Using Actions and Intentions to Evaluate Categorical Responses to Phishing and Genuine Emails

K. Parsons[1], A. McCormac[1], M. Pattinson[2], M. Butavicius[1] and C. Jerram[2]

[1] Defence Science and Technology Organisation, Edinburgh, Australia
[2] Business School, University of Adelaide, Australia
e-mail: {kathryn.parsons; agata.mccormac;
marcus.butavicius}@dsto.defence.gov.au; {malcolm.pattinson;
cate.jerram}@adelaide.edu.au

Abstract

While many studies have investigated people's susceptibility to phishing emails, little attention has been paid to how behavioural responses translate into overall intent when users are not informed they are undertaking a phishing study. This paper examines how well the quantitative multiple-choice categorisation used in such studies reflects the underlying reasoning of the users. The results of a role play scenario in which 117 participants were asked to manage 50 emails are presented. The users' multiple-choice actions were recoded based on their response to the question, *"What aspect of this email influenced your decision?"* using the Action-Intention Email Response Framework. According to this framework, intention incorporates the use of security-based reasoning, usefulness and phishing assessment. Results indicated that recoding did not significantly influence overall accuracy scores, which provides empirical support for the multiple-choice categorisation as a method of indirectly testing phishing susceptibility. However, closer examination revealed that combining the user's recommended actions with their qualitative responses provided significantly more detail on user's intent which, in many cases, changed the coding of the user's response to the email. Implications for the analysis of user performance in similar studies are discussed.

Keywords

Information security (InfoSec), Information risk, Phishing, Social engineering, Human behaviour

1. Introduction

The online threat posed by phishing has received widespread attention for almost a decade, but it remains a significant problem today (Furnell, 2013). Phishing describes a malicious attempt to deceptively acquire personal or financial information, and phishing attacks can have direct consequences to an individual or organisation, such as financial loss, as well as indirect consequences, such as damaged reputation (Parsons, McCormac, Pattinson, Butavicius, & Jerram, 2013). Recent research has demonstrated that when people are primed about phishing risks, they adopt a more diligent approach to screening emails (Pattinson, Jerram, Parsons, McCormac, & Butavicius, 2012). This has implications for the interpretation of previous studies of users' susceptibility to phishing, as it is likely that studies in which the concept of phishing was mentioned to participants may have underestimated users' susceptibility.

Hence, to more accurately assess the level of susceptibility expected in the real world (where people are infrequently reminded about the risks of phishing), it is necessary to utilise a method where participants are not directly told to make a decision regarding the legitimacy of an email. However, such studies make assumptions about a users' underlying thought process, and it may be difficult to reflect the complexity of user behaviour without revealing that participants are undertaking a phishing study.

1.1. Previous research

A number of studies have assessed users' susceptibility by directly asking participants to make a phishing decision. For example, Furnell (2007) provided respondents with the options 'illegitimate', 'legitimate' or 'don't know', and other studies asked participants to rate an email's authenticity on a five point scale, from 'Certainly phishing' to 'Certainly not phishing' (Jakobsson, Tsow, Shah, Blevis, & Lim, 2007; Tsow & Jakobsson, 2007) or asked participants to indicate if an email was a phishing attempt (Robila & Ragucci, 2006). A more recent study provided participants with the options to mark emails as important, leave them in the inbox or delete, but this study informed participants that they were assessing the legitimacy of emails (Hong, Kelley, Tembe, Murphy-Hill, & Mayhorn, 2013).

Several researchers have used a role play scenario, in which participants are not directly informed that they are taking part in a phishing study. Instead, participants are told they are participating in a study about email use or computer use. For example, Wang, Chen, Herath, Vishwanath and Rao (2012) asked participants about their likelihood to respond to an email on a five-point scale from 'Not At All Likely' to 'Very Likely'. Pattinson et al. (2012) provided participants with the options 'Leave the email in the inbox and flag for follow up', 'Leave the email in the inbox', 'Delete the email' or 'Delete the email and block the sender'.

Downs, Holbrook and Cranor (2006) presented users with eight emails, and asked them to react as they normally would in their own life. The options chosen naturally (e.g., reply by email, contact the sender by phone or in person, delete the email, save the email, click on the link, copy and paste the URL, or type the URL into a browser window) were then provided to participants in future experiments (Downs, Holbrook, & Cranor, 2007; Sheng, Holbrook, Kumaraguru, Cranor, & Downs, 2010). Although these options are comprehensive and likely to allow for the complexity of user behaviour, the specificity may indirectly prime participants about the fact the study is interested in phishing susceptibility.

1.2. Aim of this paper

The aim of this paper is to validate the quantitative multiple-choice categorisation utilised by Parsons, McCormac, Pattinson, Butavicius & Jerram (2013). To achieve this, we developed a framework, the Action-Intention Email Response Framework, through which user action and intention could be recoded. Results based on the raw (quantitative, multiple-choice) method are then compared with the recoded mixed method (quantitative and qualitative) results (Tashakkori & Teddlie, 2003) to validate and evaluate the methodology of Parsons et al. (2013).

The structure of this paper is as follows. The next section describes the research method and details of the emails utilised. This is followed by the results, in which the categories of the framework are introduced and described, and the recoded results are presented and compared to the raw scores. Finally, the conclusions summarise the significance of these findings in the context of previous literature.

2. Method

Fifty emails, consisting of 25 genuine emails and 25 phishing emails, were utilised. All emails were either actual emails received by the authors, or were found online. The emails represented a range of topics such as banking, shopping and social networking, and an effort was made to select comparable genuine and phishing emails for each topic. The emails were altered to include the details of a fictitious character, ‘Sally Jones’, as if she was the recipient of the emails. A role play based method was utilised, in which participants were told that they were viewing emails from the inbox of ‘Sally Jones’ and that the experiment was designed to study how people manage emails.

The participants consisted of 117 students from the University of Adelaide. The majority were female (90) and in the first year of their university study (93). The sample included 64 business students and 54 psychology students. They received $25 cash for their participation. Participants were presented with randomised images of 50 email messages sequentially. For each email, participants were asked to respond to the question, *“How would you manage this email?”* with one of the four replies: A) leave the email in the inbox and flag for follow up; B) leave the email in the inbox; C) delete the email; or D) delete the email and block the sender. Participants were also required to respond to the question *“What aspect of this email influenced your decision?”* in an open text field. More information regarding this scenario-based method (Erickson, 1995) is reported in Parsons et al. (2013).

3. Results

In Parsons et al. (2013) a phishing email was deemed to be correctly managed if participants responded with ‘delete the email’ or ‘delete the email and block the sender’. A genuine email was deemed to be correctly managed if participants responded with ‘leave the email in the inbox and flag for follow up’ or ‘leave the email in the inbox’.

However, a preliminary analysis of qualitative responses to the question, *“What aspect of this email influenced your decision?”* revealed that some responses did not correspond with the assumption that a phishing email would be *deleted* or *deleted* *and blocked*, and a genuine email would be *left in the inbox* or *flagged for follow up*. Some participants flagged emails for follow up to alert Sally to a potential security risk that should be investigated.

In order to examine the decision-making process more closely, a second-order analysis was conducted (Shkedi, 2004). This involved deriving the participants' underlying reasoning from their qualitative response which, in turn, was used to modify the first order responses (i.e., the categorical responses). Of the 5850 responses (50 responses for each of the 117 participants), 5817 (or 99%) included enough information to enable recoding. The other 33 responses were either insufficient or unclear, and those responses were removed from further analysis.

3.1. Reasoning provided by participants

An analysis of the responses provided by participants, together with the multiple-choice option chosen, identified four important aspects, namely:

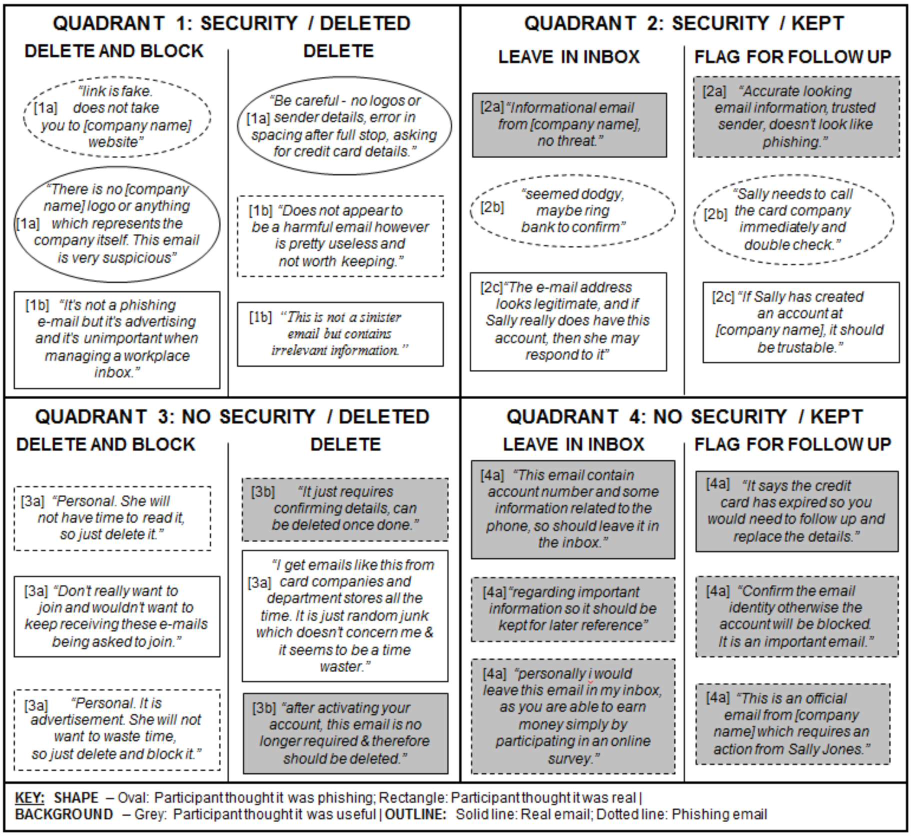
- the action recommended by participants (e.g., to delete or keep the email),
- whether participants mentioned or implied the use of security-based reasoning,
- whether participants appeared to believe the email was phishing, and
- whether participants appeared to believe the email was useful.

These aspects were used to develop the Action-Intention Email Response Framework, which represents the possible reasons for participants' decisions. The framework consists of eight categories, as shown in Table 1. It is important to highlight that categorisation for the security factor was based on whether the participant mentioned or implied the use of security-based reasoning, even if the aspects of security specified were incorrect or incorrectly implemented. For example, one participant chose to *delete and block* a genuine email from a bank, and responded with, *"There is no [bank name] logo or anything which represents the company itself. This email is very suspicious"*. This is classified as security-based reasoning, even though the assumption that a genuine email should have a professional looking logo is flawed, and the reasoning utilised did not help the individual to correctly manage the email in question.

Category Identification	**Action**	**Security-based reasoning?**	**Perceived to be phishing?**	**Perceived to be useful?**
1a	Deleted	Yes	Yes	No
1b	Deleted	Yes	No	No
2a	Kept	Yes	No	Yes
2b	Kept	Yes	Yes	No
2c	Kept	Yes	Conditional	Conditional
3a	Deleted	No	No	No
3b	Deleted	No	No	Yes
4a	Kept	No	No	Yes

Table 1: Categories of the Action-Intention Email Response Framework

Every response to the question, *"What aspect of this email influenced your decision?"* was then examined by one of three judges and was assigned a category. To ensure valid and consistent categorisation, Taylor's (1976) hermeneutics approach was utilised. Random subsets of the responses were re-examined by a second judge and contentious cases were then examined by all three judges to reach a consensus.



QUADRANT 1: SECURITY / DELETED

DELETE AND BLOCK

[1a] "link is fake. does not take you to [company name] website"

[1a] "There is no [company name] logo or anything which represents the company itself. This email is very suspicious"

[1b] "It's not a phishing e-mail but it's advertising and it's unimportant when managing a workplace inbox."

DELETE

[1a] "Be careful - no logos or sender details, error in spacing after full stop, asking for credit card details."

[1b] "Does not appear to be a harmful email however is pretty useless and not worth keeping."

[1b] "This is not a sinister email but contains irrelevant information."

QUADRANT 2: SECURITY / KEPT

LEAVE IN INBOX

[2a] "Informational email from [company name], no threat."

[2b] "seemed dodgy, maybe ring bank to confirm"

[2c] "The e-mail address looks legitimate, and if Sally really does have this account, then she may respond to it"

FLAG FOR FOLLOW UP

[2a] "Accurate looking email information, trusted sender, doesn't look like phishing."

[2b] "Sally needs to call the card company immediately and double check."

[2c] "If Sally has created an account at [company name], it should be trustable."

QUADRANT 3: NO SECURITY / DELETED

DELETE AND BLOCK

[3a] "Personal. She will not have time to read it, so just delete it."

[3a] "Don't really want to join and wouldn't want to keep receiving these e-mails being asked to join."

[3a] "Personal. It is advertisement. She will not want to waste time, so just delete and block it."

DELETE

[3b] "It just requires confirming details, can be deleted once done."

[3a] "I get emails like this from card companies and department stores all the time. It is just random junk which doesn't concern me & it seems to be a time waster."

[3b] "after activating your account, this email is no longer required & therefore should be deleted."

QUADRANT 4: NO SECURITY / KEPT

LEAVE IN INBOX

[4a] "This email contain account number and some information related to the phone, so should leave it in the inbox."

[4a] "regarding important information so it should be kept for later reference"

[4a] "personally i would leave this email in my inbox, as you are able to earn money simply by participating in an online survey."

FLAG FOR FOLLOW UP

[4a] "It says the credit card has expired so you would need to follow up and replace the details."

[4a] "Confirm the email identity otherwise the account will be blocked. It is an important email."

[4a] "This is an official email from [company name] which requires an action from Sally Jones."

KEY: SHAPE – Oval: Participant thought it was phishing; Rectangle: Participant thought it was real | BACKGROUND – Grey: Participant thought it was useful | OUTLINE: Solid line: Real email; Dotted line: Phishing email

Figure 1: Illustration of the Action-Intention Email Response Framework, showing examples of the different responses within the four quadrants

Figure 1 provides an illustration of the Action-Intention Email Response Framework, based on two factors: the action recommended by participants (i.e., whether to keep or delete the email) and the degree to which participants used security-based reasoning. These two factors (recommended action and use of security-based reasoning) were used to plot the different categories within the two-factor space, resulting in four quadrants.

The upper left quadrant (Quadrant 1) of the factor-space is associated with responses where participants used security-based reasoning and chose to *delete* or *delete and block* the email. Categories 1a and 1b fit within this quadrant. The upper right quadrant (Quadrant 2) is associated with cases where participants used security-based reasoning and chose to keep the email in the inbox or flag the email for follow up. Categories 2a, 2b and 2c fit within this quadrant. The lower left quadrant (Quadrant 3) is associated with responses where participants did not use security-based reasoning and chose to *delete* or *delete and block* the email. Categories 3a and 3b fit within this quadrant. Finally, the lower right quadrant (Quadrant 4) is associated with responses where participants did not use security-based reasoning and kept the email in the inbox or flagged the email for follow up. Category 4a fits within this quadrant.

The other aspects of user intention (i.e., usefulness and phishing assessment) are also depicted in Figure 1. The shape surrounding the responses specifies whether or not participants appeared to believe that the email was phishing; the oval corresponds with responses where participants thought the email was phishing, and the rectangle corresponds with responses where participants thought the email was genuine. The shading in the background of the responses indicates participants' opinions regarding the usefulness of the emails, where a grey background means that participants thought the email was useful. The line surrounding the shapes corresponds with the actual nature of the email. For phishing emails, the shape is surrounded by a dashed line, and for genuine emails, the shape is surrounded by a solid line. Hence, a grey rectangle with a dashed line corresponds with a phishing email that the participant thought was both useful and genuine. The number in the corner of each response indicates the number of the associated category (see Table 1 for categories).

3.2. Recoding of phishing emails

When participants were faced with a phishing email, the best responses were ones where the participants deleted the email message and explicitly provided security concerns as the reason for their decision. For example, one participant gave the following response to a phishing email: *"link is fake. does not take you to [bank name] website"*. Responses of this nature were classified as category 1a and this reasoning was used for phishing emails in 26% of cases, as shown in Table 2.

There were a minority of cases where participants kept an email, but did so because they wanted to alert Sally to a possible security threat that she should follow up with the purported company or organisation. When faced with phishing emails, responses of this nature are also extremely positive, as they suggest that the participant has a useful awareness of security. For example, one participant responded with: *"Address looks dodgy, don't' click link – call [organisation name] for confirmation"*. Responses of this nature were classified as category 2b, and this reasoning was used for phishing emails in 6% of cases.

The worst possible responses to phishing emails were ones where participants kept an email because they believed it was valid. For example, a participant responded to a phishing email with: *"I am confident in this email and will follow it up, as I know it is not a scam"*. This is particularly concerning, as it indicates that the participant is conscious of security, but has an inaccurate knowledge of what constitutes a security concern. Responses of this nature were classified as category 2a, and this category of response was provided for phishing emails in 10% of cases.

Also of concern were category 4a responses, where participants appeared to simply take emails at face value. For example, one participant gave the following response to justify a decision to keep an email: *"This is a requirement, as [company] users* *have to prove their id. If individuals have to use [company], they have to follow this email's instruction"*. This therefore suggests that the participant did not consider security when deciding how to manage that email. This was the most common response for phishing emails, and was chosen in 32% of cases.

3.3. Recoding of genuine emails

When participants were faced with genuine emails, the best response was when participants kept the email because they believed it was valid. For example, one participant gave the following (successful) response to a genuine email: *"Clearly legit, tells Sally to enter the URL herself and warns her not to click on links"*. Responses of this nature were classified as category 2a, and were given for genuine emails in 19% of cases. Category 2c responses were similar and were given in 6% of cases. These emails were kept on the basis of security reasoning, but the participants required further information before deciding whether to trust the email. An example to this is, *"The e-mail address looks legitimate, and if Sally really does have this account, then she may respond to it"*.

There were cases where participants decided to delete a genuine email, but made it clear it was not because the email was phishing, but rather, because they did not believe it was useful for Sally. For example, one participant responded to a genuine email with: *"No motivation to participate in the lengthy survey. Although it is a valid sender email address"*. This therefore indicates a good security aware decision. These responses were given for only 2% of genuine emails and were classified as category 1b.

In contrast, the worst possible response was when participants deleted a genuine email because they believed that it was illegitimate. The following provides an example of this type of response: *"The layout of the email looks illegitimate - there is no logo and the email is very short"*. Responses of this nature were classified as category 1a, and genuine emails were managed using this reasoning in 10% of cases.

3.4. Analysis of Raw and Recoded Decisions

The qualitative analysis and framework provided above captures the complexity of the decision-making process that is often overlooked in most phishing research, and the reason second-order analysis (Shkedi, 2004) was used in this study. To understand the significance of this data, the raw results based on the multiple-choice categorisation (first-order analysis) were compared to the recoded results (second-order analysis), that take into account both user action and intention (and quantitative and qualitative data). The percentage of responses in each of the raw and recoded categories can be seen in Table 2.

	Category	Phishing	Genuine
Raw	Delete and Block	15%✓	7%
	Delete	33%✓	33%
	Leave in inbox	25%	36%✓
	Flag for follow up	27%	25%✓
	Accuracy	*48%*	*60%*
Recoded	1a	26%✓	10%
	1b	2%	2%✓
	2a	10%	19%✓
	2b	6%✓	3%
	2c	4%	6%✓
	3a	16%✓	25%
	3b	4%	3%✓
	4a	32%	32%✓
	Accuracy	*48%*	*62%*

✓ Denotes a category that was deemed to be correct

Table 2: Percentage of responses in raw and recoded data

To determine whether the recoding significantly influenced the accuracy scores for the 50 emails, paired samples t-tests were conducted. For phishing emails, there was no significant difference in mean accuracy between the raw (M = 48.03, SD = 12.00) and recoded scores (M = 48.37, SD = 10.41, t(24) = -.27, p = .789, Cohen's d = -0.03). There was also no significant difference in mean accuracy between the raw (M = 60.17, SD = 11.48) and recoded scores (M = 61.86, SD = 10.82, t(24) = -1.95, p = .063, Cohen's d = -0.15) for genuine emails. This means that the quantitative multiple-choice categorisation provides an accurate reflection of the underlying reasoning or intent of users based on overall accuracy scores.

However, this must be interpreted with caution. Results were examined on an individual level to determine how much the recoding changed the accuracy scores of participants. For 23 participants (20%) the recoding changed accuracy scores by less than 1%. Accuracy scores were changed by between 1-5% for a further 59% (50%) of participants. This means that, in the majority of cases, the raw results provided by the multiple-choice categorisation was appropriate. But for almost a third (30%) of participants, the recoding revealed that their underlying intentions were not captured by the raw results. Instead, accuracy scores changed by between 6-10% for 21 participants, between 11-15% for 11 participants, and by over 20% for 3 participants.

Furthermore, the recoding facilitates closer and more useful examination of results. For example, although the overall accuracy score for phishing emails is 48%, only 32% were accuracy decisions made with security reasoning. Less than half (27%) of the correct results for genuine emails were made for security-based reasons. This information regarding incorrect security-based reasoning is also very informative. Recoded results indicate that 16% of decisions made for phishing emails and 13% of decisions made for genuine emails incorrectly used security reasoning. This has important implications for education and training, which will be highlighted in Section 4.

4. Conclusions

This study provides support for the categories ‘leave the email in the inbox and flag for follow up’, ‘leave the email in the inbox’, ‘delete the email’ and ‘delete the email and block the sender’ as a method of indirectly measuring phishing susceptibility. A role-play scenario was presented to 117 participants, and their overall accuracy scores, based on the multiple-choice options above, did not differ significantly when the results were recoded based on the Action-Intention Email Response Framework. This provides some validation of the multiple-choice categories utilised in Parsons et al. (2013).

However, since just one successful phishing attack can cause extensive damage to an individual or their organisation, examining only overall accuracy scores is not sufficient. A closer inspection of results indicated that the recoding changed accuracy scores by 11% or more for 14 of the 117 participants (12%). This means that the raw scores failed to accurately reflect the intentions of a minority of users.

The Action-Intention Email Response Framework also provides information regarding the type of mistake made by participants, which can be used for training and education purposes. An examination of the incorrect results for genuine emails revealed that 13% inaccurately used security-based reasoning, whereas 36% made no reference to security. Approximately 16% of incorrect decisions regarding phishing emails inaccurately used security-based reasoning, whereas 25% made no reference to security. It is likely that what constitutes effective training would differ based on the type of mistake made. For users who did not consider security, a simple awareness seminar might be sufficient, whereas users who considered security but inaccurately implemented the knowledge might need a more in-depth explanation of the security rules.

It is important to note that the participants in our sample consisted of business and psychology students, most of whom were in the first year of their studies. It is possible that the findings may differ in a sample of participants from a wider range of disciplines or employees rather than students. Further research into what individual differences, such as personality, experience or decision-making style, may influence the consideration of security information is warranted.

Hence, although user actions themselves are a rich source of data for analysing the results of psychological studies and inferring user intent from user action, user actions looked at in isolation may not always indicate users’ underlying thought processes. A simple multiple choice response does not allow for the complexity of human reactions to phishing emails. This study highlights that researchers should not make assumptions about decision-making processes, and should instead delve deeper into the reasoning behind users’ actions in phishing experiments. The methodology of this paper, in which action and intention are combined, could be applied in future studies to validate and evaluate user performance.

5. References

Downs, J.S., M. Holbrook & Cranor, L.F. (2007, October). Behavioral response to phishing risk (pp. 37-44). Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit, ACM.

Downs, J.S., M.B. Holbrook & Cranor, L.F. (2006, July). Decision strategies and susceptibility to phishing. Proceedings of the second symposium on Usable privacy and security (pp. 79-90), ACM.

Erickson, T. (1995). Notes on design practice: Stories and prototypes as catalysts for communication, Scenario-based design: envisioning work and technology in system development. New York: John Wiley & Sons, Inc.

Furnell, S. (2013). Still on the hook: the persistent problem of phishing. Computer Fraud & Security, 2013(10), 7-12.

Furnell, S. (2007). Phishing: can we spot the signs? Computer Fraud & Security, 2007(3), 10-15.

Hong, K.W., Kelley, C.M., Tembe, R., Murphy-Hill, E., & Mayhorn, C.B. (2013, September). Keeping Up With The Joneses Assessing Phishing Susceptibility in an Email Task. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 57, No. 1, pp. 1012-1016). SAGE Publications.

Jakobsson, M., Tsow, A., Shah, A., Blevis, E., & Lim, Y.K. (2007). What instills trust? a qualitative study of phishing. In Financial Cryptography and Data Security (pp. 356-361). Springer Berlin Heidelberg.

Parsons, K., McCormac, A., Pattinson, M., Butavicius, M. & Jerram, C. (2013). Phishing for the truth: A scenario-based experiment of users’ behavioural response to emails. In L.J. Janczewski, H. Wolf, and S. Shenoi (Eds.): Security and Privacy Protection in Information Processing Systems - IFIP Advances in Information and Communication Technology (Vol. 405, pp. 366-378). Springer Berlin Heidelberg.

Pattinson, M., Jerram, C., Parsons, K.M., McCormac, A., Butavicius, M.A. (2012). Why do some people manage phishing emails better than others? Information Management & Computer Security, 20(1), 18-28.

Robila, S.A., & Ragucci, J.W. (2006, June). Don't be a phish: steps in user education. In ACM SIGCSE Bulletin (Vol. 38, No. 3, pp. 237-241). ACM.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. (2010, April). Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 373-382). ACM.

Shkedi, A. (2004). Second-order theoretical analysis: A method for constructing theoretical explanation. International Journal of Qualitative Studies in Education, 17(5), 627-646.

Tashakkori, A., & Teddlie, C. (Eds.). (2003). Handbook of mixed methods in social & behavioral research. Sage.

Taylor, C. (1976). Hermeneutics and politics. Critical sociology, selected readings, 153-193.

Tsow, A., & Jakobsson, M. (2007). Deceit and Deception: A Large User Study of Phishing. Technical report TR649, Indiana University.

Wang, J., Chen, R., Herath, T., Vishwanath, A., & Rao, H. R. (2012). Phishing Susceptibility: An Investigation into the Processing of a Targeted Spear Phishing Email. IEEE Transactions on Professional Communication, 55(4), 345-362.