Vision: Developing a Broad Usable Security & Privacy Questionnaire

Franziska Herbert
franziska.herbert@rub.de
Ruhr University Bochum
Bochum, Germany

Florian M. Farke
florian.farke@rub.de
Ruhr University Bochum
Bochum, Germany

Marvin Kowalewski
marvin.kowalewski@rub.de
Ruhr University Bochum
Bochum, Germany

Markus Dürmuth
markus.duermuth@rub.de
Ruhr University Bochum
Bochum, Germany

ABSTRACT

We aim to develop a questionnaire that measures privacy and security knowledge, attitude, and behavior on a broad level with a wide range of topics like authentication, smart home, web tracking, operating systems, mobile devices, instant messaging, and social media. To find relevant topics, we incorporated similar questionnaires and expert advice given by researchers and public institutions. In addition, we organized a workshop with Ph.D. students and also conducted expert interviews. We generated 276 items based on the derived topics, which we are currently evaluating for comprehension. So far, 20 end-users evaluated most of our items, with almost all evaluated items being comprehensible. We conduct this pilot testing on our items to reduce the number of items. These items will be used for a first validation with a representative sample of the German population. Our final questionnaire will consist of around 30 items, which we will further validate and translate into other languages. We plan to conduct a long-term study with the final questionnaire to determine changes in people's security and privacy knowledge, attitude, and behavior over time.

CCS CONCEPTS

• **Security and privacy → Privacy protections**.

KEYWORDS

security measurement, privacy measurement, questionnaire

ACM Reference Format:
Franziska Herbert, Florian M. Farke, Marvin Kowalewski, and Markus Dürmuth. 2021. Vision: Developing a Broad Usable Security & Privacy Questionnaire. In *European Symposium on Usable Security 2021 (EuroUSEC '21), October 11–12, 2021, Karlsruhe, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3481357.3481526

1 INTRODUCTION

In a nationwide survey in 2020, the German Federal Office for Information Security found that 25 % of Germans have been victims of computer crime [10]. The participants named falling victim to online fraud, account hijacking, phishing, and being infected with malware as experienced crimes. To protect themselves against such threats, 57 % of the participants stated to use anti-virus software and 48 % indicated to use secure passwords. However, 10 % stated not to use any protective measures at all. The survey sheds some light on users security behavior and highlights the importance of understanding what users' do and experience to provide them with relevant information and even training. Getting insights into users' security and privacy behavior, knowledge, and attitude is one of the main goals of usable security and privacy research. Therefore, questionnaires covering different aspects of privacy and security have been developed [7, 8, 11, 13, 15, 19, 20]. Many of these questionnaires focus on specific target groups, like employees [19, 20]. Others have a more narrow focus on specific constructs like privacy concerns [5, 11, 15], security attitudes [8, 29], security knowledge [13, 25], or security behavior intention [7, 25]. Some questionnaires are intended to measure security on a broader level, like the HAIS-Q [19, 20], which measures security knowledge, attitude, and behavior but is tailored to a work context.

Our goal is to develop a comprehensive questionnaire, inspired by the HAIS-Q, that measures users' general privacy and security knowledge, attitude and behavior, and can be used as a screening instrument. We include both questions about privacy and security because both topics are intertwined, and we think end-users may not distinguish between the two but are rather interested in protecting their data and devices. We intend to cover the most relevant security and privacy topics for end-users, aiming only at aspects they may know and control. It shall be used as a general measurement, giving a broad overview of participants' knowledge, attitude, and behavior. We think such a measurement can be beneficial to gain insight into end-user security and privacy knowledge, attitude, and behavior on a context unrelated, broader level, asking for practices that are widely advised. This broad measurement can be used to screen for which participants should answer more specific questionnaires on, for example, special privacy attitudes. We develop this questionnaire as an instrument to measure the general security and privacy knowledge, attitude, and behavior of German end-users in a long-term study, to get insights into whether their security and privacy knowledge, attitude, and behavior change over time, e. g., due to media coverage of security incidences. Therefore, we conduct the questionnaire in German first but will translate it into other languages after validation. To identify as many relevant topics as possible, we analyzed numerous questionnaires [5, 7, 7, 11, 13, 15, 19, 20], examined privacy and security advice for end-user [10, 21, 22], and conducted a workshop with Ph. D. students from research areas like usable security, information security, network security, or privacy. Furthermore, we interviewed five security and privacy experts for different disciplines, such as information security, law, psychology, and social science. For the

EuroUSEC '21, October 11–12, 2021, Karlsruhe, Germany

ACM ISBN 978-1-4503-8423-0/21/10.
https://doi.org/10.1145/3481357.3481526

development process, we incorporate a combination of the two most popular test theories used for psychological test development, the Classical Test Theory and the Item Response Theory [3, 12, 17].

To summarize, we make the following key contribution:

(1) By developing a broader screening instrument, we aim to complement existing questionnaires on privacy and security with a measurement that gives a general impression of participants' overall knowledge, attitude, and behavior regarding both privacy and security.
(2) After the validation, our questionnaire will be used for a long-term study on end-users' privacy and security knowledge, attitude, and behavior. The study will provide insights into the current state of end-user security and privacy understanding and will help to detect trends and developments in end-user knowledge, attitude, and behavior concerning IT security and privacy.

2 BACKGROUND

The basis of our questionnaire development process is two test theories: the Classical Test Theory (CTT) and the Item Response Theory (IRT). We highlight the three steps of developing our questionnaire: conceptualization, creating an initial item pool, and testing/analyzing the items. Additionally, we overview questionnaires measuring similar constructs that we used as a guideline and source for topics and items.

2.1 Test Theories

The CTT is the theoretical background to constructing and interpreting many psychometric and diagnostic tests in the social sciences. It is called classic as it has been developed and used for more than 50 years. The CTT introduces the concepts of test score (observed score), true score, and error score [3, 12, 17]. The test score of a person consists of both a true score and an error score. A good measurement should consist of an error score as small as possible so that the measure is reliable. This theory assumes that the true score and error score are uncorrelated and that error scores on parallel tests are uncorrelated. Reliability (defined as the proportion of the true variation) is a core parameter of the CTT, which focuses on relating test scores to true scores rather than focusing on items[1]. However, item analysis with identifying the parameters item difficulty, item variance, and item discriminatory power is also part of the CTT [12, 17]. One main weakness of the CTT is that its parameters are sample-dependent.

The Item Response Theory (IRT), an extension of the CTT, can counterbalance the weaknesses of the CTT. The IRT focuses on how test performance is related to the abilities measured by the test items. Within the IRT, two layers are distinguished, manifest variables (the observable answers to the test items) and latent variables (the characteristic values of the underlying not observable abilities) [12, 17]. The IRT is based on the assumption of a probabilistic relationship between the observed test score and the underlying ability value. Therefore, the IRT might be described as "item-based," whereas the CTT might instead be described as "test-based" [12, 17].

[1]The questions or statements posed in a questionnaire or test are referred to as items.

As the IRT counterbalances some shortcomings of the CTT, we will try to use a combination of both to assess our items.

Both theories guide researchers on how to construct a reliable and valid measurement for various constructs. For our questionnaire development, we will primarily focus on validation measurements to map manifest and latent variables, but doing this is out of the scope of this vision paper. We will conduct this analysis with our first representative sample in the future, see Section 4.2. For both theories, a pool of items needs to be developed to perform all statistics and find the best fitting items.

2.2 Questionnaire Development

Essential steps in developing a questionnaire are a clear conceptualization of the target construct (including a literature review), careful item wording, a large overinclusive initial item pool, and testing the items on a sample representing the whole target population [6, 14, 17].

Conceptualization. The first step in developing a questionnaire is to define the scope of the measurement, including what it will measure and, equally important, what it will not measure. The target construct should always be embedded in a theoretical context. The construct should be developed in detail and precisely. The generality of the construct needs to be addressed as well. Therefore, a literature review needs to be done. Reviewing measures of related and similar constructs helps overcome limitations and problems of existing questionnaires and clarify the conceptualization. Additionally, the literature review helps determine if the new questionnaire is needed or if such a questionnaire already exists [6, 17].

Initial Item Pool. The next step after the conceptualization and identification of the scope is the item generation. Subsequent statistical analysis can determine items that do not fit the construct, but such analysis is powerless to detect aspects of the construct that are not included in the items. Therefore, the initial item pool should be overinclusive, including a wide variety of construct aspects and even aspects that are only tangential or unrelated to the construct [6]. For every major aspect of the construct, at least a few items should be generated. For the item generation, the basic principles of item wording should be taken into account. Items should be phrased straightforward, as short as possibles, concisely, positively, and understandably. In addition to mostly unknown technical terms, buzzwords or absolute statements (i. e., statements containing words like "always," "never," or "all") should be avoided. Researchers should also avoid using assessments, leading items, complex, and "double-barreled" items. All items should be comprehensible for the whole target population [6, 14, 17]. The response format needs to suit the item format to measure the construct adequately. Many questionnaires in social science but also in security and privacy research using rating scales due to convenience. Most of these rating scales range from 4 to 11 response options and are verbally labeled either at the endpoints or at every point.

For equidistant instruments, careful wording of response scale labels is essential. In our questionnaire, we will use extensively tested response scales (scale labels) recommended by Rohrmann [23, 24].

Testing and Item Analysis. Before a first representative assessment, a pilot testing of the questionnaire primarily focusing on the comprehensibility of the items should be conducted and often consists of a cognitive walkthrough. Afterward, all items have to be analyzed to select the most fitting items and measure the target construct. For this analysis, the questionnaire needs to be answered by people from the target population, ideally a representative sample.

During the item analysis, the item distribution, item difficulty, item variance, and the discriminatory power of every item need to be considered. Items that are too easy, too hard, with very low variance or low discriminatory power will be excluded since all items should differentiate between people with different construct values (e. g., high security knowledge vs. medium and low security knowledge). However, some items with low and high difficulty have to be taken into account if the questionnaire should also differentiate between people with extreme construct values (e. g., between two people with high security knowledge) [6, 17].

In addition to the item analysis, measures of reliability and validity are essential. For reliability, one of the most used measures is the internal consistency of both the subscales and the overall scale. For validity, factor analysis (EFA & CFA) can be used. The factor analysis will indicate if the theoretical foundation of the construct is present in the data and which items correspond to which aspects of the construct [6, 17]. We do not go into further detail here, as the statistical procedures are beyond the scope of this vision paper. The large item pool is reduced to the best fitting items in this step, and the first final questionnaire is formed. The final questionnaire needs to be evaluated further with another sample and compared to other constructs for further validation.

2.3 Privacy & Security Scales

We considered some of the most commonly used questionnaires from security and privacy research as guidelines and models to develop our new questionnaire. We outline the selection criteria in Section 3.

Internet Users' Information Privacy Concerns (IUIPC). One of the most commonly used scales to measure the privacy concerns of end-users is the Internet Users' Information Privacy Concerns (IUIPC) scale by Malhotra et al. [15]. The IUIPC scale consists of 10 items divided into three dimensions: *Control* (3 items), *Awareness* (3 items), and *Collection* (4 items). It is measured on a seven-point rating scale ranging from "strongly disagree" to "strongly agree." Items cover topics like the importance of knowing how personal data is used and concerns about online companies collecting too much personal information. Gross [11] recently conducted a factor analysis to analyze the reliability and validity of the IUIPC and found two items within the dimensions control and awareness not fitting the expected model and therefore advised to eliminate these items. He suggested a revised IUIPC scale with only eight items but still comprising of the three utilized dimensions control (2 items), awareness (2 items), and collection (4 items). His analysis highlights the importance of carefully evaluating the reliability and validity during and after the development of questionnaires.

Security Behavior Intentions Scale (SeBIS). The Security Behavior Intentions Scale by Egelman and Peer [7] was developed based on the most common computer security advice that experts offer for end-users and consists of four sub-scales: *Attitudes towards choosing passwords*, *device securement*, *staying up-to-date*, and *proactive awareness*. In total, 16 items were identified, which show good reliability using both exploratory and confirmatory factor analysis. The items consist, among other topics, of statements about using software updates, using backups, using the private browser mode, and checking one's financial accounts for fraud. Sawaya et al. [25] applied the proposed SeBIS questionnaire in a cross-cultural study to examine whether cultural differences influence end-users' security behavior. They found that participants from Asian countries tended to exhibit less secure behavior and identified factors that affect participants' security behavior. Their study revealed that, in particular, the end-users' self-confidence in their computer security knowledge influences the user's security behavior considerably more than their actual knowledge.

Human Aspects of Information Security Questionnaire (HAIS-Q). Parsons et al. [19, 20] developed the Human Aspects of Information Security (HAIS-Q) questionnaire to measure security *knowledge*, *attitude*, and *behavior* tailored for the work context. Information Security Awareness (ISA) plays an essential role in protecting an organization from several cyber threats.Through two follow-up validation studies based on the HAIS-Q questionnaire, Parsons et al. [19] evaluated the validity of this questionnaire as an effective instrument to measure ISA. Within the first study, the authors discovered that those participants who scored higher on the HAIS-Q had better performance in a phishing experiment, indicating that HAIS-Q can predict an aspect of information security behavior. In a larger follow-up study, the authors could further establish the construct validity of the instrument and conclude that the HAIS-Q is a robust measure of ISA. Items of the HAIS-Q cover aspects like sharing work passwords with colleagues, plugging USB sticks found in the streets into one's work computer, and posting things about work on social media.

Measure of End-User Security Attitudes (SA-6). To assess end-users' security attitudes, Faklaris et al. [8] developed a six-item scale. The authors wanted to contribute a lightweight method for quantifying and comparing end-users' attitudes toward using recommended security tools and practices and improving predictive modeling of who will adopt security behaviors. The items cover aspects like interest in articles about security threats and motivation for keeping one's online data and accounts safe.

Privacy Attitudes and Behavior. Buchanan et al. [5] developed measurements for privacy concern, measuring both attitude and behavior, based on the assumption that privacy concern is not a one-dimensional but rather a multi-facet construct (which could only be confirmed for the behavior part). Therefore, they developed different scales, one scale consisting of 16 items measuring *privacy concern* and a tow scale with six items each measuring privacy behavior named *general caution* and *technical protection of privacy*. General caution covers aspects like hiding one's bank card PIN when making purchases, while technical protection asks for behaviors like removing cookies. Privacy concern evaluates on a five-point rating scale how concerned people are about online identity theft. Within three studies, they developed and validated

their scales and concluded that they are both reliable and valid for use on the internet.

Internet Skills Scale (ISS). Deursen et al. [28] proposed an instrument that measures Internet skills on five scales, *operational skills*, *information navigation skills*, *social skills*, *creative skills*, and *mobile skills*. Operational skills cover, e. g., downloading files, while information navigation skill items cover aspects as retrieving a prior visited website. Social skill items consist of aspects like knowing what information should not be shared. Creative skills ask about designing a website, and mobile skill items cover the ability to install apps on mobile devices. Most items start with "I know…" and the scales consist of items that may be added for a longer version of the scale. In total, 35 items can be used.

3 METHOD

Our goal is to develop an instrument that measures end-users' security and privacy *knowledge*, *attitude*, and *behavior* on a broader level. First, we need to determine aspects of security and privacy end-users might know and generate a large item pool taking these relevant aspects into account. This focus on end-users is a constraint for our questionnaire, limiting it only to aspects known to most end-users. We did a literature review of security and privacy questionnaires and reviewed advice for end-users from researchers and public institutions to find relevant topics. In addition, we organized a workshop with experts and conducted expert interviews. After gathering the relevant aspects, we generated items for all of these aspects on the before mentioned three dimensions.

3.1 Conceptualization of the Construct

One of the most prevalent psychological models for explaining, predicting, and changing behavior is the Theory of Planned Behavior (TPB) [1, 2]. In this model, behavior intention and actual control are direct antecedents of behavior. Attitudes influence behavior intention, and background factors like knowledge can influence behavior intention and behavior. Other psychological theories of explaining and chaining behavior also include the factors knowledge and attitude [26, 30].

The knowledge, attitude, behavior framework (KAB), primarily used in environmental and health psychology, states that behavior change is affected by knowledge and attitude [26, 30]. Especially in medical psychology, education (i. e., improving knowledge) positively influenced changing behavior [16]. Knowledge, attitude, and behavior are also core concepts of many questionnaires measuring IT security and privacy [5, 7, 8, 15, 25, 28]. One validated questionnaire, the HAIS-Q, even uses the KAB model as its basis [20].

We think measuring the dimensions knowledge, attitude, and behavior of IT security and privacy is essential for developing training and information material for end-users. Measuring behavior is equally important to examine if the behavior matches the knowledge and if the behavior changed according to new information or training. We can only assess what end-users need when we have a basic understanding (in the best case representative for a population) of people's knowledge, attitude, and behavior. Therefore, we take the HAIS-Q and the KAB model as a template for our questionnaire.

The other existing privacy and security questionnaires mentioned in Section 2.3 focus on either a specific target group, like employees [19, 20], or have a more narrow focus on specific constructs like privacy concerns [5, 11, 15] and security attitudes [8]. Therefore, we see the need for a broader instrument to measure end-user security and privacy on a wide range of topics (e. g., authentication and messaging) and with the average end-user as the target group and not tailored to the work context.

3.1.1 Literature Review. For the literature review, we first searched for research papers taking security or privacy scales under consideration. We first looked at instruments we are familiar with, like the IUIPC, the SeBIS, and the HAIS-Q, and followed other questionnaires in their related work sections. Additionally, we searched for terms as "privacy scale/measurement" and "security scale/measurement" and considered papers like [18, 27, 31]. We also searched for papers advising users to stay private or secure while surfing the Internet [21, 22] to collect IT security and privacy aspects. We did also take advice from the German Federal Office for Information Security into account [9].

3.1.2 Expert Interviews. In combination with our literature analysis, we conducted an expert workshop with 12 security experts, either Ph.D. students or professors working on security or privacy-related fields but coming from different disciplines (e. g., mathematics, computer science, social science). We asked the experts to name and rate security and privacy threats and advice for end-users during the workshop. Additionally, we conducted five expert interviews with professors, post-docs, and Ph.D. students from fields like computer science (n=3), psychology (n=1), and sociology (n=1). All of these experts have a research focus on security or privacy. We asked these experts to name and rate ("name the top 3") privacy and security threats as well as their top three security and privacy advice for end-users.

The analysis of our literature review and the expert interviews revealed seven broad important aspects of end-user privacy and security so far. As we did not aim to code our content in-depth but to detect relevant aspects of end-user privacy and security our topics are rather broad. Table 1 shows examples of the derived topics and (translated) items.

3.2 Item Generation

We first assigned existing items from questionnaires as [7, 18, 19, 31] to our topic, translated all of them into German, and at least partially rephrased most of them to fit the dimensions of our scale. We also generated items on the identified threats and advice through the literature review, expert workshop, and expert interviews. To investigate which items are most comprehensible for end-users, we tried to integrate items with different examples, some inverted ones, and items with the same content phrased differently. We also included items that do not entirely fit our construct to form an overinclusive item pool for further analysis and reduction.

For the wording of our items, we followed the guidelines for item design as outlined in Section 2.2. Compared to other knowledge scales (e. g., [28]), we did not use the prefix "I know…" but generated items with correct and false statements to measure participants' knowledge.

Table 1: Evaluated Topics, Dimensions, and Example Items

Topic	Dimension	Item Examples
authentication	Attitude	*I can rely on the strengths of passwords generated by a password manager.*
messaging	Behavior	*Within the last four weeks, I used end-to-end encryption for my emails to protect my data.*
social media	Behavior	*Within the last four weeks, I have publicly shared location data (e. g., a screenshot of my running course), e. g., on social network sites.*
smart home	Knowledge	*Once I give Wi-Fi access to a smart home device, e. g., Amazon Alexa, it can be attacked over the Internet.*
operating system	Attitude	*Automated updates make it easier for me to obtain device security.*
mobile devices	Behavior	*Within the last four weeks, I had switched off GPS on my smartphone when I did not need it due to security reasons.*
web tracking	Knowledge	*You should click on "allow all" on cookie notices.*

For the attitude part, we used expressions like "…makes me feel secure". The behavior items were assessed retrospectively to measure actual behavior better. We suspect participants to remember better what they did within the last four weeks than if they did something ever. As careful item wording is the most crucial part of the item generation process, all our items are proposed in our native language, German. We will evaluate the questionnaire first in Germany before translating and testing the final questionnaire in other languages, starting with English.

So far, we have generated 276 items based on existing questionnaires, expert advice, and (potential) IT security threats. All our items consist of statements, with the behavior items focusing on a retrospective assessment of the last four weeks (e. g., "Within the last four weeks I used end-to-end encryption for my emails to protect my data"). As the other dimension, the behavior dimension measures only self-reports. We use a four-week retrospective interval for the behavior dimension to measure actual behavior despite the self-report limitation. We think that four weeks is a reasonable and appropriate interval to remember past behavior. The idea of asking for a specific interval is based on psychometric tests like the BDI-II [4], which has been validated with a two-week retrospective interval.

The statements shall be answered on equidistant five-point scales by Rohrmann [23, 24], using the agreement response scale for knowledge and attitude and the frequency response scale for the behavior items. Each topic includes items for all dimensions so that for the entire questionnaire, each dimension and each topic-specific score can be determined and compared between participants.

3.2.1 Item Comprehensibility – Current State. Before we conduct a survey to analyze the items, we used pilot testing to evaluate the comprehensibility of the items. This pilot testing is usually done via a cognitive walkthrough or thinking aloud techniques. We considered both of these techniques as not suitable for our large item set. Instead, we used surveys as a more economical and less time-consuming method.

We conducted two surveys; one addressed to experts, and one addressed to end-users. All items were presented grouped by topic, with a dichotomous response scale consisting of "This statement is comprehensible" and "This statement is not comprehensible. I would change it to:". Within the second response option, participants had the opportunity to propose a different wording of the statement. The expert survey additionally consisted of the question "Do you miss a certain aspect of the topic [topic name]? Please indicate which aspect you are missing.". Participants did not have to review all topics but could review one topic up to all topics.

We used convenience sampling for this pilot study, asking colleagues, friends, families, and students to answer the survey. We started this pilot study at the end of May, and it is still ongoing. So far, 20 end-users and 11 experts evaluated at least some of our items, and at least one person reviewed every item. We will provide some insights into our preliminary results in Section 4. Additionally, all item analysis steps will be performed with data from a representative sample and are outlined in Section 4.2.

4 ONGOING WORK

In this section, we give a brief preview of the results of our pilot test, focusing on item comprehensibility, and outline the steps that we have planned for this ongoing work.

4.1 Preliminary Results

As explained in Section 3.2.1, we used a survey as an economical and efficient way to estimate the comprehensibility of our items and to gather feedback from both experts and end-user. The experts found 75 items not comprehensible, while the end-user found only 11 items not comprehensible. Some of the items were marked as incomprehensible due to misspellings. For most of these items, participants proposed only slightly different wording, pointed out typing errors or advised that examples would be helpful. Surprisingly, some experts remarked that end-users might not understand terms as "URL" or "VPN", but none of the end-users rated these items as not comprehensible, showing that they are somewhat familiar with these terms. Of course, the pilot study is not designed to measure if the end-user correctly understands the terms. We will assess that in subsequent studies.

End-user rated the following two items as incomprehensible without proposing a different wording, "Location data (GPS) should not be posted publicly on social network sites as, e. g., Facebook or Twitter" and "To secure your data, you could, for example, disclose false data when setting up a new account (e. g., giving a fake name)". We will therefore revise these items especially carefully or even drop them.

The participating experts also pointed out typing errors and misspellings but advised using different terms (e. g., a different translation for "malware" or another word for "device security").

The experts posed some questions, e. g., what online data consists of or what commercial use means. None of the end-user raised these questions, and the corresponding items were rated as comprehensible by end-users. However, we will revise these items and provide explanations for some terms used. The latter will also help determine that every participant has at least the same basic understanding of the used terms. However, we will not explain terms like "VPN" or "URL", as the questionnaire should measure participants' understanding and knowledge about aspects like that.

We hope to gather more feedback from more end-user and experts through our pilot study. We will include all the helpful feedback of all participants and will revise our items accordingly.

4.2 Future Work

This vision paper describes the first steps on a long way to a reliable, valid, and suitable questionnaire to measure end-users' IT security and privacy knowledge, attitude, and behavior. Following the pilot testing for item comprehensibility as mentioned above, we will eliminate or revise items rated as incomprehensible and include participants' feedback. We will also need to reduce the item pool, as 276 items are too many to be answered in one survey. Our first representative survey should not take longer than 20 minutes since reading and answering items is a monotonous task. We will test the reduced item pool with a sample representative for the German population and the response scales described in Section 3.2. With the data from that study, we will perform item analysis and further reliability and validity checks based on test theories (CTT & IRT) as outlined in Section 2.2. After this first validation step, we will use other samples and measures to validate the questionnaire further. We will conduct a diary study to investigate the relationship between self-reported and actual security behavior and will report on the relation of our questionnaire with similar and completely different instruments. The final goal is to conduct a long-term study using our questionnaire to get insights into the current status and potential changes of end-user IT security and privacy knowledge, attitude, and behavior, first in Germany and subsequently in other countries. This long-term study aims to measure changes in the tested dimensions caused by more user training in these domains, occurring data breaches, media coverage about privacy threats, or even completely other factors.

ACKNOWLEDGMENTS

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2092 CASA – 390781972.

REFERENCES

[1] Icek Ajzen. 1991. The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes* 50, 2 (Dec. 1991), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

[2] I. Ajzen, D. Albarracin, and R. Hornik. 2007. *Predicting and changing behavior: A reasoned action approach.* Lawrence Erlbaum Associates, 3–21.

[3] Timo M. Bechger, Gunter Maris, Huub H. F. M. Verstralen, and Anton A. Béguin. 2003. Using Classical Test Theory in Combination with Item Response Theory. *Applied Psychological Measurement* 27, 5 (2003), 319–334.

[4] Aaron T. Beck, Robert A. Steer, and Gregory K. Brown. 1996. *Manual for the Beck Depression Inventory-II.* Psychological Corporation, San Antonio, TX.

[5] T. Buchanan, C. Paine, A. N. Joinson, and U. Reips. 2007. Development of measures of online privacy concern and protection for use on the Internet. *Journal of the American Society for Information Science and Technology* 58, 2 (2007), 157–165. arXiv:https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.20459 https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.20459

[6] L. Clark and D. Watson. 1995. Constructing validity: Basic issues in objective scale development. *Psychological Assessment* 7 (1995), 309–319.

[7] Serge Egelman and Eyal Peer. 2015. Scaling the Security Wall: Developing a Security Behavior Intentions Scale (SeBIS). In *ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 2873–2882. https://doi.org/10.1145/2702123.2702249

[8] Cori Faklaris, Laura A. Dabbish, and Jason I. Hong. 2019. A Self-Report Measure of End-User Security Attitudes (SA-6). In *Symposium on Usable Privacy and Security (SOUPS '19)*. USENIX, Santa Clara, CA, 61–77. https://www.usenix.org/conference/soups2019/presentation/faklaris

[9] Bundesamt für Sicherheit in der Informationstechnik. 2019. Surfen, aber sicher! Basisschutz leicht gemacht, 10 Tips für ungetrübtes Surf-Vergnügen.

[10] Bundesamt für Sicherheit in der Informationstechnik und Polizeiliche Kriminalprävention der Länder und des Bundes. 2020. Digitalbarometer:Bürgerbefragung zur Cyber-Sicherheit.

[11] Thomas Gross. 2021. Validity and Reliability of the Scale Internet Users' Information Privacy Concerns (IUIPC). *Proceedings on Privacy Enhancing Technologies* 2021 (2021), 235 – 258.

[12] R. Hambleton and R. Jones. 1993. An NCME Instructional Module on: Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. *Educational Measurement: Issues and Practice* 12 (10 1993), 38–47.

[13] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. "My Data Just Goes Everywhere:" User Mental Models of the Internet and Implications for Privacy and Security. In *Symposium On Usable Privacy and Security (SOUPS '15)*. USENIX, Ottawa, 39–52. https://www.usenix.org/conference/soups2015/proceedings/presentation/kang

[14] Naresh Malhotra. 2006. *Questionnaire Design and Scale Development.* The handbook of marketing research: Uses, misuses, and future advances, 176–202.

[15] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. 2004. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research* 15, 4 (2004), 336–355. http://www.jstor.org/stable/23015787

[16] N. Houston Miller. 1997. Compliance with treatment regimens in chronic asymptomatic diseases. *The American Journal of Medicine* 102, 2, Supplement 1 (1997), 43–49. https://www.sciencedirect.com/science/article/pii/S0002934397004671

[17] Helfried Moosbrugger and Augustin Kelava. 2012. *Testtheorie und Fragebogenkonstruktion 2. Auflage.* Springer.

[18] Y. Park. 2015. Do men and women differ in privacy? Gendered privacy and (in)equality in the Internet. *Computers in Human Behavior* 50 (2015), 252–258.

[19] K. Parsons, D. Calic, M. Pattinson, M. Butavicius, A. McCormac, and Tara Zwaans. 2017. The Human Aspects of Information Security Questionnaire (HAIS-Q): Two further validation studies. *Comput. Secur.* 66 (2017), 40–51.

[20] Kathryn Parsons, Agata McCormac, Marcus Butavicius, Malcolm Pattinson, and Cate Jerram. 2013. The Development of the Human Aspects of Information Security Questionnaire (HAIS-Q). In *Australasian Conference on Information Systems (ACIS '13)*. RMIT University, Melbourne,Australia. https://aisel.aisnet.org/acis2013/31

[21] Elissa M. Redmiles, Amelia R. Malone, and Michelle L. Mazurek. 2016. I Think They're Trying to Tell Me Something: Advice Sources and Selection for Digital Security. In *IEEE Symposium on Security and Privacy (SP '16)*. 272–288.

[22] Robert W. Reeder, Iulia Ion, and Sunny Consolvo. 2017. 152 Simple Steps to Stay Safe Online: Security Advice for Non-Tech-Savvy Users. *IEEE Security& Privacy* 15, 5 (2017), 55–64.

[23] Bernd Rohrmann. 1978. Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie* 9 (1978), 222–245.

[24] Bernd Rohrmann. 2007. *Verbal Qualifiers for Rating Scales: Sociolinguistic Considerations and Psychometric Data.* Technical Report. University of Melbourne, Melbourne, Australia.

[25] Yukiko Sawaya, Mahmood Sharif, Nicolas Christin, Ayumu Kubota, Akihiro Nakarai, and Akira Yamada. 2017. Self-Confidence Trumps Knowledge: A Cross-Cultural Study of Security Behavior. In *ACM Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. ACM, New York, NY, USA, 2202–2214.

[26] Barbara Schneider and Nicole Cheslock. 2003. *Measuring Results, Gaining insight on behavior change strategies and evaluation methods from environmental education, museum, health, and social marketing programs.* Technical Report. Coevolution Institute Understanding Metrics Project, San Francisco, USA.

[27] Frederic Stutzman. 2006. An Evaluation of Identity-Sharing Behavior in Social Network Communities. *iDMAa Journal* 3, 1 (2006).

[28] Alexander J.A.M. van Deursen, Ellen J. Helsper, and Rebecca Eynon. 2015. Development and validation of the Internet Skills Scale (ISS). *Information, Communication & Society* 19, 6 (2015), 804–823.

[29] T. Velki, K. Solic, and H. Ocevcic. 2014. Development of Users' Information Security Awareness Questionnaire (UISAQ) — Ongoing work. In *International*

Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO '14). 1417–1421.

[30] W. Xu, G. Sun, Z. Lin, M. Chen, B. Yang, H. Chen, and K. Cao. 2010. Knowledge, Attitude, and Behavior in Patients with Atrial Fibrillation Undergoing Radiofrequency Catheter Ablation. *Journal of interventional cardiac electrophysiology : an international journal of arrhythmias and pacing* 28 (09 2010), 199–207.

[31] Alyson L. Young and Anabel Quan-Haase. 2009. Information Revelation and Internet Privacy Concerns on Social Network Sites: A Case Study of Facebook. In *International Conference on Communities and Technologies* (University Park, PA, USA) *(C&T '09)*. ACM, New York, NY, USA, 265–274. https://doi.org/10.1145/1556460.1556499