Journal of
Cybersecurity and Privacy

MDPI

Article

The Cybersecurity Awareness INventory (CAIN): Early Phases of Development of a Tool for Assessing Cybersecurity Knowledge Based on the ISO/IEC 27032

Giorgia Tempestini [1], Ericka Rovira [2], Aryn Pyke [2] and Francesco Di Nocera [1,*]

[1] Department of Planning, Design, and Technology of Architecture, Sapienza University of Rome, 00196 Rome, Italy
[2] Department of Behavioral Sciences and Leadership, U.S. Military Academy, West Point, NY 10996, USA
* Correspondence: francesco.dinocera@uniroma1.it

Abstract: Knowledge of possible cyber threats as well as awareness of appropriate security measures plays a crucial role in the ability of individuals to not only discriminate between an innocuous versus a dangerous cyber event, but more importantly to initiate appropriate cybersecurity behaviors. The purpose of this study was to construct a Cybersecurity Awareness INventory (CAIN) to be used as an instrument to assess users' cybersecurity knowledge by providing a proficiency score that could be correlated with cyber security behaviors. A scale consisting of 46 items was derived from ISO/IEC 27032. The questionnaire was administered to a sample of college students (N = 277). Based on cybersecurity behaviors reported to the research team by the college's IT department, each participant was divided into three groups according to the risk reports they received in the past nine months (no risk, low risk, and medium risk). The ANOVA results showed a statistically significant difference in CAIN scores between those in the no risk and medium-risk groups; as expected, CAIN scores were lower in the medium-risk group. The CAIN has the potential to be a useful assessment tool for cyber training programs as well as future studies investigating individuals' vulnerability to cyberthreats.

Keywords: cybersecurity; cyber threats; awareness; knowledge; questionnaire; CAIN

check for updates

Citation: Tempestini, G.; Rovira, E.; Pyke, A.; Di Nocera, F. The Cybersecurity Awareness INventory (CAIN): Early Phases of Development of a Tool for Assessing Cybersecurity Knowledge Based on the ISO/IEC 27032. *J. Cybersecur. Priv.* **2023**, *3*, 61–75. https://doi.org/10.3390/jcp3010005

Academic Editor: Giorgio Giacinto

Received: 5 December 2022
Revised: 30 January 2023
Accepted: 13 February 2023
Published: 16 February 2023

1. Introduction

Cybersecurity is of profound importance given the dependence on technological tools for daily activities (i.e., banking, health, news, etc.), and the increase in cyber-attacks worldwide. Just in the last few years, with the global COVID-19 pandemic and the "remote" lifestyle, there has been a dramatic increase in cyber-attacks [1]. Data show that during the pandemic there was a 50.1% increase in cyber-attacks and 30,000 cyber-attacks regarding issues closely related to COVID-19 [2].

Unsafe behaviors enacted by users are a major contributing factor to the increase in breaches of cybersecurity [3]. IBM Security Service [4], for example, suggests that human error was responsible for approximately 95% of security breaches; a percentage that obviously also varies depending on the context in which the analysis is performed. El-Bably [5] investigated the behaviors of employees at a number of companies in the UK and US, noting that errors with respect to security processes were present in 43% of problematic situations within the work context. These behaviors included, for example, using weak passwords, sharing sensitive information, failing to update antivirus programs, and paying little attention to security notifications from the systems themselves.

Although there is an increasing focus on improving cybersecurity training, some studies have demonstrated that this training is not always effective. For example, Lorenz and collaborators [6] observed that despite receiving training on the best cybersecurity behaviors to adopt (i.e., complex, non-repeating passwords across devices), only a small percentage of experimental participants (2%) followed the guidance provided in the training phase when assessed at follow-up. These results could represent a starting point for understanding how training, albeit in-depth, is not enough to get users to adhere to secure practices.

In the current research, to investigate users' awareness or familiarity of cybersecurity gained by experience (namely, knowledge), we developed and administered a cybersecurity knowledge assessment tool to a sample of college students. The questionnaire is not aimed at estimating the reported behavior, but only the level of knowledge to identify markers of vulnerability.

Before specifically addressing the research design in this paper, the following sections will briefly address: (1) the topic of cybersecurity; and (2) the questionnaires already available in the literature for assessing knowledge and/or security behaviors.

2. Cybersecurity Awareness

ISO/IEC 27032 (2012) defines cybersecurity as the "protection of confidentiality, integrity and availability of information in cyberspace" [7], p. 4. Cybersecurity represents one of the main challenges that have emerged due to the development of increasingly pervasive information technologies in our daily activities [8].

It has become increasingly evident that the weakest link in cybersecurity is the human element [9]. Individuals do not always possess adequate knowledge about security protocols and are not always aware of the consequences that may arise from a failure to comply with procedures. A review by Chaudhary and colleagues [10] identified the most commonly used metrics to measure cybersecurity awareness (CSA) and noted that among all the factors measured, behavior, attitude and knowledge were the most popular. The term cybersecurity knowledge refers to awareness or familiarity with security policies, procedures, standards, guidelines, regulations, strategies, technologies/systems and best practices, gained by experience of a fact or situation. Behavior includes all those behaviors that a user engages in when interacting in cyberspace. A user can decide whether to adopt his or her cybersecurity knowledge by implementing a secure behavior. Awareness in this sense includes all these aspects and assumes precisely the meaning of the degree of knowledge and understanding of users about the importance of information security, so that sufficient levels of information security controls can be implemented to protect data and networks of the organization [11].

Ben-Asher and Gonzales [12] developed an Intrusion Detection System (IDS) to understand how individuals with different knowledge backgrounds performed at detecting security-damaging events. The results demonstrated that increased knowledge of cybersecurity facilitates the proper detection of malicious events and reduces the misclassification of benign events as malicious. Another factor that we would expect to influence cybersecurity behavior and motivate users to gain knowledge is the user's level of concern about protecting their information. Despite the fact that users do express some concern for the protection of their personal information, there is a phenomenon referred to in the literature as the "privacy paradox": even those who express the highest degree of concern often intentionally disclose their personal information online [13]. This could be due to the fact that people are unwilling to invest the time and energy necessary to ensure their information is protected as their concerns are eclipsed/overridden by the various ways in which users benefit from the dissemination of personal data (such as increasing their social value, being able to take advantage of offers, discounts, etc.). These benefits take on a negative connotation as they lead the user not only to implicitly disregard a whole range of risks, but also to ignore explicit warnings that call their attention to potential negative consequences.

Despite such warnings, many users may not be fully aware of the level of risk they are incurring, because a lack of knowledge is a double-edged sword. Low cybersecurity knowledge not only means users are less aware of sound practices and how to perform them, but also a lack of knowledge interferes with the ability of these users to recognize risky situations when they encounter them and calibrate the level of risk involved. Furnell and colleagues [14] surveyed home computer users about their estimated level of risk and found that many people were unaware of many cyber risks. IT novices, in particular, lack the knowledge to protect themselves from attacks, and as a result, if users are unaware or do not understand security risks, they are more likely to fall victim to misbehavior. Researchers have demonstrated how an individual's perception of information security and self-awareness have a decisive impact on both decision-making and behavior [15]. There are cases, however, in which knowledge of the risks and awareness of one's actions is not enough because one is not aware of what the correct behavior to adopt may be.

This difficulty may also depend on the tools used, as the more complex the tools become, involving encryption, use of access keys, and digital signatures, the greater the likelihood that people will try to circumvent the systems. This may also be due to repeated exposure to security alerts (such as those reminding people to update their antivirus) which result in a habituation effect, whereby the user no longer pays attention to what is presented. Bravo-Lillo and collaborators [16,17] precisely studied this habituation effect with respect to security dialog boxes: users ignore important messages due to repeated exposure to these dialog boxes. The user simply presses a button automatically, signaling to the system that they have read the warning, but without really paying attention to the context presented. This study highlights how a "fulfillment" mode of the security feature design (i.e., security alerts must be presented and the user must signal that he or she has read them) fails to take into account all of the processes involved to ensure that individuals are implementing safe behaviors.

3. Survey Measures in Cybersecurity

A variety of survey measures of cybersecurity awareness, habits, attitudes, and behaviors have already been proposed in the literature. However, they have several limitations. Some authors have created questionnaires that include a few items about security behaviors in order to correlate them with measures other than knowledge (e.g., personality measures [3]). Such tools do not focus on the relationship between cybersecurity awareness/knowledge and behaviors, but rather on how individual characteristics may influence the enactment of more or less secure behaviors.

Another limitation of existing questionnaires is that of the theoretical background on which the questionnaire is based. Theoretical backgrounds are often diverse and sometimes the frameworks used are outdated, as in the case of the Parkerian Hexad Model developed in the 1990s, on which the questionnaire by Arpaci and Sevinc [8] is based. Theories taken from different fields and sometimes not even specific to the domain cybersecurity are often used. The OBBSQ questionnaire by Li and collaborators [18], for example, uses a theoretical model that is not related to behaviors or knowledge about cybersecurity. Rather, the model was intended to explain non-adherence to healthy behaviors in general (Health Belief Model). Although there is value in investigating whether models can generalize across domains, there are many distinctive aspects to cybersecurity knowledge and contexts that are key to understanding cybersecurity behavior. For example, in comparing physical security behaviors and cybersecurity behaviors, individuals may be more likely to routinely close their curtains than cover their webcams.

Another major limitation is the lack of a clear definition of what aspects are intended to be investigated. In the case of the CS-S [8], for example, most of the proposed factors contain items related to secure and unsecure behaviors, with only one factor composed of items related to knowledge/awareness.

Existing cybersecurity questionnaires have undoubtedly increased knowledge and promoted research in this field. They have also been administered to very large samples from organizational settings, e.g., [8,18,19]. However, there generally seems to be little consensus among these various surveys with respect to the underlying theory, the aspects to be investigated, and the most appropriate response scale to use.

Given the limitations of the questionnaires reported above, we decided to adopt a more comprehensive, atheoretical, and updated solution for developing a specific measure of cybersecurity awareness: relying on the guidance provided by ISO/IEC 27032:2012 (Information Technology–Security Techniques–Guidelines for Cybersecurity).

Using a standard reference (ISO/IEC 27032) for item structuring compensates for the limitations of some of the questionnaires just described. First, it is an international standard that provides for its application to different countries, although these may differ in technological readiness or technological advancement. By creating a framework and identifying a specific terminology, it allows those involved in cybersecurity to collaborate in solving cybersecurity problems, providing a shared vocabulary and shared knowledge that it is good for all to acquire if they intend to interface with such a complex and multidetermined topic.

Second, the purpose of the ISO/IEC 27032 is to provide a guide to improve knowledge and address issues related to cybersecurity, offer an overview of the topic, identify the parties and roles involved in cybersecurity, and differentiate it from other types of security. The standard highlights these particularities and distinguishes cybersecurity from other security domains, such as: information security, network security, Internet security, and critical information infrastructure protection (CIIP).

4. The Cybersecurity Awareness Inventory (CAIN)

4.1. Overview

Users must understand and properly use systems to ensure the effectiveness of any security strategy implemented [20]. The Cybersecurity Awareness Inventory (CAIN) described below is aimed at one thing: assessing users' knowledge related to cybersecurity by providing a single proficiency score. CAIN items do not require subjects to express opinions or report their computer behaviors; rather, it is a knowledge test to which the user can provide both right and wrong answers. The advantage of survey measures is evident when considering their ease of use, minimal cost, and usefulness in obtaining meaningful information from users that would otherwise be inaccessible. Moreover, questionnaires can be administered inexpensively to large samples, making it easy to monitor the "health" status of an organization. The CAIN is intended as a proficiency test providing information about people's awareness of cyber risks, to be used as a measure of individuals' vulnerability. However, the ability to implement the correct behavior in risky situations may be unrelated or only partly related to knowledge. Indeed, the CAIN score (i.e., sum or proportion of correct answers) is an index to be used in a great variety of research and applied settings by correlating it to performance data, individual difference measures, and behavioral reports of security practices.

In the current research, to investigate users' level of knowledge and to identify markers of vulnerability, we developed and administered a cyber knowledge assessment tool to a sample of college students. The student sample affords an independent measure of vulnerability in that some students have been flagged by IT department security procedures as having been exposed to/engaged in cybersecurity risk, while other students in the sample have not been flagged.

4.2. Methods

4.2.1. Development of the Questionnaire

The items of the CAIN were generated from the ISO/IEC 27032:2012 (Information Technology–Security Techniques–Guidelines for Cybersecurity). The document was carefully examined, and salient components were transformed into potential items. The potential items were constructed trying to cover all sections of the standard, starting with the most general definitions and ending with more specific aspects, such as those concerning stakeholders.

An initial set of 116 potential items was evaluated by four judges (two authors and two experts not involved in this study), who rated all items on a scale from 1 (keep) to 3 (drop). The criteria for dropping were: lack of comprehensibility to the general public (e.g., "Cross-border legal issues are not a problem in the investigation"), lack of generality (the item was specific to organizational rather than personal issues: e.g., "It is recommendable to share with other stakeholders the reports related to risks"), and redundancy (the item's content was already addressed by other items: e.g., "A computer can be remotely hacked" was removed and "Computers can be controlled remotely" was kept).

Raters assessed each item using a 1 to 3 score. Scores were then averaged, and items that needed attention were further checked in order to decide whether to keep or drop them. After these sessions, the set was reduced to a draft version of 51 items. A True/False response scale was implemented. That version was then administered to a small convenience sample to test the comprehensibility of the items. This phase led to a further reduction of the items to 46.

The following is the description of the ISO/IEC 27032:2012 sections, along with the corresponding items that were generated and retained. It should be noted that sections do not represent dimensions or latent factors: they are only the product of a categorization leading to the table of contents of the ISO document.

Terms and definitions. This section reports terms and definitions given in ISO/IEC 27000 (Information Security Management Systems Family of Standards) that apply for the purposes of the ISO/IEC DIS 27032 (items #1, #2, and #3). Some of these terms and definitions (and related examples reported in this section) provided a cue for generating items.

Overview. This section provides an overview related to introductory aspects of ISO/IEC 27032, especially regarding how individuals manage their online identity (items #4, #5, #6, and #7). Some people are more careful about what they share online while others tend to upload online (e.g., on social networks) personal information that third parties can also appropriate. The items derived from this subsection refer specifically to the knowledge that both individuals and organizations have with respect to sharing their personal information and awareness of how their actions involve consequences of their own responsibility.

Assets in cyberspace. ISO/IEC 27032 defines assets as "anything of value to the individual and the organization that [...] is classified into personal and organizational assets" (p. 15). This section specifically reports items that were generated from the information on personal assets (items #8, #9, #10, and #11).

Threats against the security of cyberspace. This section includes items that address the threats that exist in cyberspace in relation to the previously mentioned assets (items #12, #13, #14, #15, #16, #17, #18, #19, #20, #21, #22, and #23). Threats are divided into two areas: threats to personal assets and threats to organizational assets. The former mainly relate to identity issues revolving around the leakage or theft of personal information, while the latter relate to those threats that organized cybercrime groups deployed to cause damage to organizations (e.g., stealing an organization's URL and reviewing data to another organization unrelated to the real one). The items in this section refer to both types of threats that are reported by ISO/IEC 27032, as well as related to how these attacks are generated. *Roles of stakeholders in cybersecurity.* The items that were generated from this section deal with the roles that stakeholders have in cybersecurity, starting with those of consumers and ending with those of organizations and providers (items #24, #25, and #26). This subsection is important because stakeholders play an active role in the development and use of the Internet; thus, it is important that we have knowledge of what their respective roles are. *Guidelines for stakeholders.* This sub-section of ISO/IEC 27032 focuses on its three main aspects: (1) security guidance for consumers, (2) internal management of an organization's information security risk, and (3) security requirements that suppliers should specify for consumers to implement (items #27, #28, #29, #30, #31, and #32). The standard proposes recommendations for consumers and organizations that were taken as a "cue" for item formulation.

Cybersecurity controls. Once cybersecurity threats have been identified and appropriate guidelines have been drafted, controls should be selected and implemented that support the security requirements. This section provides an overview of the major cybersecurity controls so as to support the guidelines set forth in this standard. Thus, the items that were generated from this section cover the implementation of these control layers, server protections, and end-user controls, and controls against social engineering attacks (items #33, #34, #35, #36, #37, #38, #39, #40, #41, #42, #43, #44, and #45).

Framework of information sharing and coordination. Cybersecurity incidents often cross national geographic and organizational boundaries, and the speed of information flow and changes that occur following the incident often leave responding individuals and organizations with a limited time to act. It seems important to report a general item related to these issues so as to assess whether there is awareness regarding the timing in which the consequences of an attack may show up (item #46).

4.2.2. Participants

A total of 277 college students (76 females, 192 males, and 9 preferred to not report; mean age = 19.6 years, st.dev. = 1.5) for whom vulnerability data were available (see the section below) volunteered and participated in this study. Participants academic majors ranged across liberal arts, sciences, and engineering. All students are mandated to take specific U.S. Department of Defense cybersecurity training annually because they are accessing government systems. Specifically, all students must take and pass an annual, on-line training called the "Cyber Awareness Challenge" which presents video vignettes on topics such as Social Engineering, Social Networks, and Website Use, and includes a multiple-choice test. Ethical approval was obtained from the IRB at the U.S. Military Academy.

4.2.3. Risk Level Measure

To obtain an independent measure of participants' level of cybersecurity risk (vulnerability), we accessed via the institution's IT department automated reports generated by Microsoft™ Azure Active Directory Identity Protection, which detects risk based on users' system interactions (https://docs.microsoft.com/en-us/azure/active-directory/identity-protection/overview-identity-protection, accessed on July 2022). We used reports based on our users' prior nine months of activity. User risk may be detected if, for example, a user's activity is abnormal or suspicious, or evidence suggests that their credentials have been leaked (e.g., the user's email and password were found posted in a public location). One reason leaked credentials could occur is as a consequence of a user providing personal information in response to a phishing email. Azure's internal Identity Protection risk-level calculations are not entirely transparent, but risk is classified into three tiers: low, medium, and high.

The higher the level for a flagged event, the higher the confidence that the user or sign-in is compromised. Occasionally, a single user might have more than one risk event flagged in the report. If a user was associated with more than one event, and those events had different risk levels, the user was classified into the highest associated risk level. Our sample could then be subdivided into three groups: (i) those with no risk detected (N = 162); (ii) those categorized as low risk (N = 68); and those categorized as medium risk (N = 45).

Two participants were categorized as high risk, and they were omitted from analyses involving the risk-level factor. In the analyses, we will test the directional hypothesis that lower CAIN scores will correspond to higher risk.

4.3. Procedure

Participants filled out the questionnaire posted online using the Qualtrics platform. In addition to completing the questionnaire, information was collected about the risks participants incurred in the past 9 months.

5. Data Analysis and Results

5.1. CAIN Scores

Overall, participants correctly responded to most of the items (Table 1). A few items had a low proportion of correct responses. For example, item #33 "*Cookies are a mechanism to* *steal sensitive information*" received a correct response, FALSE, only in half of the cases. That item may be somewhat ambiguous, as cookies may be used to steal sensitive information but are not by themselves a malicious piece of code. Item #40 "*Firewalls slow down the performance of devices*" also showed correct responses in half of the cases. However, this is more likely a knowledge gap rather than an item-related issue.

Table 1. Proportion of correct responses for each CAIN item.

Item	Correct Response	Correct Responses
1. Some programs can send a large number of emails automatically.	TRUE	98%
2. Computers can be controlled remotely.	TRUE	99%
3. It is not possible to secretly install software on a computer.	FALSE	87%
4. Personal data shared on the Internet (including social networks) can be eventually accessed by others even if the user deletes them.	TRUE	99%
5. Criminal organizations use spyware to exploit any weakness in cyberspace.	TRUE	99%
6. Activities of a single user cannot affect the cybersecurity of an entire system.	FALSE	82%
7. Users are responsible for their actions in cyberspace, even when such actions are unintentional.	TRUE	93%
8. Avoiding using real names or pictures keeps a user safe from possible threats.	TRUE	73%
9. Virtual money has no value in the real world.	FALSE	91%
10. A hacker stealing a user's virtual information cannot access the user's sensitive information.	FALSE	92%
11. Banks may accept transactions using virtual currency.	TRUE	78%
12. Credit information can be sold on the black market.	TRUE	98%
13. Cyber criminals can prevent users from accessing their personal applications and services.	TRUE	91%
14. Cyber criminals are not interested in the personal information recorded in online games.	FALSE	93%
15. An organization's URL is a critical resource.	TRUE	90%
16. Phishing emails can be used to steal users' personal information.	TRUE	97%
17. Documents received by email cannot convey viruses.	FALSE	93%
18. Cyber-attacks cannot be delivered by legitimate websites.	FALSE	88%
19. Individuals may receive cyber-attacks from people they know.	TRUE	96%
20. A free Wi-Fi Internet access can be used to steal personal information.	TRUE	97%
21. Hackers can remotely capture the keypresses from a computer (e.g., for stealing passwords).	TRUE	93%
22. Intranets (organizations' private networks) are secure from cyber-attacks.	FALSE	80%
23. It is not possible to sneak into another computer just by knowing the IP (Internet Protocol) address.	FALSE	78%
24. A user should report to the appropriate authority if they inadvertently gain access to a site that requires authorization.	TRUE	94%
25. When buying or selling online, one may unintentionally participate in criminal transactions.	TRUE	90%
26. An organization is not responsible for its employees' security education.	FALSE	83%
27. Reading the security policy of a website is useless to prevent cyber-attacks.	FALSE	82%

Table 1. *Cont.*

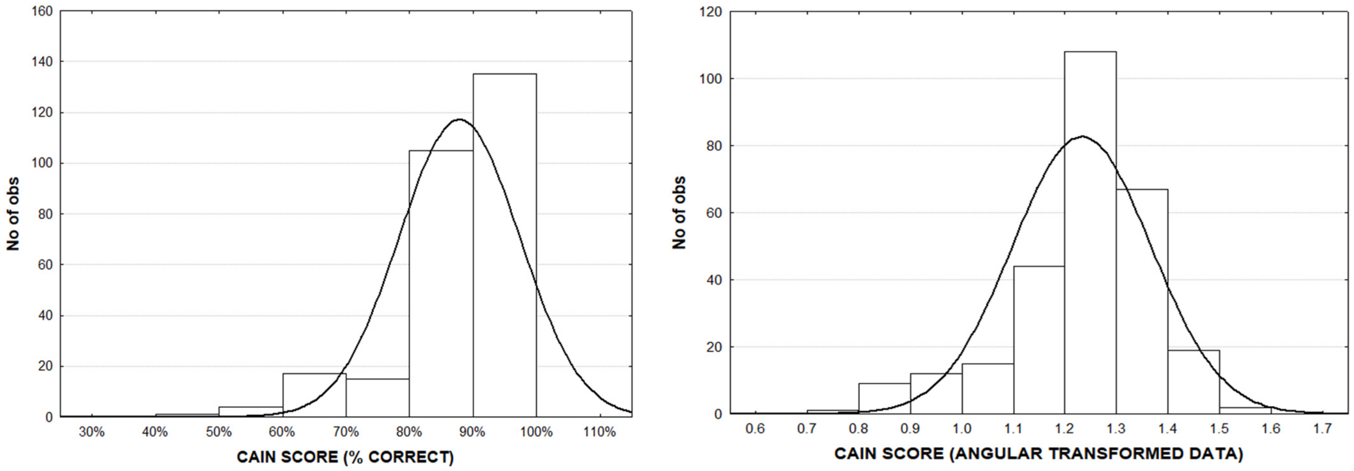
Item	Correct Response	Correct Responses
28. When a user is not aware of the risks, it is better to ask someone trusted before acting online.	TRUE	96%
29. Online stores have no specific rules regarding security management.	FALSE	79%
30. Some online payment mechanisms are not reliable.	TRUE	95%
31. A blog owner is not responsible for materials uploaded by other users.	FALSE	59%
32. An organization does not need an authorization to release sensitive information of its employees on any website.	FALSE	78%
33. Cookies are a mechanism to steal sensitive information.	FALSE	48%
34. Software updating is not relevant for cybersecurity.	FALSE	91%
35. Antivirus software works even if temporarily disabled.	FALSE	75%
36. Unsolicited opened window can transmit a virus.	TRUE	92%
37. Malicious pieces of code (scripts) hidden in some websites can infect a device.	TRUE	97%
38. Browsers have tools for protecting against cyber-attacks.	TRUE	92%
39. Current operating systems have embedded firewalls.	TRUE	92%
40. Firewalls slow down the performance of devices.	TRUE	49%
41. A fake account can be created to contact a user with the aim to steal their sensitive information.	TRUE	96%
42. Emails requesting username and password are legitimate.	FALSE	88%
43. A freebies offer can be a way to steal personal information.	TRUE	93%
44. A two-factor authentication (e.g., password + sms code, authenticator app) is a security tool.	TRUE	96%
45. A device involved in suspicious activity should be quarantined (i.e., isolated for preventing infections of other devices).	TRUE	83%
46. If a cyber-attack happens today, the consequences may be visible weeks away.	TRUE	94%

The average total score was 40.26 (st.dev. = 4.34). Internal consistency (reliability) of the scale was assessed through the Kuder–Richardson's coefficient (KR20), which is a special case for Cronbach's alpha when items are dichotomous (i.e., have only two response levels, shown here as TRUE or FALSE). The index is equivalent to Chornbach's alpha and provides values ranging from 0 to 1. The closer the KR20 coefficient is to 1.0 the greater the internal consistency of the items in the scale. In this very case, the KR20 coefficient was 0.78, indicating a fairly high reliability. Item-total statistics indicated that none of the items was crucial (if eliminated) for dramatically increasing reliability (see Table 2). However, eliminating item #33, that we reported as ambiguous, would increase reliability to 0.80, and to 0.81 if item #40 was also eliminated.

Table 2. KR20 coefficient if item deleted.

1	0.784	13	0.781	25	0.776	37	0.780
2	0.783	14	0.772	26	0.776	38	0.776
3	0.778	15	0.785	27	0.776	39	0.782
4	0.781	16	0.783	28	0.780	40	0.798
5	0.783	17	0.775	29	0.777	41	0.777
6	0.771	18	0.773	30	0.780	42	0.777
7	0.779	19	0.775	31	0.788	43	0.778
8	0.790	20	0.781	32	0.776	44	0.777
9	0.782	21	0.779	33	0.798	45	0.781
10	0.774	22	0.770	34	0.772	46	0.780
11	0.787	23	0.784	35	0.779		
12	0.783	24	0.778	36	0.780		

The distribution of the CAIN scores was asymmetrical (skewness = −1.89) and narrow (kurtosis = 3.49) (Figure 1). Therefore, scores were transformed using the arcsin square root method for successive analyses. This type of transformation is typically used when dealing with proportions/percentages, and corrects distribution issues (see Figure 1a,b). Percent of correct responses were instead plotted for readability.

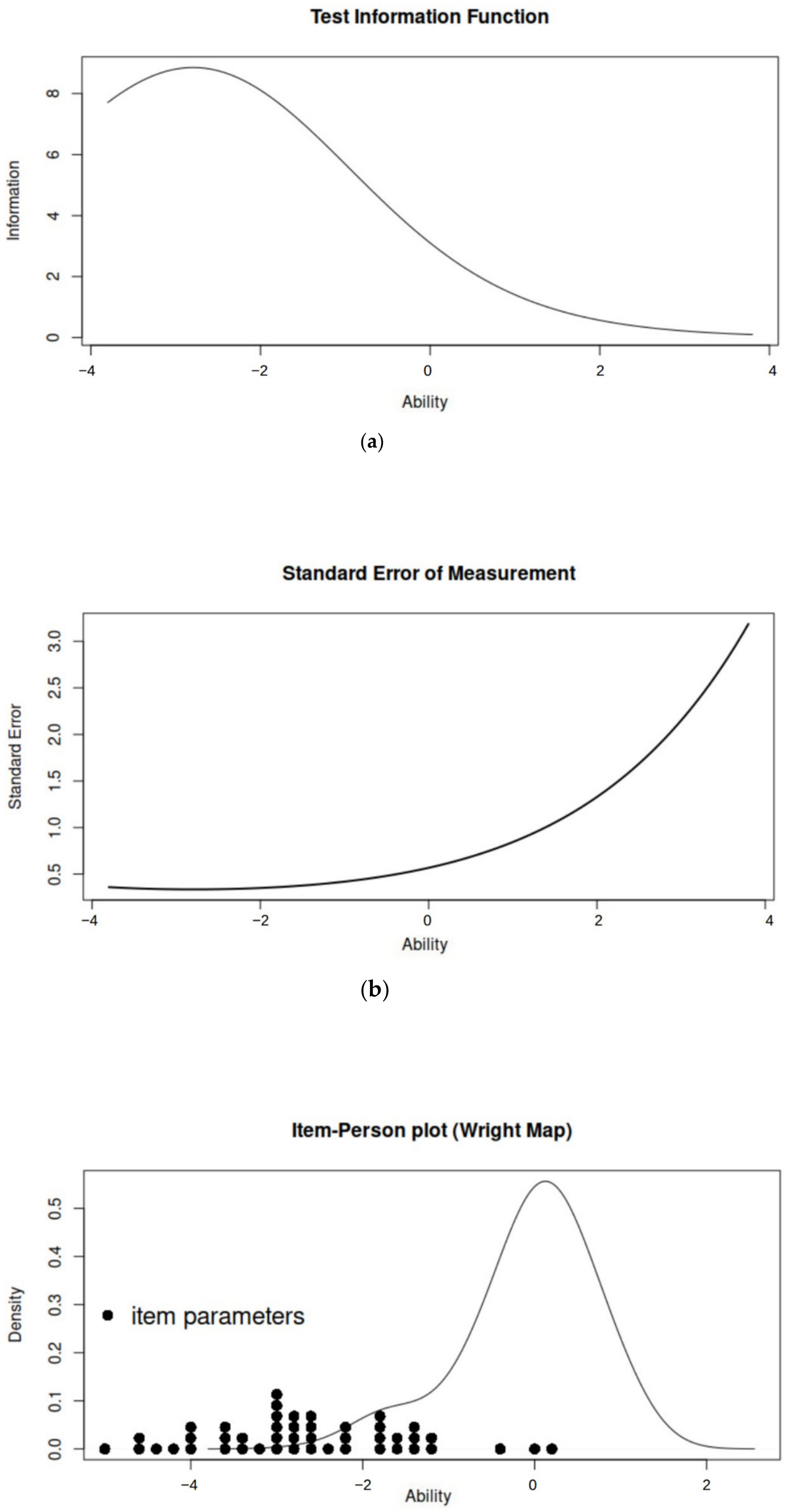


(**a**) (**b**)

Figure 1. (**a**) Distribution of the CAIN scores. The line indicates the expected normal distribution; (**b**) Distribution of the CAIN scores (angular transformed data). The line indicates the expected normal distribution.

5.2. Item Response Model Analysis

We run a one-parameter Item Response Model analysis (Rasch Model) using all 46 items of the CAIN and on the whole sample (N = 277). Analyses were run in R [21] with the package ltm [22]. Global fit indices of the model were not completely satisfactory: Bootstrap (rep = 1000) Goodness-of-Fit using Pearson chi-squared p = 0.001, AIC = 7856.41, BIC = 8023.11. However, the overall Cronbach alpha was good (alpha = 0.78). The total amount of information, assuming that the cybersecurity ability values have an interval ranging from −4 to 4, was about 63.01%, meaning that the CAIN is able to reliably measure the latent cybersecurity ability score of about 63% of the whole sample of participants (Figure 2a).



(**a**)

(**b**)

(**c**)

Figure 2. (**a**) Test information plot; (**b**) precision of the test measure (standard error of the test) as function of the estimated cybersecurity vulnerability score; (**c**) item–person plot or Wright Map.

Moreover, as shown in Figure 2b, the CAIN reaches the greatest precision (lowest standard error) when it is used to estimate low cybersecurity ability levels. Both Figure 2a,b confirm that the range of the estimated item difficulties (95% C.I.: −3.78; −1.85) is centered on the range of the estimated person abilities (95% C.I.: −1.03; 0.76). As shown in Table 3, proportions of the correct answer for each item were quite high, and the corresponding estimated Rasch difficulty parameters were all negative, indicating that the 46 CAIN items assessed the easiest segment of the latent cybersecurity ability. Figure 2c, the item–person map, confirms this conclusion. As can be observed, the estimated item difficulties (shown as black dot in the graph) were all located below the average cybersecurity vulnerability estimated on the whole sample (i.e., the zero on the graph represents the estimated average cybersecurity ability).

Table 3. Item difficulty and item fit statistics (Monte–Carlo simulation).

	Proportion	Logit	Difficulty	se	z	$\chi2$	Pr (>$\chi2$)
1	0.982	3.996	−4.415	0.462	−9.567	5.537	0.672
2	0.989	4.515	−4.943	0.588	−8.403	3.867	0.796
3	0.870	1.901	−2.185	0.199	−10.985	9.742	0.429
4	0.986	4.223	−4.649	0.513	−9.057	4.398	0.780
5	0.986	4.223	−4.649	0.513	−9.057	2.231	0.966
6	0.823	1.538	−1.775	0.179	−9.920	16.201	0.093
7	0.931	2.609	−2.961	0.255	−11.621	5.775	0.742
8	0.726	0.973	−1.126	0.158	−7.148	18.936	0.055
9	0.910	2.311	−2.637	0.228	−11.559	13.424	0.111
10	0.917	2.402	−2.738	0.236	−11.607	13.028	0.138
11	0.776	1.244	−1.439	0.166	−8.645	16.167	0.105
12	0.978	3.810	−4.226	0.424	−9.964	8.915	0.283
13	0.906	2.267	−2.590	0.225	−11.527	7.624	0.581
14	0.928	2.553	−2.902	0.250	−11.630	17.411	0.034
15	0.895	2.146	−2.457	0.216	−11.402	8.440	0.497
16	0.971	3.515	−3.920	0.371	−10.554	4.447	0.851
17	0.928	2.553	−2.902	0.250	−11.630	10.190	0.288
18	0.881	2.001	−2.296	0.205	−11.182	9.488	0.394
19	0.957	3.095	−3.479	0.310	−11.232	12.488	0.113
20	0.971	3.515	−3.920	0.371	−10.554	4.593	0.831
21	0.931	2.609	−2.961	0.255	−11.622	6.883	0.635
22	0.801	1.395	−1.613	0.173	−9.350	26.978	0.002
23	0.776	1.244	−1.439	0.166	−8.646	3.953	0.970
24	0.939	2.728	−3.089	0.267	−11.576	5.474	0.783
25	0.903	2.226	−2.544	0.221	−11.490	13.376	0.124
26	0.830	1.588	−1.832	0.181	−10.100	5.089	0.914
27	0.823	1.538	−1.775	0.179	−9.920	8.664	0.552
28	0.964	3.285	−3.679	0.336	−10.955	6.188	0.652
29	0.787	1.307	−1.512	0.169	−8.952	4.744	0.922
30	0.950	2.933	−3.308	0.290	−11.418	9.017	0.354
31	0.588	0.358	−0.407	0.145	−2.802	23.247	0.039
32	0.780	1.264	−1.463	0.167	−8.747	6.943	0.795
33	0.480	−0.080	0.104	0.143	0.726	39.473	0.003
34	0.913	2.355	−2.687	0.232	−11.586	20.986	0.012
35	0.747	1.084	−1.255	0.161	−7.801	7.239	0.786
36	0.921	2.450	−2.790	0.240	−11.622	8.868	0.437
37	0.975	3.653	−4.061	0.395	−10.291	10.114	0.210
38	0.924	2.501	−2.844	0.245	−11.630	5.762	0.764
39	0.924	2.501	−2.845	0.245	−11.630	2.670	0.986
40	0.491	−0.036	0.054	0.143	0.373	42.400	0.002
41	0.964	3.285	−3.679	0.336	−10.956	16.242	0.041
42	0.881	2.001	−2.296	0.205	−11.183	7.747	0.617
43	0.928	2.553	−2.902	0.250	−11.630	7.381	0.583

Table 3. *Cont.*

	Proportion	Logit	Difficulty	se	z	χ2	Pr (>χ2)
44	0.964	3.285	−3.679	0.336	−10.955	8.498	0.377
45	0.827	1.563	−1.804	0.180	−10.011	12.522	0.227
46	0.942	2.792	−3.158	0.274	−11.537	8.256	0.451

5.3. Relation of CAIN Scores to Risk Level

The proportion of correct responses for each subject (arcsin transformed) was used as a dependent variable in an ANOVA design with risk level (none, low, and medium) as the factor. As shown in Figure 3, the numerical difference between the medium-risk group and the other two is visible, but the overall main effect did not reach statistical significance (F2,272 = 2.160, p = 0.117). Nevertheless, the Duncan post hoc analysis demonstrated a significant difference between the no-risk and medium-risk levels (p = 0.029) and a marginal significance between low- and medium-risk levels (p = 0.081).

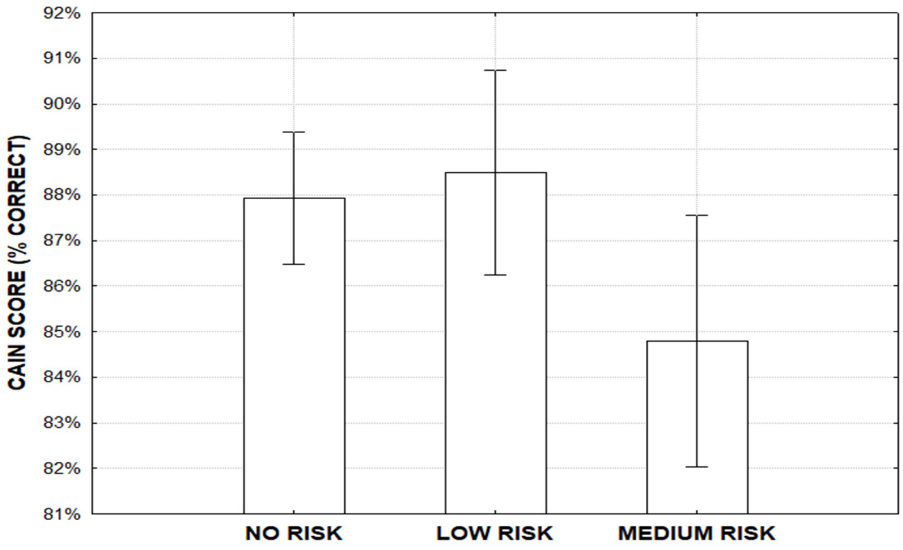


Figure 3. CAIN scores by risk levels.

6. Discussion

The purpose of this study was to develop an instrument to assess users' awareness (here intended as a synonym of knowledge) of cybersecurity issues and to provide a proficiency score that could be eventually correlated with other measures (e.g., risky behaviors). Knowledge influences behavior, although it is not the only variable that determines it. Estimating knowledge is an important step in reducing user vulnerability by intervening at the organizational level. For example, Razaque and colleagues [23] developed a web-based Blockchain-enabled cybersecurity awareness program in which estimating the user knowledge was a key element. The user was tested on common cybersecurity questions and tasks, such as e-mail phishing and weak password policies, and then the program tested whether the user had a good knowledge base by conducting tests on advanced cybersecurity topics.

Generally speaking, one would expect that knowledge plays a crucial role in predicting and mitigating vulnerability. Therefore, a standard measurement tool is highly desirable.

To make sure that the measure was as standardized as possible, we started by reviewing the questionnaires already existing in the literature and found that many of them are composed of either a very large [18,19] or a very small [3] set of items, making them either cumbersome for the user or superficial for the purpose of testing users' vulnerability. Most importantly, none of the cybersecurity surveys reviewed are considered broad/comprehensive cybersecurity knowledge tests. For example, the CS-S is mainly oriented to obtain information about users' behaviors, and includes only a few knowledge items.

Our aim, instead, was to develop a questionnaire specifically oriented to measure users' cybersecurity awareness, composed of a fair number of items covering as many aspects as possible, and based on a comprehensive knowledge base rather than a specific theoretical approach. Based on the analysis of ISO/IEC 27032, the standard guideline in the field of cybersecurity, a pool of 46 items was selected and great attention was devoted to wording, by trying to use terminology that would be understood by users with different cybersecurity knowledge levels.

Data analysis demonstrated that participants attained fairly high CAIN scores and generated a distribution of scores that was skewed and narrow. The high percentage of correct responses may indicate that college students have fairly in-depth knowledge regarding aspects of cybersecurity. Our particular population of students also underwent specific U.S. Department of Defense cybersecurity training annually because they are accessing government systems. Additionally, in some instances high scores could also result from real-life situations that students personally encountered: for example, for items #2 and #45, the percentage of correct answers was numerically higher in the medium-risk group. This could be because those who were issued a medium-risk report may have been contacted by IT to take action on their devices, through remote IT control or quarantining their devices, thereby affording medium-risk users first-hand knowledge about such practices.

In terms of the relation of CAIN scores to risky behaviors, although the omnibus comparison of the three risk level groups (no risk, low risk, and medium risk) did not reach statistical significance, a significant difference between the medium-risk group and the other two was found. Those who have more knowledge of procedures, consequences of their actions, policies, and responsibilities are more careful about the actions that they take, protecting their devices from different kinds of threats. Future studies with larger and more balanced samples may address the existence of a subset of items showing internal consistency and for which differences between subgroups could be investigated.

Our results supported our expectation that the medium-risk group would be associated with the lowest CAIN (knowledge) scores. However, we had also initially expected that no-risk users would have higher CAIN scores than low-risk users. It was interesting that scores for the low-risk group were actually relatively high and very similar to those of the no-risk group. One possible explanation for this similarity is that relatively high knowledge may not lead users to avoid risk completely. As mentioned previously, having more knowledge allows the user to better calibrate the level of risk at play. Users who had been flagged for only low risk events may have been aware of the risk associated with their behavior, but also the fact that it was only a low-level risk, and so they deemed it an acceptable level of risk in the face of other goals. Another possible reason the scores may not have revealed a clear difference between the low-risk and no-risk groups is the possibility of a ceiling effect given the sample's high performance overall, which is likely a result of annual training. The low-risk group may have had less knowledge than the no-risk group. One possibility to add more variability in the scores could be to have users not only provide an answer (TRUE/FALSE) but also to rate their confidence in their answer (e.g., between 1 and 100).

One of the most difficult aspects in estimating the validity of cybersecurity questionnaires is the availability of concurrent behavioral measures for the users. Many validation approaches compare the questionnaire data to subjective measures (e.g., subjective reports on the users' secure behaviors). Here, we attempted to use an objective indicator of the secure behavior, namely the risk detected by Microsoft™ Azure Active Directory Identity Protection. Although further validation of CAIN is still pending, our results are encouraging. Future studies should administer the CAIN to other categories of users (e.g., workers, traditional college students, or other age groups) for shedding light on the general distribution and could maybe lead to standardized scores. In its current form, the CAIN is a reliable questionnaire showing a fairly high internal consistency, and it is a good candidate measure in any cybersecurity study in which the vulnerability level of the users should be assessed.

7. Conclusions

Our aim was to develop a questionnaire for assessing peoples' awareness of cybersecurity knowledge and threats. CAIN items were generated based on a standard document that represents the state of the art in this field. An initial administration to a sample of college students demonstrated that the scale is reliable and we also provided indications of its validity, although more studies are needed in that regard. In its present form, the CAIN could provide a useful addition in any cybersecurity study for assessing individuals' vulnerability to cyberthreats. Applications are endless, as the CAIN score can be correlated with any objective and/or subjective measures, and can be used for comparing groups of users, and for assessing improvements in cybersecurity awareness (e.g., after training). As a final note, we need to acknowledge that as we are submitting this article, the ISO/IEC 27032 is currently undergoing a revision. When the final version of the standard is published, it will be necessary to make a comparison between the current version and the new standard. This comparison will help us to evaluate whether the CAIN should be kept in its present form or whether some items should be changed, removed, or added.

Author Contributions: G.T., E.R., and F.D.N. conceptualized the study and developed the items; E.R. managed and supervised data collection; G.T. wrote the initial draft; F.D.N. and A.P. ran the data analysis. All authors contributed to the interpretation of the results and manuscript editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We are grateful to COL Christopher Morrell for his valuable technical support on this project and to Fabio Presaghi for his precious inputs on item response theory.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lallie, H.S.; Shepherd, L.A.; Nurse, J.R.; Erola, A.; Epiphaniou, G.; Maple, C.; Bellekens, X. Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic. *Comput. Secur.* **2021**, *105*, 102248. [CrossRef] [PubMed]
2. World Economic Forum. COVID-19 Risks Outlook: A Preliminary Mapping and its Implications. 2020. Available online: https://www.weforum.org/reports/covid-19-risks-outlook-a-preliminary-mapping-and-itsimplications (accessed on 1 July 2022).
3. Kennison, S.M.; Chan-Tin, E. Taking risks with cybersecurity: Using knowledge and personal characteristics to predict self-reported cybersecurity behaviors. *Front. Psychol.* **2020**, *11*, 3030. [CrossRef] [PubMed]
4. IBM Security Services 2014 Cyber Security Intelligence Index. Available online: https://media.scmagazine.com/documents/82/ibm_cyber_security_intelligenc_20450.pdf (accessed on 1 July 2022).
5. El-Bably, A.Y. Overview of the Impact of Human Error on Cybersecurity based on ISO/IEC 27001 Information Security Management. *J. Inf. Secur. Cybercrimes Res.* **2021**, *4*, 95–102. [CrossRef]
6. Lorenz, B.; Kikkas, K.; Klooster, A. "The four most-used passwords are love, sex, secret, and god": Password security and training in different user groups. In Proceedings of the International Conference on Human Aspects of Information Security, Privacy, and Trust, Las Vegas, NV, USA, 21–26 July 2013; pp. 276–283.
7. *ISO/IEC 27032:2012*; Information Technology e Security Techniques e Guidelines for Cybersecurity. International Organization for Standardization: Geneva, Switzerland, 2012.
8. Arpaci, I.; Sevinc, K. Development of the cybersecurity scale (CS-S): Evidence of validity and reliability. *Inf. Dev.* **2021**, *38*, 026666692199751. [CrossRef]
9. Chandarman, R.; Van Niekerk, B. Students' cybersecurity awareness at a private tertiary educational institution. *Afr. J. Inf. Commun.* **2017**, *20*, 133–155.

10. Chaudhary, S.; Gkioulos, V.; Katsikas, S. Developing metrics to assess the effectiveness of cybersecurity awareness program. *J. Cybersecur.* **2022**, *8*, tyac006. [CrossRef]
11. Shaw, R.S.; Chen, C.C.; Harris, A.L.; Huang, H.J. The impact of information richness on information security awareness training effectiveness. *Comput. Educ.* **2009**, *52*, 92–100. [CrossRef]
12. Ben-Asher, N.; Gonzalez, C. Effects of cyber security knowledge on attack detection. *Comput. Hum. Behav.* **2015**, *48*, 51–61. [CrossRef]
13. Barth, S.; De Jong, M.D. The privacy paradox–Investigating discrepancies between expressed privacy concerns and actual online behavior–A systematic literature review. *Telemat. Inform.* **2017**, *34*, 1038–1058. [CrossRef]
14. Furnell, S.M.; Bryant, P.; Phippen, A.D. Assessing the security perceptions of personal Internet users. *Comput. Secur.* **2007**, *26*, 410–417. [CrossRef]
15. Huang, D.L.; Rau, P.L.P.; Salvendy, G.; Gao, F.; Zhou, J. Factors affecting perception of information security and their impacts on IT adoption and security practices. *Int. J. Hum.-Comput. Stud.* **2011**, *69*, 870–883. [CrossRef]
16. Bravo-Lillo, C.; Cranor, L.; Komanduri, S.; Schechter, S.; Sleeper, M. Harder to Ignore? Revisiting {Pop-Up} Fatigue and Approaches to Prevent It. In Proceedings of the 10th Symposium On Usable Privacy and Security (SOUPS 2014), Menlo Park, CA, USA, 9–11 July 2014; pp. 105–111.
17. Bravo-Lillo, C.; Komanduri, S.; Cranor, L.F.; Reeder, R.W.; Sleeper, M.; Downs, J.; Schechter, S. Your attention please: Designing security-decision UIs to make genuine risks harder to ignore. In Proceedings of the Ninth Symposium on Usable Privacy and Security, Newcastle, UK, 24–26 July 2013; pp. 1–12.
18. Li, L.; He, W.; Xu, L.; Ivan, A.; Anwar, M.; Yuan, X. Does explicit information security policy affect employees' cyber security behavior? A pilot study. In Proceedings of the 2014 Enterprise Systems Conference, Shanghai, China, 2–3 August 2014; pp. 169–173.
19. Parsons, K.; McCormac, A.; Butavicius, M.; Pattinson, M.; Jerram, C. Determining employee awareness using the human aspects of information security questionnaire (HAIS-Q). *Comput. Secur.* **2014**, *42*, 165–176. [CrossRef]
20. Furnell, S. Why users cannot use security. *Comput. Secur.* **2005**, *24*, 274–279. [CrossRef]
21. R Core Team. *R: A Language and Environment for STATISTICAL Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022; Available online: https://www.R-project.org/ (accessed on 4 November 2022).
22. Rizopoulos, D. ltm: An R package for latent variable modeling and item response analysis. *J. Stat. Softw.* **2022**, *17*, 1–25.
23. Razaque, A.; Al Ajlan, A.; Melaoune, N.; Alotaibi, M.; Alotaibi, B.; Dias, I.; Oad, A.; Hariri, S.; Zhao, C. Avoidance of Cybersecurity Threats with the Deployment of a Web-Based Blockchain-Enabled Cybersecurity Awareness System. *Appl. Sci.* **2021**, *11*, 7880. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.