2024

# Addressing the Challenge to Measure Information Security Behavior: Toward a Holistic Metric with Scavenger Hunts

Martin Brehmer
*University of Augsburg, Germany*, martin.brehmer@uni-a.de

Raphaela Melanie Stöckl
*University of Augsburg, Germany*, raphaela.stoeckl@uni-a.de

# Addressing the Challenge to Measure Information Security Behavior: Toward a Holistic Metric with Scavenger Hunts

**Research in Progress**

Martin Brehmer[1], Raphaela Stöckl[1]

[1] University of Augsburg, Faculty of Business and Economics, Augsburg, Germany
{martin.brehmer, raphaela.stoeckl}@uni-a.de

**Abstract.** Cyberattacks and data breaches lead to high costs for organizations worldwide. Information security education and training awareness programs are one of the most important countermeasures. Here, assessing individuals' level of information security awareness is a crucial task. Regarding this, one of the major challenges is to measure security behavior, a core dimension of security awareness. This is because it is often assessed indirectly through questionnaires, which could bias metrics. Therefore, our overarching goal is to develop a more holistic metric that considers and integrates actual human behavior. In this design science research study, we present the status quo of our research, namely a prototypical instance for such a measurement approach, and initial meta-requirements based on two design iterations and pilot tests: a scavenger hunt to measure the consequences of real-world interactions, based on the Human-Aspect-of-Information-Security-Questionnaire as a scientific foundation.

**Keywords:** information security awareness, metric, SETA programs, DSR

## 1 Introduction

The annual cost of cyberattacks for German businesses has exceeded EUR 200 billion (Wintergerst 2023) for three consecutive years since 2021, and data breaches have led to an average cost of USD 4.45 million per incident worldwide (IBM 2023). Hereto, information security education, training, and awareness (SETA) programs are one of the major countermeasures (IBM 2023) mitigating these tremendous business risks (Hiscox Ltd 2023). These programs aim to foster secure behavior, typically through classroom or online courses that focus on knowledge transfer as well as hands-on training, such as learning how to store sensitive data. Measuring and interpreting individuals' information security awareness (ISA) level is an essential part of these programs (Kruger and Kearney 2006) as it indicates both the level of awareness and the learning progress within the organization. As a result, evaluated SETA efforts reveal deficiencies in the security chain and indicate areas to improve in subsequent sessions. However, the reliability, accuracy, and transparency of the applied measurement methods,

19th International Conference on Wirtschaftsinformatik,
September 2024, Würzburg, Germany

as well as the expressive power of the results, are arguable and consequently pose an enormous challenge for researchers, security educators, and practitioners (Chaudhary et al. 2022). A key reason is that security awareness assessments are mostly based on questionnaires in research (Rohan et al. 2023; Fertig et al. 2020) and practice (KnowB4 2023) which means that information security behavior (ISB), a core dimension of ISA (Kruger and Kearney 2006), is measured indirectly. SETA programs market leaders like KnowB4 aim to tackle this issue by including behavioral security aspects, e.g., providing an additional "personal risk score", predominantly based on responses to phishing simulations or training participation (KnowB4 2024). However, this neglects other critical threats such as brute force attacks on weak passwords, or dumpster diving, where criminals search for sensitive information in discarded files. In essence, questionnaires fail to measure actual ISB (Fertig et al. 2020; Rohan et al. 2023), and other metrics consider behavior aspects unsatisfactory. This can lead to unwarranted confidence in organizations' resilience against cyber threats due to the inaccurate ISB measurement. For instance, knowing that a found USB stick is potentially malicious does not guarantee that it is not plugged in within the daily work routine. Consequently, organizations are vulnerable to future cyberattacks and data breaches. Beyond practical implications, researchers aiming to improve SETA programs struggle to rigorously evaluate new approaches and generate meaningful theoretical insights. Therefore, research on ISA metrics that considers different aspects of an individual's **actual ISB** (**aISB**) is essential for rigorous ISA assessment (Rohan et al. 2023).

Our overarching research project aims to address this challenge by developing an instance and design principles for a more holistic ISA metric which includes measuring aISB as a critical dimension of ISA. Thereby, we see the potential to strengthen the effectiveness of SETA programs and their value for both practice and research. In this research-in-progress article, we present our initial steps toward this goal by answering the following research questions:

1. *How can actual security behavior be measured in an authentic environment based on a scientifically grounded information security awareness assessment?*
2. *What are initial meta-requirements and next steps for future research toward a more holistic metric for an information security awareness assessment considering both, questionnaires and actual security behavior?*

To answer these questions, we present an assessment approach implemented as a gamified version of the well-accepted and validated Human-Aspect-of-Information-Security-Questionnaire (HAIS-Q) by Parsons et al. (2017). Specifically, we create tasks for a scavenger hunt based on this questionnaire to assess the participants' level of ISA, including aISB, in an authentic environment. This means, participants follow a storyline and complete security-relevant tasks to earn assessment points. In doing so, this article contributes to design science research (DSR) by providing initial insights from an evaluated instance of a novel approach that addresses the real-world problem of inaccurate ISA metrics. Additionally, we outline meta-requirements for developing a more holistic metric and outline the next steps to revise the artifact and achieve our overarching goal. As we follow the DSR methodology of Peffers et al. (2007), this article focuses on two iterations (see Figure 1). The first iteration addresses the process of a) how we derive game tasks from a validated questionnaire to cover knowledge,

attitudes, and actual behavioral aspects of ISA (Parsons et al. 2017), and b) how we implement these tasks within a game prototype to measure aISB including a pilot test. The second iteration includes a) design considerations as well as practical implementation aspects, and b) meta-requirements derived from a second pilot test.


**Figure 1**. Metric design process (f. Peffers et al. (2007)) - dashed lines show ongoing research

## 2 Theoretical Background and Related Work

Measuring ISA is a challenge discussed in several recent literature-review-based studies (Fertig et al. 2020; Chaudhary et al. 2022; Rohan et al. 2023). ISA integrates three core dimensions: security-related "gains in ***knowledge*** and positive changes in ***attitudes*** and ***behaviors***" (Chaudhary et al. 2022) of individuals in organizations. This definition is rooted in the well-accepted knowledge-attitude-behavior model (KABM) of Kruger and Kearney (2006). Later research addresses criticism for lacking context information and a standardized measurement approach (McCormac et al. 2017) by building on these three core dimensions. For instance, Parsons et al. (2017) introduce the HAIS-Q, a modular questionnaire with contextualized items for all three core dimensions, validated in several iterations. However, while it seems feasible to measure knowledge and attitudes through standardized questionnaires, measuring ISB indirectly is heavily criticized as it captures self-reported expected behavior instead of aISB (Rohan et al. 2023). Thus, more mature and reliable ISA scales that consider aISB are required (Rohan et al. 2023), alongside standardized questionnaires (Chaudhary et al. 2022).

So far, according to Chaudhary et al. (2022), who reference the literature review of Fertig et al. (2020), previous studies on metrics mainly address two types of ISA metrics: 'knowledge-based metrics' and 'behavior-based metrics'. Knowledge-based metrics measure, e.g., aspects of the KABM by applying the HAIS-Q. In contrast, behavior-based metrics rely on data logs of security-relevant real-life events, such as the length of an employee's chosen password resulting from a change request. (Chaudhary et al. 2022; Fertig et al. 2020) From our perspective, these terms are not entirely distinct, as knowledge-based metrics also appear to indirectly measure ISB according to the KABM, although not aISB. However, they differ in that behavior-based metrics focus on real-world data rather than knowledge-based reporting. (Chaudhary et al. 2022)

Thus, to develop a more holistic metric, we aim to expand the existing understanding of metrics by building on the concept of knowledge-based metrics, specifically the KABM, while also considering aISB: We apply the HAIS-Q as a theoretical foundation to measure the ***knowledge*** and ***attitude*** dimensions of ISA, and further measure actual ***behavior*** through behavioral data that represent the corresponding HAIS-Q items.

Therefore, we gamify the HAIS-Q (Parsons et al. 2017) by using a digital scavenger hunt (SH) game as a medium. Here, gamification serves as a motivator to a) engage participants in a SETA context that is often perceived as burdensome (Hu et al. 2022),

and b) encourage a careful completion of the assessment tasks, potentially enhancing the survey data quality (Harms et al. 2015). The SH employs gamification through various game-design functionalities adapted for a non-game context (Deterding et al. 2011). Technically speaking, SH is a game concept where players solve a set of tasks while interacting with the real world, e.g., searching for specific objects at a designated location. The term 'digital SH' refers to the delivery method using, e.g., a mobile application to present tasks in a motivating manner and to capture the corresponding players' responses. Thus, we utilize the SH mobile application called 'Actionbound' (AB). This platform allows, for example, the inclusion of *storytelling* (to motivate and inform where to go and what to do), *time pressure (*to foster natural behavior and a strong focus on the given task), simple single- or multiple-choice questions that are rated with *points* (to measure the ISA dimensions ***knowledge*** and ***attitude***). Additionally, this allows for tasks where ***behavior*** can be measured through the consequences of their real-life interactions, also rated with *points*. This means we develop tasks based on the HAIS-Q that have to be solved hands-on in the real world. For instance, one task is to find a specific password to access further information for the game. Here, the SH's players first have to walk to a specific location at the university. Second, once there, various general options are presented to solve the task, e.g., extracting a password from a (prepared) USB Stick that is conveniently available at the players' location, or investigating another location for hidden hints on the password. Third, the players evaluate these options, and fourth, they choose an option. Thereby, each option results in a password, but only one of them is the correct password which is awarded with points for the individuals' ranking in the *leaderboard* (to ensure automatic ISA assessments (Fertig et al. 2020)). In summary, our SH includes game-design elements such as *storytelling*, *time pressure*, scorable *points*, and providing a *leaderboard,* which is known to potentially work in a SETA context (Brehmer and Reinelt 2023).

This research-in-progress article presents the current state of our research, focusing on meta-requirements for rigorously deriving meaningful SH tasks rather than a final metric. In the following, we offer a preliminary instance of how we strived to measure ISA including aISB. However, a final metric that interprets the measurements requires, e.g., an assessment of the validity and confirmability of the scoring system.

## 3 Scavenger Hunt Approach to Measure Security Awareness

In order to develop the SH tasks that are implemented in AB, we applied the HAIS-Q (Parsons et al. 2017) as a scaffolding construct. This means that the overall structure of the SH was based on the HAIS-Q's focus areas and all tasks that have been developed related to the content and purpose of the corresponding HAIS-Q items. With that, we basically followed the process (phase 1) for developing an ISA metric of Rohan et al. (2023), but with a focus on converting an existing scale (HAIS-Q) toward a more realistic measurement approach that considers not only cybersecurity knowledge and attitudes but also aISB as impact indicators for SETA programs (Chaudhary et al. 2022).

Consequently, our **first design iteration** focused on a) developing SH tasks based on the HAIS-Q and b) evaluating the approach with a pilot study as a proof of concept

to detect technical and procedural errors. To derive the SH tasks, we conducted two initial team workshop days. This included reviewing and implementing the tasks. Since the HAIS-Q is modular (Parsons et al. 2017), we considered 4 out of the 7 focus areas (information handling, social media use, email use, password management) to set the duration of the SH at less than 1.5 hours, as the longer the duration, the more likely the results are to be biased. A pilot study was conducted (n=17 participants, students, and research assistants) to find technical and procedural errors and to receive qualitative feedback. The participants reported some technical problems that occurred and found it motivating but slightly too long. It is noteworthy, that no participant mentioned the measurement itself to be burdensome. Examples of SH tasks are provided in Table 1.

**Table 1**. Examples of SH tasks that are derived from HAIS-Q items of Parsons et al. (2017)

| HAIS-Q reference item | SH tasks based on the HAIS-Q items |
|---|---|
| **Focus Category:** Information handling;<br>Sub-category: Inserting removable media.<br>1. If I find a USB stick in a public place, I shouldn´t plug it into my work computer.<br>2. If I find a USB stick in a public place, nothing bad can happen if I plug it into my work computer.<br>3. I wouldn´t plug a USB stick found in a public place into my work computer. | Once you have found the papers and USB sticks, you will receive a cryptic email that gives you hints to find the password needed to open the encrypted file.<br>You have the choice to use one of the USB sticks in the room or go to the office mentioned in the email to get the password. Return the USB stick. What is the password? |
| **Focus Category:** Email use; Sub-category: Clicking on links in emails from known senders.<br>1. I am allowed to click on any links in emails from people I know.<br>2. It´s always safe to click on links in emails from people I know.<br>3. I don´t always click on links in emails just because they come from someone I know. | […] Your mobile phone is making a noise...<br>What was that? - While walking across the campus, you notice that your work colleague, with whom you have already done some projects has sent you an email. Check your student email inbox. […]<br>Get the password for the […] Information Hub, where all information about the university is stored. Enter the password here: […] |

In our **second design iteration**, we focused on a) shortening the game and eliminating technological errors, e.g., an email that participants were supposed to receive during the game did not reach everyone. Further, we strived to b) assess the feasibility of our SH regarding its practical operability and potential for a more holistic ISA metric. To achieve a), we dropped the focus category password management which led to a new duration of ca. 35-45 minutes. For b), we followed the first steps of phase 2 (scale development) from Chaudhary et al. (2022) and considered the criteria of good ISA metrics (Spitzner 2024) for developing a valuable metric approach for practitioners and researchers. To develop a lightweight metric, we adjusted the maximum points for the focus areas among 3 different levels (400 points, 961 points, 1175 points). This enabled us to check whether there is a minimum sum of achievable points without bias. In both tests, the participants were informed that they were in a test situation for research purposes. The use of a pseudonym also ensured no further conclusions concerning individual participants could be drawn. We then conducted a second pilot test (n=21) to receive data for a score comparison. In more detail: All participants completed a pre-test covering the three selected focus areas of the HAIS-Q. Thereby, we calculated Cronbach's α to ensure overall reliability for further comparison. Given that the HAIS-Q is validated (Parsons et al. 2017), the meta-analysis on reported alpha values in leading academic journals by Taber (2018), and considering our small sample sizes,

we state our values to be acceptable: information handling: α = .605; social media use: α = .736; email use: α = .698. Subsequently, the participants played the SH. Then, we checked whether the ISA scale of Kruger and Kearney (2006), e.g. applied by Firsty Arisya et al. (2020), could be applied to interpret the results of the SH scores. This means, we calculated the corresponding score (HAIS-Q vs. SH) of each construct per participant and interpreted it with the help of the HAIS-Q scale. In 28 construct comparisons (out of *3 focus areas* x *n=21* in total) both scales lead to the same conclusion. Then, in 33 comparisons, the SH scale was lower than the corresponding HAIS-Q scale. We state that the SH scale is still valid because the data reveal that these participants lost their points mostly in behavior-based tasks of the SH. This can be an indicator that our approach is successfully measuring and differentiating aISB. In only two comparisons, the SH scaled better than the corresponding HAIS-Q scale, but it must be mentioned that both would have scaled equally if the participants had reached only one point more in the HAIS-Q. Thus, we assume our scales to be comparable to a certain degree. However, the results also indicate that there should be at least ca. 1175 points per AB category to improve overall accuracy (2% avg. difference between HAIS-Q and SH vs. 19-23% avg. diff.). Based on these findings, we suggest six initial meta-requirements for the further development of a more holistic metric which can be seen in Figure 2.

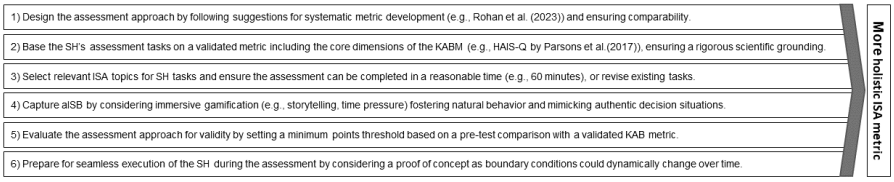


**Figure 2.** Meta-requirements toward a more holistic ISA metric that considers aISB

## 4 Limitations and Future Research

This article provides insight into how to measure ISA, including aISB. Our pilot studies show that the assessment prototype is functional and indicate that the HAIS-Q-based metric might be applicable to our SH scales. Nevertheless, due to the limitations of this study, e.g. the small sample size of our pilot tests, further validation is needed, e.g., by repeating the study with more participants. We also plan to revise the SH prototype toward a more valid reward system. Additionally, expert interviews will be conducted to validate the correctness and practical relevance and to reduce subjectivity that might have influenced the derivation process from the HAIS-Q items to the SH tasks. Moreover, we did not control for existing knowledge. This is consistent with real assessment situations but should be considered as it provides information for further validation of a final metric. We acknowledge that developing such an SH is resource-intensive. However, an SH concept like ours could be reused and customized, e.g., by considering further time savings through the use of generative artificial intelligence to develop content such as a motivating narrative or quizzes (Brehmer and Buonassisi 2024).

## References


Brehmer, M. & Buonassisi, V. (2024), Educators' Friend - Applying Generative AI to Create Effective Digital Learning Objects for Information Security Education: Toward Initial Design Principles. In: Proceedings of the 57th Annual Hawaii International Conference on System Sciences, 03.-06.01.2024.

Brehmer, M. & Reinelt, R. (2023), Gamifying a Learning Management System: Narrative and Team Leaderboard in the Context of Effective Information Security Education. In: Proceedings of the 56th Annual Hawaii International Conference on System Sciences, 03.-06.01.2023.

Chaudhary, S., Gkioulos, V. & Katsikas, S. (2022), Developing metrics to assess the effectiveness of cybersecurity awareness program, Journal of Cybersecurity **8** (1). https://doi.org/10.1093/cybsec/tyac006.

Deterding, S., Dixon, D., Khaled, R. & Nacke, L. (2011), From game design elements to gamefulness. In: Artur Lugmayr/Heljä Franssila/Christian Safran et al. (Eds.). Proceedings of the 15th International Academic MindTrek Conference Envisioning Future Media Environments, MindTrek '11: Academic MindTrek 2011, Tampere Finland, 28 09 2011 30 09 2011. New York, NY, ACM, 9–15.

Fertig, T., Schütz, A. E. & Weber, K. (2020), Current Issues Of Metrics For Information Security Awareness. In: Proceedings of the 28th European Conference on Information Systems (ECIS), 15.-17.06.2020.

Firsty Arisya, K., Ruldeviyani, Y., Prakoso, R. & Lailatul Fadhilah, A. (2020), Measurement of Information Security Awareness Level: A Case Study of Mobile Banking (M-Banking) Users. In: 2020 Fifth International Conference on Informatics and Computing (ICIC). IEEE.

Harms, J., Biegler, S., Wimmer, C., Kappel, K. & Grechenig, T. (2015), Gamification of Online Surveys: Design Process, Case Study, and Evaluation. In: Simone Barbosa/Mirko Fetter/Julio Abascal González et al. (Eds.). Proceedings of the 15th IFIP TC 13 International Conference, Bamberg, 14.09.-18.09.2015. Cham, Springer, 219–236.

Hiscox Ltd (2023), Hiscox Cyber Readiness Report 2023. https://www.hiscox-group.com/sites/group/files/documents/2023-10/Hiscox-Cyber-Readiness-Report-2023.pdf. Accessed 26.06.2024.

Hu, S., Hsu, C. & Zhou, Z. (2022), Security Education, Training, and Awareness Programs: Literature Review, Journal of Computer Information Systems **62** (4), 752–764. https://doi.org/10.1080/08874417.2021.1913671.

IBM (2023), Cost of a Data Breach Report 2023, Ponemon Institute & IBM Security. https://www.ibm.com/reports/data-breach. Accessed 26.06.2024.

KnowB4 (2023), KnowBe4 Technical Documentation for the Security Awareness Proficiency Assessment (SAPA). https://www.knowbe4.com/hubfs/KnowBe4-Security-Awareness-Proficiency-Assessment-SAPA-Technical-Document_EN-US.pdf. Accessed 26.06.2024.

KnowB4 (2024), Risk Scores. https://support.knowbe4.com/hc/en-us/articles/360001358728-Virtual-Risk-Officer-VRO-and-Risk-Score-Guide.

Kruger, H. A. & Kearney, W. D. (2006), A prototype for assessing information security awareness, Computers & Security **25** (4), 289–296. https://doi.org/10.1016/j.cose.2006.02.008.

McCormac, A., Zwaans, T., Parsons, K., Calic, D., Butavicius, M. & Pattinson, M. (2017), Individual differences and Information Security Awareness, Computers in Human Behavior **69**, 151–156. https://doi.org/10.1016/j.chb.2016.11.065.

Parsons, K., Calic, D., Pattinson, M., Butavicius, M., McCormac, A. & Zwaans, T. (2017), The Human Aspects of Information Security Questionnaire (HAIS-Q): Two further validation studies, Computers & Security **66**, 40–51. https://doi.org/10.1016/j.cose.2017.01.004.

Peffers, K., Tuunanen, T., Rothenberger, M. A. & Chatterjee, S. (2007), A Design Science Research Methodology for Information Systems Research, Journal of Management Information Systems **24** (3), 45–77. https://doi.org/10.2753/MIS0742-1222240302.

Rohan, R., Pal, D., Hautamäki, J., Funilkul, S., Chutimaskul, W. & Thapliyal, H. (2023), A systematic literature review of cybersecurity scales assessing information security awareness, Heliyon **9** (3), e14234. https://doi.org/10.1016/j.heliyon.2023.e14234.

Spitzner, L. (2024), Security Awareness Metrics. https://www.sans.org/blog/security-awareness-metrics/. Accessed 26.06.2024.

Taber, K. S. (2018), The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education, Research in Science Education **48** (6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2.

Wintergerst, R. (2023), Wirtschaftsschutz 2023. https://www.bitkom.org/sites/main/files/2023-09/Bitkom-Charts-Wirtschaftsschutz-Cybercrime.pdf. Accessed 26.06.2024.