinguishing between phishing and legitimate emails as well as how confident students feel regarding their own judgment provide further insight into the level of student's practical cyber security skills. Including both types of assessments universities evidence based educational programs in cyber security.

REFERENCES

[1] G. Desolda, L. S. Ferro, A. Marrella, T. Catarci, and M. F. Costabile, "Human factors in phishing attacks: A systematic literature review," *ACM Computing Surveys*, vol. 54, no. 8, Art. no. 173, pp. 1–35, 2022. https://doi.org/10.1145/3469886

[2] S. Zhuo, R. Biddle, Y. S. Koh, D. Lottridge, and G. Russello, "SoK: Human-centered phishing susceptibility," *ACM Transactions on Privacy and Security*, vol. 26, no. 3, Art. no. 24, pp. 1–27, 2023. https://doi.org/10.1145/3575797

[3] L. Gallo, D. Gentile, S. Ruggiero, A. Botta, and G. Ventre, "The human factor in phishing: Collecting and analyzing user behavior when reading emails," *Computers & Security*, vol. 139, p. 103671, 2024. https://doi.org/10.1016/j.cose.2023.103671

[4] A. K. Gwenhure, "University students' security behavior against email phishing attacks: Insights from the health belief model," *Journal of Cybersecurity*, vol. 11, no. 1, p. tyaf034, 2025.

https://doi.org/10.1093/cybsec/tyaf034
[5] K. Singh, P. Aggarwal, P. Rajivan, and C. Gonzalez, "Cognitive elements of learning and discriminability in anti-phishing training," *Computers & Security*, vol. 127, p. 103105, 2023. https://doi.org/10.1016/j.cose.2023.103105
[6] M. Mossano and M. Volkamer, "Literature review: Misconceptions about phishing," in *Human Aspects of Information Security and Assurance*, N. Clarke and S. Furnell, Eds. Cham: Springer, 2025, pp. 215–228. https://doi.org/10.1007/978-3-031-72559-3_15
[7] M. Mutlutürk, M. Wynn, and B. Metin, "Phishing and the human factor: Insights from a bibliometric analysis," *Information*, vol. 15, no. 10, Art. no. 643, 2024. https://doi.org/10.3390/info15100643
[8] Z. Morić, V. Dakić, M. Plećaš, and I. Ogrizek Biškupić, "Evaluating end-user defensive approaches against phishing using education and simulated attacks in a Croatian university," *Journal of Cybersecurity and Privacy*, vol. 5, no. 3, Art. no. 38, 2025. https://doi.org/10.3390/jcp5030038
[9] B. Naqvi, K. Perova, A. Farooq, I. Makhdoom, S. Oyedeji, and J. Porras, "Mitigation strategies against the phishing attacks: A systematic literature review," *Computers & Security*, vol. 132, p. 103387, 2023. https://doi.org/10.1016/j.cose.2023.103387
[10] L. Ribeiro, I. S. Guedes, and C. S. Cardoso, "Which factors predict susceptibility to phishing? An empirical study," *Computers & Security*, vol. 136, p. 103558, 2024. https://doi.org/10.1016/j.cose.2023.103558
[11] R. Jabir, J. Le, and C. Nguyen, "Phishing attacks in the age of generative artificial intelligence: A systematic review of human factors," *AI*, vol. 6, no. 8, Art. no. 174, 2025. https://doi.org/10.3390/ai6080174
[12] N. N. S. Ismail, F. I. F. Rikzan, N. Katuk, N. L. Hashim, and N. A. Mohd Zulkefli, "Enhancing information security awareness on phishing among IT students: A pilot test case study at Politeknik Tuanku Syed Sirajuddin," *Journal of Digital System Development*, vol. 1, 2023. https://doi.org/10.32890/jdsd2023.1.2
[13] K. Althobaiti and N. Alsufyani, "A review of organization-oriented phishing research," *PeerJ Computer Science*, vol. 10, p. e2487, 2024. https://doi.org/10.7717/peerj-cs.2487
[14] Z. Shahbazi, R. Jalali, and M. Molaeevand, "AI-based phishing detection and student cybersecurity awareness in the digital age," *Big Data and Cognitive Computing*, vol. 9, no. 8, Art. no. 210, 2025. https://doi.org/10.3390/bdcc9080210
[15] M. Mossano and M. Volkamer, "Misconceptions about phishing," in *Human Aspects of Information Security and Assurance*, N. Clarke and S. Furnell, Eds. Cham: Springer, 2025, pp. 215–228.
[16] A. Alsaqer, H. Almajed, K. Alarfaj, and M. Frikha, "Phishing simulation as a proactive defense: A customizable platform for training and behavioral analysis," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 6, 2025. https://doi.org/10.14569/IJACSA.2025.01606104
[17] J. Prümmer, T. van Steen, and B. van den Berg, "A systematic review of current cybersecurity training methods," *Computers & Security*, vol. 136, p. 103585, 2024. https://doi.org/10.1016/j.cose.2023.103585
[18] J. Mungo, "Self-paced cybersecurity awareness training educating retail employees to identify phishing attacks," *Journal of Cyber Security Technology*, vol. 8, no. 2, pp. 71–119, 2024. https://doi.org/10.1080/23742917.2023.2244210
[19] C. Tian, M. L. Jensen, and A. Durcikova, "Phishing susceptibility across industries: The differential impact of influence techniques," *Computers & Security*, vol. 135, p. 103487, 2023. https://doi.org/10.1016/j.cose.2023.103487
[20] D. Gonzalez-Jimenez, F. Capozza, T. Dirkmaat, E. van de Veer, A. van Druten, and A. Baillon, "Falling and failing (to learn): Evidence from a nation-wide cybersecurity field experiment with SMEs," *Journal of Economic Behavior & Organization*, vol. 230, p. 106868, 2025. https://doi.org/10.1016/j.jebo.2024.106868
[21] J. Kuem, D.-H. Kwak, and S. Goel, "How to enhance antiphishing training performance? The roles of performance facilitators and message-specific coping responses," *Information & Management*, vol. 62, no. 8, Art. no. 104234, 2025. https://doi.org/10.1016/j.im.2025.104234

[22] T. Malloy, F. Fang, and C. Gonzalez, "Improving online anti-phishing training using cognitive large language models," *Computers in Human Behavior*, vol. 176, p. 108760, 2026. https://doi.org/10.1016/j.chb.2025.108760

[23] D. Hillman, Y. Harel, and E. Toch, "Evaluating organizational phishing awareness training on an enterprise scale," *Computers & Security*, vol. 132, p. 103364, 2023. https://doi.org/10.1016/j.cose.2023.103364

[24] D. Yin, M. T. Mullarkey, G.-J. de Vreede, and M. Limayem, "Timing of feedback after phishing simulations: Evidence from a randomized field experiment," in *Proc. 45th Int. Conf. Inf. Syst. (ICIS 2024)*, Paper 1435, 2024. https://aisel.aisnet.org/icis2024/security/security/2/

[25] M. Merritt, S. Hansche, B. Ellis, J. Nethery Snyder, K. Sanchez-Cherry, and D. Walden, "Building a cybersecurity and privacy learning program," NIST Special Publication 800-50, Rev. 1, National Institute of Standards and Technology, Gaithersburg, MD, USA, 2024. https://doi.org/10.6028/NIST.SP.800-50r1

[26] J. Prümmer, T. van Steen, and B. van den Berg, "Assessing the effect of cybersecurity training on end-users: A meta-analysis," *Computers & Security*, vol. 150, p. 104206, 2025. https://doi.org/10.1016/j.cose.2024.104206