Identifying the Effect of Data Breach Publicity on Information Security Awareness Using Hierarchical Regression

HUI NA CHUA\(^1\), (Senior Member, IEEE), JIA SHENG TEH\(^2\), AND ANTHONY HERBLAND\(^3\)

\(^1\)Department of Computing and Information Systems, Sunway University, Selangor 46150, Malaysia
\(^2\)Department of Risk Advisory, Deloitte Business Advisory Sdn Bhd, Kuala Lumpur 60000, Malaysia
\(^3\)School of Health and Social Work, University of Hertfordshire, Hatfield, Hertfordshire AL10 9AB, UK

Corresponding author: Hui Na Chua (huinac@sunway.edu.my)

This research was supported by the Ministry of Education Malaysia Fundamental Research Grant Scheme [FRGS/1/2019/ICT04/SYUC/02/2].

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Sunway University Research Ethics Committee under Application No. SUREC 2020/074.

ABSTRACT The technological evolution has formed new challenges for organizations to safeguard their information as digital assets. Information Security Awareness (ISA) is the cognitive state where individuals comprehend information security, threats, and the capability to develop preventive strategies. Prior studies discovered that human mistakes or misbehavior is the most vulnerable link in information security due to insufficient security awareness. There were massive data breaches reported throughout the years globally. Literature shows that individuals will develop their evaluations of risks and sense of security awareness when receiving security risk information such as data breach incidents. These indications motivated us to examine the effect of an unexplored factor, that is, data breach publicity (DBR) on ISA. The purpose of this research is to discover if DBR significantly improves a model’s ability to predict ISA and its magnitude in influencing ISA. A 3-stage hierarchical linear regression approach was used to build up the model with prior known influential factors to predict ISA. To the extent of our knowledge, there is no study reported to date regarding the implication of DBR on ISA. Our main findings reveal that DBR significantly explains 6.7% of ISA and achieves the highest coefficient comparing with prior known factors. Our research contributes to a novel discovery of a new factor that significantly influences ISA and its magnitude in increasing ISA. This discovery implies the need to incorporate the knowledge of data breach incidents into ISA-related educative programs or strategies to increase ISA.

INDEX TERMS Data breach publicity, hierarchical regression, information security awareness, information privacy.

I. INTRODUCTION

As organizations in this era heavily rely on Information Systems (IS) to function and ensure high productivity, the concern for information security has emerged as one of the top priorities in any organization’s security management [1], [2] to maintain information availability [3], confidentiality and integrity [4]. As a result, information security issues have become a concern for most businesses, as they would result in severe consequences, for instance, corporate liability, financial loss, reputation, and credibility damage [5]. To minimize the security risks, organizations have deployed technical measures such as implementing security technologies to safeguard business information assets [6]. Nonetheless, it was found that investing in such technologies solely is inadequate to eliminate security risks and is not fully adequate to guarantee information security [7].

Researchers asserted that the weakest link in the information security system is usually due to the employees’ misused behavior; more specifically, employees’ naïve mistakes and accidental or intentional harm are the most prevalent factors leading to security breaches [4], [8].

Information Security Awareness (ISA) is important as it would significantly impact employees’ security behavior and
their adherence to the organization’s security policies and regulations [9]. ISA can be denoted as an individual’s cognition whereby he or she processes information regarding information security, which can be described by understanding the essentialness of information security and developing consciousness and awareness on security objectives, issues, vulnerabilities; and possessing interest towards acquiring the necessary skill sets and knowledge to utilize information systems responsibly [10].

Findings from several pieces of research demonstrated that the primary cause of information system misuse behavior and its repercussions among employees is insufficient security awareness regarding Information Security Policies (ISP) and security best practices [11]. Although previous literature has extensively presented the significance of employees’ ISA, several investigations showed that ISA is still a crucial subject to comprehend. Most employees in the organizations do not possess adequate ISA on security incidents, regulations, and policies [12].

In Malaysia, the Personal Data Protection Act 2010 (PDPA) is enforced to safeguard individuals’ data, but it does not have obligations for notification if a data breach occurs in an organization. In prior studies, organizations’ practices and policy implementation for ensuring personal data security were found inadequate [13]. A massive data breach from Malaysian local Telco companies was first discovered that there was a leak of information on phone numbers and personal information for more than 46 million users. Nonetheless, the leak was estimated to have happened between 2012 and 2013, and there were at least 178 cases of data breaches reported until 2019 [14]. Internationally, some of the high-profile data breaches that happen throughout the years involve large corporations. [15], [16].

Work in [17] showed that individuals would develop and form their subjective perceptions and evaluations of risks when they receive news about these risks, making these risks more prominent in an individual’s mind. A prior study in [18] further supported that a security breach can immediately develop an individual’s perception of vulnerability and susceptibility to risks and harms. Literature [19] also revealed that the frequency of data breach news dissemination affects data protection awareness.

Previous studies have focused on different perspectives such as architectures of security awareness programs, initiatives, and strategies [20] for predicting ISA. The known factors studied by prior works and found significantly affecting ISA [11], [21] include information security policy, ISA knowledge, training, negative experience, media influence, peer behavior, and culture. Further, prior studies indicated that when an individual recognizes a threat through reading security-related publications [45], it will develop a sense of awareness [37] and negative experience [46]. This reasoning will indirectly enable them to pay more attention to information security.

Nonetheless, to the extent of our knowledge, no scientific study reported thus far has examined the implication of data breach publicity on ISA. As such, our research seeks to answer the following:

RQ1) Does DBR have a significant influence in improving a model’s ability to predict ISA?
RQ2) How does DBR affect ISA?

Answering these research questions provides novel evidence to discover if a new factor (DBR) can significantly improve a model’s ability to predict ISA and its magnitude in affecting ISA. This evidence will lead to the implication if there is a need to incorporate data breach incidents into an ISA educative program or strategy for improving ISA.

The rationale for choosing hierarchy regression modelling as the primary approach is based on its capability in enabling more than one variable (i.e., factor) to be added into a predictive model in separate stages called “blocks” to statistically “control” for certain variables, to investigate whether adding a predictor variable(s) (i.e., the DBR) significantly improves the model’s ability to predict the target variable (i.e., ISA). Further, hierarchy regression also allows the analysis of the DBR effect on ISA through observing the regression coefficients (β). The regression coefficient represents the mean change in the target variable (i.e., ISA) for one unit of change in the predictor variable (i.e., DBR) while holding other variables in the model constant. These capabilities suit the context of our research questions setting in RQ1 and RQ2, respectively. Our approach of using hierarchy regression contributes to a methodological demonstration of how this approach can be applied in a similar research scenario.
The second element in describing ISA pertains to the degree to which employees’ motivation, commitment, and behavior reflect on the organization’s core information security mission, in line with meeting the prerequisites for information security best practices [4], [23].

Bulgurcu et al. [23] also categorized ISA as two different components, which are General Information Security Awareness (GISA) and Information Security Policy Awareness (ISPA), respectively. GISA is described as an individual’s general comprehension and knowledge of information security issues and threats, whereas ISPA is an employee’s knowledge and comprehension to meet and achieve the organization’s information security policy prerequisites. Thereupon, exploring and understanding the concept of ISA, factors influencing ISA are imperative to develop practical security knowledge and consciousness to mitigate security risks.

Researchers in the information technology domain have presented different approaches to investigate ISA [13], [26]. For example, studies on exploring information security awareness on social media were conducted. Apart from that, Stanton et al. [27] examined ISA by performing a survey to study security behaviors on password settings. Moreover, studies investigating smartphone users’ ISA were also conducted [28].

ISA is not a behavior factor, but it usually leads to developing and cultivating security behavior [9], [30]. On the contrary, security behavior is built up based on behavioral theories such as General Deterrence Theory, Protection Motivation Theory, and Theory of Planned Behavior that affect an individual’s behavior to safeguard information security, for instance, an individual’s intention to conform with information security policies [6].

B. RELATED STUDIES OF INFLUENTIAL FACTORS (KNOWN FACTORS) ON ISA

Literature has shown that ISA positively affects security policy compliance behavior [12], [31]. Therefore, it is crucial to identify influential factors that significantly impact ISA. Prior studies discovered numerous influential factors. They are: i) information security policies, ii) Security, Education, Training and Awareness (SETA) programs, iii) information system knowledge, iv) negative experience of security incidents, v) media influence, vi) peer behavior, v) organizational culture and information security culture. Each of these factors is discussed as follows.

Establishing Information Security Policies (ISP) is essential in an organization’s information security management [32]. ISP is referred to as a “direction-giving document” [33] for employees to follow and adhere to the organization’s regulations [33]. Several scholars described ISPs as a plethora of roles, responsibilities, regulations, and guidelines that are stipulated to safeguard an organization’s crucial digital assets, resources, and information technology [34], [35]. Previous studies have proven that ISP provisions that are accessible and understandable positively impact an individual’s ISA and knowledge [29].

SETA is a program that aims to educate and train employees to develop their information security principles awareness and potential risks to an organization’s valuable assets [36]. SETA programs encompass a wide range of different approaches, methods, and measures that involve implementing security education, training, and awareness-raising activities [36], [37]. Prior researches have contended that the reinforcement and implementation of SETA programs have helped to reduce and discourage user’s misuse and risky security behavior, subsequently lowering the information security-related errors caused by employees [22], [34], [38]. Findings presented in prior studies revealed that security training carried out in organizations strongly influences employees’ ISA to the cognitive and behavioral extent [32], [39] that help to increase employee’s ISA substantially through the inculcation of security knowledge. Several pieces of research have also suggested different methods, such as employee involvement [40], discussion, and phishing training [41] that helped raise ISA. Security education helps to form employees’ interest and consciousness about security issues, thereby raising awareness and helping them to develop preventive measures on the organization’s vulnerability to security threats [42].

The term “IS Knowledge” is often referred to as the comprehensive understanding of an information system encompassing a broad range of components, including computer knowledge, self-efficacy, and innovativeness with computing technology [43]. Kruger et al. [3] asserted that an individual’s ability to protect and enhance integrity, confidentiality and information availability is based on understanding the constructs of information security, which indeed IS knowledge is necessary. For example, to prevent security incidents from happening, it is vital to understand the consequences that virus infection may result from an individual’s careless security behavior when using the Internet, using strong passwords to protect personal information, and practicing regular backups to recover from a loss caused by security threats [44]. There are prior studies [30], [37] that empirically have proven that an individual’s degree of IS knowledge has a positive association with ISA.

Individuals might have encountered different security incidents in their lives due to misbehavior and carelessness when handling information. Security incidents may happen in numerous forms. For example, individuals themselves might experience passwords being stolen, accounts being hacked, credit card fraud, or even fall for a phishing mail when using the Internet. Additionally, they might also gain security experiences from their peers, family members and through reading security-related news articles [45]. These personal experiences might be a valuable lesson for developing a sense of awareness [37]. It can be deduced that when individuals have a better overall understanding of security incidents, they will be more careful and warier in terms of cognition and behavior in ensuring information security [45]. Furthermore,
individuals must recognize that it will lead to negative implications [46]. Consequently, this cognitive reasoning will indirectly enable them to pay more attention to it.

Spreading security awareness involves more than providing security education and establishing security policies. Hence, the critical factor is that it must be accessible and reachable to its audience. Happer and Philo [47] proclaimed that the mass media has supported developing and shaping public understanding, which also leads to the essential change in behavior. The media landscape has tremendously evolved and transformed over the years [48]. Social media, as one of the communication mediums, is progressively superseding traditional media [49]. In the business context, a prior study showed that social media marketing positively impacts raising awareness and buying behavior [50]. However, there is scant research attention devoted to exploring social media’s impact on awareness in the information security domain despite other application fields that presented concrete evidence that social media impact an individual’s awareness.

Behavioral literature has demonstrated the importance of subjective norms, which indicates that the perceptions of individuals considered significant to an employee influence how the employee behaves [51]. It is found that employees tend to adhere to an organization’s security policies if they notice that people around them, particularly managers, supervisors, colleagues, and subordinates, are also fulfilling their responsibilities in adhering to security provisions and regulations [32]. The literature claimed that if those deemed an important person to the employee in advocating and supporting the employee to adhere to the organization’s security regulations, the employee will follow their advice [32]. Prior studies also denoted that the expectations of an employee’s peers indirectly form a persuasive pressure and influence on his or her positive security adherence attitude, and this attitude will also derive a significant impact on other employees [51], [52]. Prior studies also showed that employees develop security awareness through conversation and discussion with their peers [39], [40].

Information security culture relates to the implemented measures and employees’ behavior to protect the organization’s digital assets [53]. Research has also revealed that solid leadership support has an extensive impact on information security management and the cultivation of a better security culture [54]. Literature also claimed that security culture is strongly associated with the organization’s mission and vision [55]. Several pieces of research have made significant contributions to the association between organizational culture and information security culture [2], [56], [57]. Connolly et al. [58] had proven that organizations that focus more on employee interaction and culture tend to have a positive direction towards ISA. Several studies have also implicated that information security culture notably influences an individual’s ISA [56], [59]. Parsons et al. [60] found that employees who are more knowledgeable and aware of information security issues tend to conform with established policies when an organization has a robust and vigilant security culture. Wiley et al.’s [61] findings further confirmed that security culture plays a role as a mediator between the association of ISA and organization culture.

C. DATA BREACH PUBLICITY

A data breach is a security incident, which involves unauthorized access, disclosure, usage, or disposal of data, often personal data [62]. A data breach may lead to illegal access or exposure of confidential, sensitive, or any protected information that encompasses personally identifiable information, health-related information, trade secrets, or intellectual property [63]. Data breaches can happen in different ranges of scopes, from impacting a few persons to a large group of customers [63].

Some of the high-profile data breaches that happen throughout the years involve large corporations such as Facebook, Yahoo, Equifax, and Marriott. For example, in 2018, Facebook reported that around 87 million users’ data had been inappropriately harvested by Cambridge Analytica [15]. Apart from that, in 2018, Marriott International, an American multinational hospitality organization that operates a large group of hotels, revealed that records from 500 million guests were breached and stolen by an unauthorized party on its hotel reservation systems [16].

Martin et al. [18] argued that a security breach is capable of immediately developing an individual’s perception of vulnerability and susceptibility to risks and harms. Kahneman [17] showed that individuals would develop and form their subjective perceptions and evaluations of risks when they receive information from the news regarding these risks, making these risks more prominent in an individual’s mind. Additionally, data breach information dissemination frequency was found to impact data protection awareness [19] significantly.

Despite the potential of data breach publicity on ISA, as far as we know, there are no studies reported that have empirically proven how DBR can improve a model’s ability to predict ISA.

III. METHODOLOGY

A. STUDY DESIGN

For data collection, an online survey questionnaire was formulated according to the factors we wanted to include in the hierarchical regression model for testing, including the demographic and known factors studied by literature, and the new factor, DBR, that we are interested in investigating.

Known factors of ISA were adopted based on previous literature as the research framework, except for the Data Breach Publicity factor, which was proposed for investigation in this study. The survey instruments were adopted and adapted from several previous works of literature [23], [29], [39], [61], [64]–[71].

The scale for the target variable, i.e., ISA, was adopted based on Bulgurcu et al.’s [23] research dimensions of GISA, which measures the employee’s general information.
security knowledge and comprehension. All the factors were measured using four items and employing a 5-point Likert scale ranging from ‘Strongly Disagree’ (1) to ‘Strongly Agree’ (5).

A survey test run was also carried out before officially rolling out the survey. A total of 8 respondents were involved in the test run to provide their suggestions and opinion regarding the survey questionnaire. The distribution of their age group falls between 20 to 50 years old across different ethnicities. In addition, the respondents come from different working backgrounds, such as three respondents were reported from Banking and Finance fields, two were working in the Food and Beverage field, and the other three were from Telecommunications field.

Several refinements were done based on the test run from the respondents’ opinions and recommendations; for instance, the working industries were expanded to cover a few other prominent working fields, such as service and healthcare. Furthermore, to avoid confusion when reading “other platforms” as compared to social media platforms to stimulate awareness on information security in one of the questions, the “other platforms” terms were changed to “TV/radio/newspapers”. Apart from that, other comments stated that the questionnaire is well-designed and understandable.

The questionnaire was distributed and managed from June to August 2020, using Google Forms, to examine individual’s Information Security Awareness (ISA). Participants were not required to have an email or Google account to access the Google Forms distributed. The survey was done through a snowball sampling approach. Respondents were recruited from invitations through social media platforms and email solicitations. The responses were kept private and confidential to assure the respondents’ anonymity. The respondents were requested to provide their demographic characteristics comprising of their age, ethnicity, gender, and employment by industry sector.

To serve as validation, approval for the study and the overall research procedure was obtained from the Sunway University Research Ethics Committee (Reference No.: SUREC 2020/074).

B. MODELLING AND ANALYSIS DESIGN

We performed a 3-stage hierarchical regression analysis to determine if DBR significantly improves a model’s ability to predict ISA and its magnitude in influencing ISA by considering all other known factors (i.e., demographic and known factors). As contrasted with multivariate linear regression, this hierarchical approach was assessed by testing the change in R-square ($R^2$) from one stage to the next. We built three regression models in three stages using this model comparison framework by adding previously known ISA factors studied by prior literature, i.e., demographic and known factors, as variables to a preceding model in each stage.

Our interest in this study is to conclude whether the newly added variable, i.e., DBR, significantly improves explaining ISA. The $R^2$ indicates the proportion of variance in ISA, which DBR accounts for in the final stage of the multiple regression analysis [75]. The first stage model (Model in Stage-1) includes demographic information such as age, gender, ethnicity, and the working industry. In the next stage (Model in Stage-2), we added and replicated known variables (namely the prior studied influential ISA factors discussed in Section II-B) reported by previous research. In the final stage (Model in Stage-3), we added the variable of DBR that we are interested in to answer if DBR can significantly improve a model’s ability to predict ISA by considering all other known factors. The 3-stage hierarchical regression model is presented as follows:

\[
\text{Model in Stage-1:} \\
\text{ISA} = \text{Intercept}_1 + \text{demographic variables} \\
\text{Model in Stage-2:} \\
\text{ISA} = \text{Intercept}_2 + \text{demographic variables} + \text{known ISA variables} \\
\text{Model in Stage-3:} \\
\text{ISA} = \text{Intercept}_3 + \text{demographic variables} + \text{known ISA variables} + \text{DBR} \tag{1}
\]

Because our first research question (RQ1) aims to answer if DBR can significantly improve a model’s ability to predict ISA, in other words, our interest is to determine if the final Model in Stage-3 explains ISA better than Model in Stage-2. If there is a statistically significant difference of $R^2$ between Model 2 and 3 (i.e., $R^2$ change), it can be concluded that the added DBR explains ISA beyond the demographic and known factors in Model Stage-2. Our first research question (RQ1) can be answered by referring to this conclusion. Therefore, we do not consider the options of entering DBR into Stage-1 or Stage-2 models without considering the demographic and known ISA factors. The significance level for the model procedures was set to $p < 0.05$

For answering our second research question (RQ2) of how DBR affects ISA, we will observe the coefficient ($\beta$) of DBR in the Stage-3 model. The coefficient ($\beta$) indicates how much ISA is expected to increase when the DBR variable increases by one, holding all the demographic and known variables constant.

IV. RESULTS

A. DATA COLLECTED

Table 1 presents the demographics of respondents. A total collection of 529 responses were gathered in this survey. The gender population was considered equal, comprising 263 (49.7%) male and 266 (50.3%) female respondents. The age distribution for the respondents ranges from below 21 years to 51 years old and above.

The respondents’ mean and standard deviation (SD) age were 36.7 and 11.0, respectively, showing a considerably balanced distribution whereby there was no age group dominating the sample. Similarly, with the occupation factor, we
observed no dominance among the industries. We observed that most of our respondents were Malaysian Malay and Malaysian Chinese, which totaled 413 respondents (78.1%), followed by Malaysian Indians. As there was no data available showing specifically the population distribution between the age of 18 years old and above from the Malaysian Statistics Department, it was unfeasible to derive and confirm the statistical significance of the racial balance in the ratio of the Malaysian population.

Based on the multigroup analysis on the ISA score in Table 1, the ‘21-30’ years old respondents were the most aware of information security (ISA mean = 4.30). Female respondents scored higher ISA than males (4.18). The Malay ethnicity respondents were the most information security-aware (4.39). The respondents working in the Banking & Finance industry scored the highest ISA with a mean score of 4.42 compared to other industries, followed by the Telecommunication and Information Technology industries.

### B. VALIDITY AND RELIABILITY

Table 2 represents the questionnaire instruments utilized to evaluate each item with their respective loadings, average variance extracted (AVE), and Composite Reliability (CR). All the item loadings were significant, in which all the constructs do not fall below 0.7, the recommended minimum cut-off value [74].

The AVE value measures variation captured by a factor as a construct concerning the amount of variation due to measurement error. Finally, the value of CR is a measure of internal consistency in scale questionnaire items. Moreover, reliability and internal consistency validation were also performed for the latent variables based on Cronbach’s Alpha scores with the minimum threshold cut-off value of 0.7.

The CR assessment scores indicated that the items met the prerequisites of accounting good reliability and were well-developed. Results also demonstrated that the AVE values exceeded the suggested threshold of 0.50 [70]. Therefore, we fully adopted or partially adapted questionnaire items of ISA, ISP, SETA, ISK, NEX, SMI, PEB, ORC, and SEC listed from prior studies presented in Section III-B.

For the Data Breach publicity factor as a new construct, we developed the questionnaire items as follows:

- **DBR1**: The publicity of data breach incidents allows me to be aware of the potential risks and consequences if the information is not appropriately protected.
- **DBR2**: The publicity of data breach incidents draws my attention to implement necessary security measures to protect my information.
- **DBR3**: Sharing publicized data breach incidents with my families and friends is important to help them increase information security awareness.
- **DBR4**: It is essential for the relevant authorities to increase the publicity of data breach incidents timely to reduce financial loss in individuals and organizations.

### C. PRELIMINARY ANALYSES

Preliminary analyses were performed to confirm there was no violation of the four assumptions of linear regression [72], which are: i) normality (to determine if our sample is well-modelled by a normal distribution), ii) linearity (to check if there is a relationship between a factor variable and ISA significantly without bias), iii) multicollinearity (to confirm there are no factor variables that are too highly correlated with each other), and iv) homoscedasticity (to ensure a condition in which the variance (or error) in our regression model does not vary much as the value of the factor variables change, in other words, the spread of the errors is constant across the modelling process.). Except for a multicollinearity violation between age and working experience, no other significant violations were encountered. Hence
parametric tests were conducted. Consequently, the ‘working experience’ variable was discarded for the subsequent analyses.

**D. CORRELATION ANALYSIS**

We performed correlation analyses to identify the strength of the linear relationship between the main factor variables. Table 3 presents the correlation matrix, which examines the relationship between ISA, ethnicity (ENC), age (AGE), gender (GEN), industry (IDT), ISP, SETA, ISK, NEX, SMI, PEB, ORC, SEC, and DBR. Because a correlation matrix is symmetrical, only half of the correlation matrix is displayed. As there are no highly correlated variables (correlation coefficient ≥ 0.8) with each other in the model, we ruled out the potential multicollinearity problem.

1) **ISA AND DEMOGRAPHIC FACTORS**

Referring to Table 3, there is inconclusive evidence about the significance of the association between ISA and Ethnicity and Gender because their effects did not achieve statistical significance (p < 0.05). While the relationship between ISA and age was statistically significant, the small magnitude of −0.09 indicates the association between ISA and age is trivial, practically not correlated. On the contrary, the correlation between ISA and industry is 0.16, implying little association between the two. The Kruskal-Wallis H test was performed to gain a better insight into the association between ISA and industry. Based on the test results, there was a divergence in the respondents’ ISA across different industries. We observed respondents from industry sectors of ICT (Information Technology and Telecommunications) and Banking & Finance (mean scores of both sectors = 4.4) were likely to have higher ISA, followed by Healthcare (mean score = 4.3), compared to other sectors (mean scores ranging from 4.0-4.1).

2) **ISA AND NON-DEMOGRAPHIC FACTORS**

There were significant associations between ISA and other non-demographic factor variables, with the magnitude of correlation coefficients between 0.5 and 0.7 (that is, Table 3 cells highlighted in grey), suggesting that those variables can be considered moderately positively correlated. In addition, it is found that the data breach publicity (DBR) posed the strongest strength in the relationship with ISA significantly, compared to all other variables. This observation indicates a strong positive association between ISA and DBR.

**E. HIERARCHICAL REGRESSION MODELLING**

A summarized 3-stage hierarchical regression is presented in Table 4. The 3-stage hierarchical regression was applied to examine the extent to which factor variables predicted ISA. The summarized table also presents the estimates for the parameters encompassed in the final predictive model. The coefficients (β) of the final regression model Stage-3 described the relationship of each studied factor held against ISA mean score, while all other variables are kept constant.

<table>
<thead>
<tr>
<th>Items</th>
<th>Questions</th>
<th>Loading</th>
<th>CR</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISA1</td>
<td>Information Security Awareness (partially adopted from [23])</td>
<td>0.826</td>
<td>0.870</td>
<td>0.627</td>
</tr>
<tr>
<td>ISA2</td>
<td></td>
<td>0.757</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISA3</td>
<td></td>
<td>0.812</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISA4</td>
<td></td>
<td>0.770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISP1</td>
<td>Information Security Policy (adapted from [65] [66])</td>
<td>0.700</td>
<td>0.826</td>
<td>0.543</td>
</tr>
<tr>
<td>ISP2</td>
<td></td>
<td>0.739</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISP3</td>
<td></td>
<td>0.743</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISP4</td>
<td></td>
<td>0.763</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SETA1</td>
<td>Security, Education, Training &amp; Awareness (adapted from [34])</td>
<td>0.831</td>
<td>0.908</td>
<td>0.712</td>
</tr>
<tr>
<td>SETA2</td>
<td></td>
<td>0.882</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SETA3</td>
<td></td>
<td>0.831</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SETA4</td>
<td></td>
<td>0.829</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISK1</td>
<td>Information Systems Knowledge (adapted from [69])</td>
<td>0.816</td>
<td>0.866</td>
<td>0.618</td>
</tr>
<tr>
<td>ISK2</td>
<td></td>
<td>0.811</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISK3</td>
<td></td>
<td>0.766</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISK4</td>
<td></td>
<td>0.749</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEX1</td>
<td>Negative Experience (partially adapted from [68])</td>
<td>0.772</td>
<td>0.861</td>
<td>0.608</td>
</tr>
<tr>
<td>NEX2</td>
<td></td>
<td>0.779</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEX3</td>
<td></td>
<td>0.789</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEX4</td>
<td></td>
<td>0.778</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMI1</td>
<td>Social Media Influence (adapted from [64])</td>
<td>0.708</td>
<td>0.839</td>
<td>0.566</td>
</tr>
<tr>
<td>SMI2</td>
<td></td>
<td>0.820</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMI3</td>
<td></td>
<td>0.751</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMI4</td>
<td></td>
<td>0.725</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEB1</td>
<td>Peer Behavior (adapted from [65])</td>
<td>0.757</td>
<td>0.851</td>
<td>0.587</td>
</tr>
<tr>
<td>PEB2</td>
<td></td>
<td>0.787</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEB3</td>
<td></td>
<td>0.779</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEB4</td>
<td></td>
<td>0.742</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORC1</td>
<td>Organizational Culture (adapted from [71])</td>
<td>0.770</td>
<td>0.859</td>
<td>0.603</td>
</tr>
<tr>
<td>ORC2</td>
<td></td>
<td>0.771</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORC3</td>
<td></td>
<td>0.773</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORC4</td>
<td></td>
<td>0.792</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEC1</td>
<td>Security Culture (adapted from [65])</td>
<td>0.700</td>
<td>0.845</td>
<td>0.578</td>
</tr>
<tr>
<td>SEC2</td>
<td></td>
<td>0.807</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEC3</td>
<td></td>
<td>0.770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEC4</td>
<td></td>
<td>0.760</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBR1</td>
<td>Data Breach Publicity</td>
<td>0.719</td>
<td>0.857</td>
<td>0.600</td>
</tr>
<tr>
<td>DBR2</td>
<td></td>
<td>0.796</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBR3</td>
<td></td>
<td>0.781</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBR4</td>
<td></td>
<td>0.800</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The effects of demographics information, discovered by prior studies to predict ISA [61], [73], were tested at Stage 1. Our results on the demographics information model were statistically significant (p = 0.001) but only explained 2.7% of the variation of ISA. As highlighted in prior
TABLE 3. Correlation matrix: gender, age, ethnicity, industry, ISA, ISP, SETA, ISK, NEX, SMI, PEB, ORC, SEC, and DBR.

<table>
<thead>
<tr>
<th></th>
<th>ISA</th>
<th>ENC</th>
<th>AGE</th>
<th>GDR</th>
<th>IDT</th>
<th>ISP</th>
<th>SETA</th>
<th>ISK</th>
<th>NEX</th>
<th>SMI</th>
<th>PEB</th>
<th>ORC</th>
<th>SEC</th>
<th>DBR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENC</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>-0.09*</td>
<td>-0.13**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDR</td>
<td>0.07</td>
<td>0.13**</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDT</td>
<td>0.16**</td>
<td>0.20**</td>
<td>-0.18**</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISP</td>
<td>0.65**</td>
<td>0.02</td>
<td>-0.08</td>
<td>0.09*</td>
<td>0.17**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SETA</td>
<td>0.57**</td>
<td>0.04</td>
<td>-0.14**</td>
<td>-0.03</td>
<td>0.19**</td>
<td>0.62**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISK</td>
<td>0.58**</td>
<td>0.09*</td>
<td>-0.22**</td>
<td>-0.02</td>
<td>0.23**</td>
<td>0.56**</td>
<td>0.63**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEX</td>
<td>0.60**</td>
<td>-0.01</td>
<td>-0.002</td>
<td>0.04</td>
<td>0.16**</td>
<td>0.54**</td>
<td>0.47**</td>
<td>0.47**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMI</td>
<td>0.63**</td>
<td>0.15**</td>
<td>-0.19**</td>
<td>0.03</td>
<td>0.20**</td>
<td>0.59**</td>
<td>0.55**</td>
<td>0.62**</td>
<td>0.61**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEB</td>
<td>0.60**</td>
<td>0.06</td>
<td>-0.09*</td>
<td>-0.01</td>
<td>0.17**</td>
<td>0.67**</td>
<td>0.77**</td>
<td>0.64**</td>
<td>0.50**</td>
<td>0.63**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORC</td>
<td>0.58**</td>
<td>0.12**</td>
<td>-0.23**</td>
<td>0.01</td>
<td>0.21**</td>
<td>0.64**</td>
<td>0.64**</td>
<td>0.59**</td>
<td>0.46**</td>
<td>0.60**</td>
<td>0.72**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEC</td>
<td>0.72**</td>
<td>0.05</td>
<td>-0.12**</td>
<td>0.03</td>
<td>0.16**</td>
<td>0.70**</td>
<td>0.70**</td>
<td>0.61**</td>
<td>0.55**</td>
<td>0.61**</td>
<td>0.72**</td>
<td>0.66**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBR</td>
<td>0.75**</td>
<td>0.06</td>
<td>-0.09*</td>
<td>0.14**</td>
<td>0.09*</td>
<td>0.67**</td>
<td>0.44**</td>
<td>0.49**</td>
<td>0.61**</td>
<td>0.57**</td>
<td>0.54**</td>
<td>0.59**</td>
<td>0.64**</td>
<td></td>
</tr>
</tbody>
</table>

Note: *p < 0.05; **p < 0.001

TABLE 4. Summary of the hierarchical regression analysis.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Cumulative Adjusted R², F-change, p value-change</th>
<th>Simultaneous (the final model in Stage-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entered at Stage-1</td>
<td>0.027, F(4, 524) = 4.6, 0.001</td>
<td>Coefficient (β) 0.012, p-value 0.652</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>0.028, 0.289</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.001, 0.977</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>0.018, 0.499</td>
<td></td>
</tr>
<tr>
<td>Industry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entered at Stage-2</td>
<td>0.588, F(8, 516) = 100.1, &lt; 0.001</td>
<td>Coefficient (β) 0.016, p-value 0.686</td>
</tr>
<tr>
<td>ISP</td>
<td>0.091, 0.036</td>
<td></td>
</tr>
<tr>
<td>SETA</td>
<td>0.092, 0.012</td>
<td></td>
</tr>
<tr>
<td>ISK</td>
<td>0.075, 0.035</td>
<td></td>
</tr>
<tr>
<td>NEX</td>
<td>0.120, 0.002</td>
<td></td>
</tr>
<tr>
<td>SMI</td>
<td>-0.030, 0.525</td>
<td></td>
</tr>
<tr>
<td>PEB</td>
<td>-0.038, 0.348</td>
<td></td>
</tr>
<tr>
<td>ORC</td>
<td>0.254, &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>SEC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entered at Stage-3</td>
<td>0.067, F(1, 515) = 111.4, &lt; 0.001</td>
<td>Coefficient (β) 0.413, p-value &lt; 0.001</td>
</tr>
<tr>
<td>DBR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data Breach Publicity (DBR) was hypothesized as a sub-component of ISA, and it was entered at Stage 3. The result shows that it explains an additional 6.7% of the variation in ISA. The final regression model accounted for a total of 68.2% variance (R²) in ISA. The F-change with a significance of p < 0.001 means that data breach publicity added in Model Stage-3 significantly improved the ISA prediction. In Table 4, the β values and p values represent the results of the final complete model in Stage-3. The coefficient (β) of DBR achieves the highest score of 0.413 compared with other

...
known factors. This finding indicates that one unit change of DBR will significantly ($p < 0.001$) affect 0.413 times the mean change in ISA while holding other variables in the model constant. Remarkably, regardless of the initial significant correlation with ISA, the Age (AGE), Industry (IDT), Information Security Policy (ISP), Peer Behavior (PEB), and Organizational Culture (ORC) were insignificant ($p > 0.05$) covariates in the final regression model.

V. DISCUSSION

Not reported in prior studies, our main finding confirmed and answered this study’s first research question (RQ1) that data breach publicity is a significant factor for improving a model’s ability to predict ISA by observing the positive increment of $R^2$ change in the hierarchical regression model. Furthermore, after controlling for demographic factors and other prior studies reported factors, data breach publicity predicted an additional 6.7% of ISA variability in the hierarchical regression analysis.

The DBR achieves the highest coefficient ($\beta$) with a positive 0.413, comparing with prior known factors. Hence, we can infer that data breach publicity has the highest positive effect on ISA than prior known factors. This finding answered our second research question (RQ2).

Contradictory to previous literature findings by Schlienger and Teufel [59] and Da Veiga et al. [56], as indicated in our final model, organizational culture was an insignificant factor, though security culture showed otherwise. However, these findings align with the recent study of Wiley et al. [61] that explained a more compound association between ISA, security culture, and organization culture, where security culture poses a mediator role in the relationship between ISA and organization culture. Furthermore, the correlation analysis in our study also confirmed the association between ISA, security culture, and organization culture. This finding implies that regardless of organizational culture, better security culture is possibly a stronger predictor of an individual’s ISA.

Significant relationships between ISA and factors of negative experience, information systems knowledge, and SETA were found in the final model. Similar results were reported by Haeussinger & Kranz [29] and Mani et al. [39]. On the other hand, the model analysis shows no significant relationship between ISA and peer behavior, which is inconsistent with previous research that stated employees develop security awareness through conversation and discussion with their peers [39], [40]. Further examination is required to investigate the interplay of peer behavior influence with other factors, especially the socio-environmental antecedents of ISA, to understand this contradiction.

Also, no significant relationship was found between ISA and security policy provision in the final model. This finding contradicts prior studies that suggested that information security policy provisions positively impact an individual’s ISA. Organizations in Malaysia are devoting less emphasis to promoting the cruciality of information security policies in their respective organizations, which is a significant cause of why most employees are unaware of their company’s security policies [12]. Without a well-established and informed security policy, employees may not be able to implement robust security measures that could help evaluate and identify potential weaknesses of information security [13].

Prior research findings [29], [39] revealed the significant effect of information obtained from traditional media on ISA. However, there is inadequate research attention dedicated to exploring the influence of social media on ISA despite other application fields presenting evidence of social media’s impact on individual’s awareness [50], and this communication medium is gradually replacing traditional media [49]. Our finding confirms the significant relationship between ISA and social media influence.

Our study shows that demographic factors were not significant covariates in the final hierarchical model based on our hierarchical analysis. However, compared with previous research conducted in a different country [61], a significant relationship was found between ISA, age, and gender. Thus, the variation in national characteristics between different countries may affect the relationship between ISA and demographic factors. However, the validity of this proposition needs further investigation to confirm.

A. CONTRIBUTION AND IMPLICATIONS

The results of our study encompass both theoretical and practical implications. From the aspect of theoretical implications, our findings contribute to the theoretical research work by presenting empirical evidence and confirming the association between ISA and data breach publicity, which has not been reported in the study of ISA theoretically or empirically.

We discovered a significant positive association between data breach publicity (DBR) and ISA. Essentially, individuals with the knowledge of data breach incidents were likely to have higher ISA. This finding revealed that DBR is a significant predictor and possesses the highest effect in shaping an individual’s ISA. By increasing the efforts of publicizing data breach incidents, individuals are inclined to learn from these experiences and eventually develop a stronger sense of protection and vulnerability to online harms [17], [18]. Therefore, relevant authorities such as governments and media corporations should stress their role in increasing data breach incidents’ publicity.

According to our research findings, security information derived from social media was discovered to carry a significant effect on an individual’s awareness of information security. Thus, our findings show that social media is an effective and potent tool in building information security awareness. Moreover, because of the rapid evolution of technology, people today tend to engage and spend more time on social media. This evidence is crucial from a practical perspective. As a result, government and organizations would be more efficiently utilizing social media to disseminate data.
breach incidents that would be more widespread and quicker for individuals to obtain the latest information.

This study presents a significant contribution to information security literature. The results of our study can be applied in a future investigation to examine other related factors and the effects of their interplay. For instance, it is perhaps practically useful to discover the association between DBR and other known factors relevant to information dissemination, for example, media influence, SETA, and how it may impact the perception of an individual’s negative experience. More prominently, our findings also present a practical contribution. Government and industrial organizations can rely on empirical evidence to guide their strategy in disseminating information security newscasts, including data breach publicity through an effective medium such as social media.

This study also contributes to a practical exemplar case of how a 3-stage hierarchical regression approach can confirm if a newly studied factor has a significant impact in improving a predictive model’s ability and how the new factor impacts the prediction. Finally, using hierarchy regression provides a methodological demonstration of how this approach can be applied in a similar research scenario.

B. LIMITATIONS AND FUTURE WORK

Our research encompasses several limitations. First, the data collection approach of our research was geographically constrained to Malaysia. Henceforth, to populate the findings, future researchers should consider the study in other countries and regions, where cultural variation might be an external factor that yields a distinct perspective on findings. Second, one of the plans to extend this study is collecting randomized sample data with more balanced demographics and dimensions. Moreover, the present research does not explore the interlinking effects that each influential factor possesses. Therefore, future research should consider investigating the effects between the factors affecting ISA, as these predictors might have intertwined impacts when affecting an individual’s ISA.

Another avenue for future work is that researchers could examine the effects of data breach publicity on influencing an individual’s information privacy and security compliant behavior since the proposed factor was a relatively new finding in the field. Besides, demographic characteristics of different data samples might reflect divergence in findings. For example, the correlation analysis of our study shows respondents from different industries posed different levels of ISA. Therefore, a future study may be conducted with a larger data sample to analyze different industry sectors.

VI. CONCLUSION

Our research provides novel evidence that data breach publicity has a significant positive impact on a model’s ability to predict ISA, and it also demonstrates the highest positive effect on ISA compared to prior known factors. Relevant authorities such as governments and media corporations should stress their role to increase the publicity of data breach incidents. To improve employees’ ISA, organizations should strategize their ISA programs to include data breach publicity. Using hierarchy regression contributes to a methodological demonstration of how this approach can be applied in a similar research scenario that requires empirical testing on an unknown factor’s effect if it can improve a model’s ability for prediction by considering prior known factors.

REFERENCES


HUI NA CHUA (Senior Member, IEEE) received the Ph.D. degree from the University of Nottingham, U.K. She joined academia after working in the industry for more than ten years. She is currently an Associate Professor with the Department of Computing and Information Systems, Sunway University. She had contributed and managed various Malaysian national research projects, such as the FRGS and MDEC funds and industry-based funded research and development projects when/where she worked as a full-time Technical Lead and filed more than ten patents. Her research interests include applied machine learning, text analytics, data mining, data management, and information security/privacy.

JIA SHENG TEH received the Bachelor of Science degree (Hons.) in information systems (business analytics) from Sunway University and Lancaster University, U.K. He is currently a Business Analyst with Deloitte Risk Advisory, IT, and specialized assurance field. His research interests include information privacy, data mining, data modeling and visualization, and big data analysis methods.

ANTHONY HERBLAND received the Ph.D. degree from the University of Hertfordshire, U.K. He is currently an Educational Technologist and a Senior Lecturer with the University of Hertfordshire. He has published numerous articles contributing to health science, e-learning, and statistical methodology. His research interests include data analysis and statistical methods.

...