Home Is Where the Smart Is: Development and Validation of the Cybersecurity Self-Efficacy in Smart Homes (CySESH) Scale

Nele Borgert
nele.borgert@rub.de
Ruhr University Bochum
Bochum, Germany

Oliver D. Reithmaier
oliver.reithmaier@itsec.uni-hannover.de
Leibniz University Hannover
Hannover, Germany

Luisa Jansen
luisa.jansen@rub.de
Ruhr University Bochum
Bochum, Germany

Larina Hillemann
larina.hillemann@rub.de
Ruhr University Bochum
Bochum, Germany

Ian Hussey
ian.hussey@rub.de
Ruhr University Bochum
Bochum, Germany

Malte Elson
malte.elson@rub.de
Ruhr University Bochum
Bochum, Germany

ABSTRACT

The ubiquity of devices connected to the internet raises concerns about the security and privacy of smart homes. The effectiveness of interventions to support secure user behaviors is limited by a lack of validated instruments to measure relevant psychological constructs, such as self-efficacy – the belief that one is able to perform certain behaviors. We developed and validated the Cybersecurity Self-Efficacy in Smart Homes (CySESH) scale, a 12-item unidimensional measure of domain-specific self-efficacy beliefs, across five studies ($N = 1247$). Three pilot studies generated and refined an item pool. We report evidence from one initial and one major, preregistered validation study for (1) excellent reliability ($\alpha = 0.90$), (2) convergent validity with self-efficacy in information security ($r_{SEIS} = 0.64, p < .001$), and (3) discriminant validity with outcome expectations ($r_{OE} = 0.26, p < .001$), self-esteem ($r_{RSE} = 0.17, p < .001$), and optimism ($r_{LOT-R} = 0.18, p < .001$). We discuss CySESH's potential to advance future HCI research on cybersecurity, practitioner user assessments, and implications for consumer protection policy.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; • **Security and privacy** → *Human and societal aspects of security and privacy.*

KEYWORDS

self-efficacy, cybersecurity, smart homes, scale development, validation

ACM Reference Format:
Nele Borgert, Oliver D. Reithmaier, Luisa Jansen, Larina Hillemann, Ian Hussey, and Malte Elson. 2023. Home Is Where the Smart Is: Development and Validation of the Cybersecurity Self-Efficacy in Smart Homes (CySESH) Scale. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3580860

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3580860

1 INTRODUCTION

The popularity of smart home devices has increased so dramatically over the past decade that consumer households are now bristling with smart fridges, app-based video doorbells, smart heating systems, smart ambient lighting, all of which often interconnected through voice-activated assistants [72]. Despite their wide range of functions, one can define smart homes as residences "equipped with a high-tech network, linking sensors and domestic devices, appliances, and features that can be remotely monitored, accessed or controlled, and provide services that respond to the needs of its inhabitants" [7]. The rapid adoption of smart home devices is accompanied by increasingly voiced privacy concerns [33, 55, 132], and high profile security incidents [52, 88, 124]. Studies found that even if users express concerns and low disclosure attitudes, their security behaviors do not always correspond respectively, a phenomenon known as privacy paradox [68, 89]. This is exacerbated by privacy compromising default settings and dark patterns [82] that capitalize on the instinctive trust of consumers [133] or that of other people involved in the use of such devices [128].

Facilitating self-efficacy, which is the belief about one's own ability to enact certain skills [11], is a psychological solution to improve security behaviors that is promoted by decades of extensive research in HCI [28, 41, 100]. Self-efficacy is formed by motivational, cognitive, emotional, and choice-related mechanisms [11, 13], allowing for multiple pathways to strengthen users' self-efficacy beliefs, e.g., via interface designs that are scaleable across the consumer population [131]. It is a concept that is per definition subjective (i.e. latent) without an obvious behavioral counterpart. As such, self-efficacy cannot be measured directly like one would measure manifest behavior, e.g. setting a password, keystrokes, etc., simply by the fact that the self-appraisal of one's ability in those behaviors is not equivalent to the actual performance. Rather, as proposed by Social Cognitive Theory (SCT) [10, 14, 15], self-efficacy has reciprocal effects on behaviors and socio-structural factors (see Figure 1). Accordingly, self-efficacy significantly impacts a person's interests, behavioral choices, endurance when faced with obstacles (such as high user burden [114]), and ultimately selected or constructed

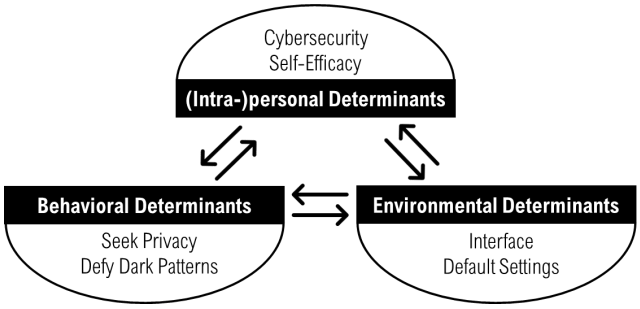

Figure 1: Causal Model of Social Cognitive Theory for Cybersecurity Self-Efficacy [14]

environments [11, 13]. Notably, self-efficacy is not a generalized characteristic, but a context specific belief [11].

While there is arguably some similarity between certain individual smart home devices and, for example, smartphones, there are other important differences that make smart home environments unique: (a) smart homes are complex, remotely networked ecosystems of multicomponent IT devices [116]; (b) they involve simultaneous use, multiple co-existing user roles, and are frequently not limited to one consumer with different rights and needs associated [59, 128]; (c) the high level of automation in smart homes might lead to perceptions of devices' agency [64, 95] with various important effects and (para)social mechanisms [99], e.g., over(-calibrated) trust in security default settings; and (d) smart homes collect diverse types of sensitive data in remarkable abundance with dramatic consequences for users in case of a breach [4, 27]. The combination of all these aspects distinguishes smart homes as a unique domain of technology interaction, and it is conceivable that lay users treat smart home devices differently than other technology. Therefore, we specifically define cybersecurity self-efficacy in smart homes as the belief in one's capability to control information processed by smart home devices and systems against unauthorized disclosure, modification, loss, or destruction.

Considering the impact of cybersecurity self-efficacy on security behaviors in various other IT domains (medical information systems [110], organisational data [48], software design [5], personal data [37]), it is presumed to also affect security behaviors of smart home owners [86]. Interventions designed to increase users' cybersecurity self-efficacy in smart homes could be arranged to feasibly reach the entire heterogeneous user base. The question of the success of self-efficacy interventions demands a well validated instrument to assess differences in smart home owners' cybersecurity self-efficacy beliefs. Existing assessment methods routinely rely on ad-hoc scales about some abstract or generalized experience of self-efficacy not specific to a particular class of devices, task, or skill [18]. This could cause problems on two levels: (a) self-efficacy beliefs are domain-specific, i.e., a person may report different self-efficacy strengths for more abstract and more specific contexts [14]; additionally (b) ad-hoc scales could lead to low replicability due to lack of validity information, increase the heterogeneity of findings in the field, show limited generalizability, and add to potential jingle-jangle fallacies [47]. The jingle fallacy is the false assumption that two similarly named scales measure the same trait, while the jangle fallacy is the false assumption that differently named scales actually measure dissimilar traits. Both fallacies can obfuscate the valid, empirical relationships between factors of interest [76, 126]. Having a validated psychodiagnostic scale of cybersecurity self-efficacy beliefs would enable meaningful insights and foster consensus on human factors in cybersecurity.

Here, we therefore report the development and validation of the Cybersecurity Self-Efficacy in Smart Homes (CySESH) scale across five studies. In three pilot studies ($Ns = 5, 23,$ and 82), we generated an item pool and established its content validity and comprehensibility. In the two main studies ($Ns = 166$ and 971), we examined the scale's psychometrics, including reliability and validity characteristics. Reliability results showed excellent coefficients, and we report robust evidence of convergent as well as discriminant validity. The final scale consists of 12 items that can be adopted for user or field studies. CySESH could become an important tool to assess the effectiveness of interventions targeted to improve self-efficacy, implement usable interfaces to assist cybersecurity tasks, and support evidence-based decision making regarding policy measures to sustainably improve security and privacy in the home environment at scale.

2 RELATED WORK

2.1 Latent Influences on Cybersecurity Behavior

The fundamental premise of improving security behaviors is that cybersecurity is both a technical and a psychological challenge [6, 91], as even the most sophisticated technical measures fail to protect data privacy and security when not used as intended by the designer. Expertise [44, 122] and awareness [65, 103] trainings have proliferated, but the heterogeneity of users, their prior knowledge, and abilities [49, 105], pose an insurmountable challenge to these interventions when implemented on a larger-scale.

Self-efficacy, the believed ability to perform certain behaviors, influences security behavior through the activation of multiple other latent processes (e.g., goal setting, analytic strategies, causal attribution, affect regulation) [11]. These open the spectrum of intervention possibilities to more feasible solutions reaching the heterogeneous user base [104, 131]. A meta-review on human aspects of cybersecurity found users' self-efficacy as the sole construct that consistently predicted security behaviors [41].

2.2 Self-Efficacy and Its Measurement

The goal of this paper is to present the development of a measure of self-efficacy beliefs specific to the domain of cybersecurity in smart homes. This domain-specificity is necessary in self-efficacy scales since a person's reports about their self-efficacy beliefs can vary to a great extent between, for example, the belief to be able to play chess or to skateboard [11]. Self-efficacy scales so far have been developed for a multitude of domains, for example finance [78], breastfeeding [120], alcohol resistance [130], online learning [115], entrepreneurship [83], driving [50], and many more. The underlying domain-specificity necessitates the construction of a fitting instrument if one wants to measure self-efficacy in said domain with high validity [121]. To successfully develop a new self-efficacy

measure, three critical recommendations regarding the exact wording of self-efficacy statements need to be realized [12, 14]: items must concern (a) the possibility of a behavior, not the intention, (b) the person's confidence in performing the behavior regardless of the expected outcomes, and (c) present skill perceptions, not hypothetical assumptions about those acquired in the future. Furthermore, best practices [87] advise to assure comprehensibility (avoid strong generalizations, multiple meanings, and negations) and aim for a clear test score interpretation (unidimensional items, intermediate item difficulty, and fitting response formats). Results from a systematic literature review [19] showed that published scales do not consistently comply with these recommendations [3, 58, 63]. This stresses two interconnected needs that we aim to meet: (a) a succinct, low-cost, transparent, publicly available scale for cybersecurity self-efficacy, and (b) comprehensive evidence regarding three psychometric quality criteria of that scale as defined by state-of-the-art test evaluation entities [38, 87, 98]. These three criteria include (1) objectivity, (2) reliability, and (3) validity.

2.3 Scale Development Criteria

First, objectivity is defined by the standardization of the test's procedure, analysis, and interpretation. A scale lacking objectivity would yield different conclusions due to unsystematic effects [1], e.g. different implementations across research labs or situations. For this reason, we provide a standardized scale with clear instructions, a computational strategy, and an interpretation note to allow global scale implementations. Second, reliability concerns the precision of measurement. An unreliable test could be due to an inconsistent performance of the user across items [21], and raise the question whether all items actually reflect the same self-efficacy belief that the scale is supposed to capture as a general factor [117]. Multiple items - or the replication of measurement instances - are necessary to estimate the degree of reliability [21]. In consequence, the CySESH scale contains multiple items that are generated and selected through iterative inputs of varied experts to arrange its precision. Finally, objectivity and reliability are prerequisites for the third quality criterion: validity of the obtained scores as a meaningful measure of the latent self-efficacy construct [36]. Scores would not be valid if lower self-efficacy beliefs did not correspond with lower scores on the scale. This would effectively imply that study results and any conclusions drawn from them are also not valid [46]. We validate our scale by estimating the relation between self-efficacy and other constructs as put forward by self-efficacy theory. If CySESH has no tendency to be distinct from those concepts, it is a conflated, redundant [62], or even invalid measure.

The steps we took to develop and validate the CySESH scale were based on the procedure described by Boateng et al. [16]:

(1) **Item Development**: We defined contexts of cybersecurity self-efficacy measures (i.e. device usage, coping or compliance, and security implementation) through emerging themes from a literature review. An expert workshop ($N = 5$) on smart home security was conducted to establish initial content validity, after which, we designed the first set of items. We did so with possible theoretical challenges for construct validity in mind; for example, we defined CySESH and its clear distinctions to similar psychological constructs (as self-efficacy scales tend to be confounded by e.g., outcome expectation [80] if they are not carefully designed). For this, we adhered to the assumptions of SCT and self-efficacy item generation guidelines by Bandura [12, 14].
(2) **Scale Development**: We piloted the items with two different samples: experts on test development ($N = 23$) were instructed to further review content validity, and native English speakers ($N = 82$) ensured the comprehensibility of items and instructions. We evaluated the original set of items with regard to qualitative responses and psychometric calculations, selected the best performing items, and preregistered the main validation study with complete methodology as well as analyses plans on the Open Science Framework.
(3) **Scale Evaluation**: We conducted an initial validation study ($N = 166$) and a main validation study ($N = 971$) to examine reliability and construct validity characteristics. The final scale contains of 12 items that unidimensionally measure user's cybersecurity self-efficacy in smart homes.

Investing in this scale development and validation process will allow further advances in understanding the complexity of cybersecurity issues by facilitating standardized research methods, an important prerequisite to effective meta-analytic research synthesis [40] that can provide vital policy implications for smart home consumer protection.

3 ITEM DEVELOPMENT

We employed a deductive strategic approach (in contrast to inductive, external, prototype etc.) for diagnostic item-development [1, 24]. A deductive method builds upon the existence of a specific theory that describes the trait's characteristics and determines the item format. To implement this approach, we began by considering assumptions of the SCT on self-efficacy [10, 15] and the resultant guidance on domain-specificy [11] and item-development [12, 14]. Items were not formulated inductively on the basis of prior published scales.

3.1 Identification of the Domain

3.1.1 Literature Review. An overarching literature survey on cybersecurity self-efficacy was conducted to gain a general review data set on the topic (file link: preregistration of literature survey). Multiple simultaneous literature reviews originate from this data set each with a different scope, such as a psychometric quality assessment [19], discussion on implications for practitioners, or as in this case a contextual overview of self-efficacy measures. The review data set has two main contributions to this scale development work. First, it reinforces the posited need of low-cost, standardized, and validated research methods in the field of cybersecurity self-efficacy by indicating the current extent of measurement heterogeneity. Second, it serves to identify a candidate for convergent scale analysis by assessing the context fit to CySESH and general representativeness of previously published scales. Consequently, it was the goal of the literature review to analyze current heterogeneity or item overlap in measures and identify relevant scales for convergent analysis.

Methods for gathering and extracting data of the literature were preregistered on OSF. Items were collected by three trained coders.

Training was completed when an inter-rater agreement coefficient of $\iota > 0.6$ for main variables was reached. Each coder extracted data from 2/3 of the publications, so that all publications were coded twice to ensure high quality review data. Given the interdisciplinary nature of cybersecurity self-efficacy research, we systematically queried nine scientific databases (EBSCOhost, IEEE Xplore, ACM Digital Library, Science Direct, dimensions.ai, arXiv, Scopus, Web of Science, and Wiley Online Library). To decrease bias in our keyword selection, we implemented a quasi-automated text mining and keyword co-occurrence network method as introduced by Grames et al. [54] (R version: 4.0.3; litsearchr version: 1.0.0). As a result, we used the following keyword string:

> "self-efficacy" AND ("cybersecurity" OR "cyber security" OR "information security" OR "IT security" OR "information technology security" OR "IS security" OR "information system security" OR "wireless security" OR "home wireless security" OR "usable security" OR "computer security" OR "data protection" OR "data security" OR "personal data" OR "privacy" OR "security threat" OR "wireless network" OR "device security")

We identified 173 different self-efficacy measures across 174 studies published in the past decade (list available at: data of literature review). Measures relied on ad-hoc development and different conceptual understandings of self-efficacy stemming from multiple theories, such as reactive control [97] or as a general trait [112]. Hence, we did not use the listed scales of the literature review as sources for our item development. We coded scales for their item-based heterogeneity. Specifically, we collected all provided items and then subjectively coded the mutual content of each scale's items to categorize the scale's overall context. For that reason, heterogeneity of scales is determined by the content of the respective items. A thematic synthesis strategy, involving open coding with emergent categories, was applied to identify three contexts [17]: (a) computer or internet usage, (b) coping or compliance behavior, and (c) privacy or IT security implementation. Computer or internet usage scales (4 scales) focus rather on the primary tasks of usage [67, 127]. Coping or compliance behavior scales (56 scales) aim at emotional, cognitive, or behavioral processes used in stressful situations [23, 69]. Privacy or IT security implementation scales (65 scales) use indicators of an individual's cybersecurity performance or application [100, 129]. Consequently, we focus on scales from the last context that would concern cybersecurity self-efficacy in smart homes.

3.1.2 Convergent Construct Validity. To determine whether CySESH is sufficiently close to content-similar constructs and in consequence a meaningful measure of latent cybersecurity self-efficacy, we analyzed its convergent validity. For this purpose, we selected a suitable self-efficacy scale from the context of privacy or IT security implementation: the Self-Efficacy in Information Security (SEIS) scale by Rhee, Kim and Ryu [100]. The scale's representative and general character is highlighted by recurring references within other ad-hoc scale development works [39, 75, 94] and significant correlations (ranging between $r = .60$ and $r = .85$) with a number of alternative cybersecurity self-efficacy measures. Rhee, Kim and Ryu [100] report several steps taken to ensure validity in item construction. The SEIS items involve skills related to the protection of information or systems more generally and hence, overlap with our own construct definition. Therefore, we expected a moderate correlation between SEIS and CySESH suggesting evidence for convergent construct validity.

However, the SEIS lacks context specificity. This would raise the question to which degree the SEIS scale is reliable and valid for the domain of cybersecurity in smart homes. One reason for difference is that the SEIS scale, published in 2009, was not designed to include present-day security issues relevant to smart homes, such as linkage, cookies, and cloud services. Items of the SEIS scale focus on more abstract protection skills related to personal computers and browser usage [100]. Some SEIS items refer to behaviors not broadly applicable to smart home devices, or they may even tap into other psychological traits that should be distinguished from self-efficacy [12, 14], e.g. outcome expectation or hypothetical assumptions about future skill acquisitions. Based on results from an expert workshop on smart home security stages, CySESH provides new domain-appropriate and contemporary items that are intended to reflect the recommended construct specificity. Items were exclusively generated by deductions from SCT theory as well as the workshop results. We piloted the scale in two expert studies ($N = 105$) and evaluated the remaining items with another two studies ($N = 1,137$). Results indicate CySESH's psychodiagnostic quality in terms of reliability and validity.

3.2 Validity of Items

3.2.1 Content Validity. Given the importance of content validity, its several sources (representativeness [35], relevance [118], domain clarity [96], and technical quality [57]) were consulted in three consecutive pilot studies to balance the approaches. We included a mix of content experts (workshop study), test development experts (first pre-testing study), and native English speakers (second pre-testing study) to generate items that provide good domain coverage with adequate psychometric performance and comprehensibility.

In the workshop study, security experts discussed potential indicators of CySESH's operationalization with the goal to achieve content representativeness and relevance [35, 118]. That is, the expert workshop was used to design the scale, not the literature review. This included six steps following Bandura's guide for scale construction [12]: (1) insights about multicausality, i.e. what dynamic factors influence the successful configuration of security in smart homes; (2) phases of pursuit; i.e. what chronological stages exist over which users can exercise control; (3) content validity of latent psychological constructs - behavioral, cognitive, affective - involved in cybersecurity; (4) definitions, response scales, and practice items; (5) smart home security tasks and activities, i.e. what abilities, skills or knowledge can be applied by users to succeed; and (6) challenges and impediments hindering regular security behavior. The five workshop participants were security experts in different areas of smart home cybersecurity: one software engineer with security expertise, two HCI designers with security expertise, one security expert, and one legal practitioner that is a privacy expert. We recruited experts through our affiliated partners of the funding consortium to identify crucial aspects of secure smart home technology use and adoption that are tied to self-efficacy. To achieve

psychometric validity, the scale is required to cover user actions actually relevant to the implementation and maintenance of smart home security – including important actions and behaviors that lay users may simply not be aware of. Further, lay people might have misconceptions about actions they frequently take but that do not actually improve security (e.g., scheduled password changes with complicated password rules). Lay people were involved in the pre-testing study 2 to make sure that the items that cover relevant behaviors identified by security experts are phrased comprehensibly.

The qualitative analyses of the unstructured workshop data, which were recorded through written notes, combined multiple synthesis schemes. Descriptive grouping contributed to analyzing the cybersecurity stages. Results revealed five stages in which users can apply abilities, knowledge, and skills to increase the cybersecurity of their smart home devices: (1) purchase, (2) commissioning, (3) daily use, (4) maintenance tasks, and (5) decommissioning. As a primary goal, participants generated a concept map to collect relevant security tasks and activities (file link: concept map of expert workshop). We used a thematic synthesis of the security tasks and activity data to examine the views of involved experts and identify common topics. The resulting 16 common topics were: data deletion, device connection, data collection, investments in devices, privacy policies, third party communication, setup assistant, settings, installation, password management, professional help, meaning of settings, understanding manufacturer declarations, updates, dark patterns, and threats. Other common themes of the discussion focused on: (a) performance demands, such as self-regulatory processes of users, (b) limitations of user control because their security would rely on technological solutions, and (c) limitations of circumstantial user control, i.e. people that did not choose to use a smart home device, such as children or guests.

Items were derived from the resulting five stages and 16 key topics of the discussion using the framework synthesis method, which is a structured technique that fits key topics to the cybersecurity stages. For each stage and each topic one or more items were developed. This analysis is only one of three aspects on which we based the generation of items (the other two being of theoretical nature: definition and assumptions of CySESH). Data of the framework analysis and resultant items are accessible on OSF (file link: data of expert workshop).

3.2.2 Discriminant Construct Validity. Self-efficacy items should be domain-specific and discriminant, i.e. allow distinguishing them from other, related psychological processes [62]. Discriminant validity is tested by assessing CySESH's distinction from constructs that should be closely related to self-efficacy but have established differences on a definitional dimension or by theoretical considerations [8, 79, 81]. This so-called nomological network of construct relations [22, 36, 62] puts forward trait-level differences from which insights can be best obtained.

Consequently, we expected small correlations between CySESH and selected discriminant constructs: self-esteem, optimism, and outcome expectations. Self-esteem is typically defined as a self-reflective overall evaluation of one's value or worth as a person [26, 102]. If a person believes to be capable of a certain behavior (i.e., high self-efficacy) and if that capability is valued, then the self-efficacy belief can positively impact one's self-esteem [79]. Optimism is the dispositional expectancy that positive outcomes will prevail throughout one's life [106, 108]. Even though optimism influences persistence behavior similarly to self-efficacy, there is no agency sensitivity to attain the positive outcome state [81]. Relations between both constructs are still expected since highly self-efficacious people tend to think more optimistically [14]. Outcome expectations are relatively defined as "believed consequences of a person's behavior" [42] and can be loosely categorized in physical, affective, and social outcomes that are either positive, negative, or neutral [42, 77]. These concern the results of performing a behavior while self-efficacy is the believed capability of performing it [8].

3.3 Test Construction

The complete materials of CySESH's pilot version are publicly accessible on the OSF (file link: CySESH scale pilot version). CySESH is a specific measure for cybersecurity self-efficacy in smart homes by means of two strategies. Items address smart home relevant security aspects (e.g., device linkage, third party access, cloud services) identified through the expert workshop. Next, four researchers, knowledgeable in item generation, test-theoretical aptitude of items, scientific standards required for questionnaires, and test theory in general, formulated 10 to 15 items each independently, which later in a process of discussion were combined to a total of 45 items (see Table 1). Since there is little validity evidence for other scales identified by the literature review, we based our first set of items not on prior published items, but on SCT theory and the expert workshop results. Participants are instructed that "[...] you will be asked about your smart home devices [...] always think about your complete smart home system".

Items used one of two response formats: 32 items are answered by indicating one's certainty to be able to perform the behavior of interest (e.g., "highly certain can do") and 13 items are answered by indicating one's agreement with the statement described by the item (e.g., "strongly agree"). The response scale ranged from 1 to 7 with lower values characterizing lower self-efficacy beliefs and higher values characterizing higher self-efficacy beliefs. This did not apply for one inverted item (item A42), which we phrased oppositely to control response biases [92]. For similar reasons, we also included two classic attention check items [111]. Two trial items unrelated to the topic of cybersecurity followed the instructions to let participants familiarize themselves with the response scales. All elements of the scale were implemented in an online questionnaire (via Qualtrics) to enhance standardization.

4 SCALE DEVELOPMENT

4.1 Pre-Testing the Scale

4.1.1 Procedure. In September 2021, we conducted two pilot studies to thoroughly test the developed item sets. The objective of the first study was to refine content validity (domain clarity and technical quality of the generated items [57, 96]) for which experts on test development - with at least a Bachelor's degree in psychology - were sampled. Expertise on theory-based, latent psychological constructs and their measurement is of high relevance to e.g., reassure the fit of generated items and the conceptual definition of CySESH, recognize missingness in structure, test instructions, or

trial items, and be aware of deviations in scoring or other flaws. Participation advertisements were distributed on social media and university-internal job platforms. The experts received *15 EUR* as compensation.

In the second study, native English speakers who preferably owned smart homes devices were recruited to ensure the comprehensibility of items and instructions as the final part of technical quality. This sample was acquired through Prolific and rewarded with the recommended payment rate of *7.50 GBP* per hour (i.e., *1.75 GBP* for participation in this study). The minimum required age for participation in both studies was 18 years.

Both survey flows began with an informed consent form, followed by sociodemographic questions, sample-specific instructions (focus on validity vs. comprehensibility), and the pilot version of CySESH. After each CySESH pilot item, experts were asked to provide remarks on the item in a text box.

4.1.2 Samples. A total of 23 psychologists (13 with a Bachelor's degree, 10 with a Master's degree) participated in the first study (17 female, 5 male, and one non-binary person with a mean age of $M = 25$ ($SD = 2.32$)). Participants also reported their English proficiency (lowest B2).

In the second study, we recruited 82 native English speakers (55 female, 26 male, one non-binary person). The sample's mean age was $M = 33.16$ ($SD = 10.47$). For the educational level, 34 participants stated that they had a Bachelor's degree, 33 finished the secondary school, 9 had a Master's degree, and 6 graduated from trade school. Three participants indicated they have or pursue a degree in psychology. Most participants (91%) owned at least one smart home device.

4.2 Evaluation and Item Reduction

First, we separately evaluated the CySESH pilot items from two perspectives - qualitative and quantitative analysis, respectively - and subsequently, consolidated the results to construct a validation version of CySESH. This process of pre-evaluation decisions is outlined in Table 1.

4.2.1 Qualitative Data Analysis. Items with negative remarks in the two pilot studies ($N = 105$) were either adjusted (7 items: A1, A5, A11, A12, A16, A28, A33), marked for exclusion (9 items: A3, A14, A15, A21, A23, A24, A32, A37, A39) pending further evidence of quantitative data, or in fact excluded (10 items: A17, A20, A22, A25, A27, A30, A34, A40, A41, A46) regardless of quantitative results. 19 items without negative remarks were eligible for continued inclusion (A2, A4, A6 - A10, A18, A19, A26, A29, A31, A35, A36, A38, A42 - A45).

Generally, participants commented on potential ambiguities, unknown technical terms, or perceived misfits of items and the provided CySESH definition, e.g. "maybe specify [...]" (participant #17) or "I think I would find it good if similar terms were always used [...]" (participant #12). Accordingly, we aligned terms ("devices" and "attack"), reworded items to reduce ambiguity ("safety" and the source of collection for A5), and simplified sentence structures (A11 and A5). Items with multiple remarks or remarks that required exhaustive changes were marked for exclusion, e.g. "the term resources may be unclear in this context" (participant #13). Strong ambiguity was perceived with regard to the terms "arrange" and "things" (A14) as well as "disadvantageous" and "trick me" (A23 and A24). This applies also to item A21 (setting up multiple devices), where the extent of the statements seemed unclear to participants. Comments on A3, an item about deletion of local data, were resolved by specifying the instructions of the scale on referred devices. The item was kept after confirming results of the quantitative decision stage to balance the content of A2, which regards the deletion of cloud data. Another item about the deletion of cloud data (A15) had remarks on the ambiguous meaning of the type of data and the definiteness of deletion. Instead of rewording the item, A15 was removed because A2 outperformed A15 and both statements cover the deletion of cloud data. Severe negative remarks, for example "fine, but motivation =/= CySESH. Don't know if the definition above is confusing?" (participant #08) or "what exactly are bad settings?" (participant #60), lead to the exclusion of the respective pilot item. In hindsight, we realized that A17, which has to do with open ports, used terms that might be too technical for lay users. A27 (concerning the protection of information) was formulated very vaguely and pointed rather towards successful consequences of the behavior. Severe concerns were also voiced for item A34, which is about the detection of bad settings. Participants remarked that responses could refer to different subjects; either themselves, the functionality of the device, or the manufacturer.

4.2.2 Quantitative Data Analysis. Due to the small sample size, the quantitative analyses were only performed on the data collected from the native English speakers in the second pilot study ($N = 82$). Results can be accessed on the OSF (file link: results from CySESH pilot studies). We calculated an initial factor estimate using Horn's parallel analysis. It demonstrated a one factor solution, which supported our goal of developing a unidimensional scale. An exploratory factor analysis with a promax rotation partly confirmed this finding with mostly satisfactory factor loadings, but not yet exclusively acceptable model fit indices ($TLI = 0.734; RMSEA = 0.070; RMSR = 0.09; BIC = -2831.60; \chi^2(945) = 1332.75, p < 0.001$).

Table 2 shows the three psychometric criteria that items had to satisfy to be considered for subsequent scale validations. Item selection based on these criteria leads to scales that achieve a centralized distribution of responses, best differentiate between high-scoring and low-scoring users, and include homogeneous items. Item analysis results for all criteria are shown in Table 1. We dropped 22 items (A16, A19, A20, A22, A25, A26, A29 - A33, A35 - A41, A43 - A46) because responses to those items indicated issues with ceiling effects (possibly due to low complexity of the security task). We removed another item (A42) that did not meet the criteria for minimum item-total correlation coefficients. This was also the only item in the pilot version which was inverted. No remaining items showed lower standard deviations than defined. Consequently, 22 items (A1 - A12, A14, A15, A17, A18, A21, A23, A24, A27, A28, A34) were left for further selection processes.

4.2.3 Finalizing Item Selection. Items were required to pass both analyses perspectives to be eligible for the scale validation stage. Of the 13 eligible items (A1, A2, A4 - A12, A18, A28), 2 were redundant (A18, A28) in content with qualitatively slightly better performing items (A4, A12) and therefore excluded for reasons of test economy.

Table 1: Selection Process of CySESH Pilot Version

#	Item	Remark	*M*	*SD*	*ITC*	Result
A1	I can use the device's privacy policy for risk assessment of my privacy.	AD	4.18	1.82	.65	IN
A2	I can delete the data stored in the cloud if I no longer want to use my device.	IN	4.79	2.03	.67	IN
A3	I can get the information I need to delete my data stored on my device.	MX	4.99	1.87	.72	IN: B
A4	I can find out which third parties have access to the data my device collects.	IN	3.48	1.87	.60	IN
A5	I can get the information I need on what data are collected.	AD	4.50	1.77	.72	IN
A6	I can learn the technological know-how to understand my device's technical data sheet.	IN	4.45	1.96	.56	IN
A7	I can keep track of my privacy implications when I link multiple devices.	IN	4.01	1.77	.70	IN
A8	I can detect when interfaces are designed to influence my decisions about security options.	IN	4.22	1.89	.65	IN
A9	I can identify violations of my privacy rights by a device feature.	IN	3.46	1.53	.63	IN
A10	I can get in touch with a manufacturer's data protection officer when necessary.	IN	3.68	1.87	.52	IN
A11	I feel confident that I know about the existing privacy implications before buying a new device.	AD	4.73	1.69	.66	IN
A12	I know how to safely disable my device in case of a security flaw.	AD	4.83	1.96	.67	IN
B13	–	–	–	–	–	IN: B
A14	I can do the things needed to arrange a secure smart home.	MX	4.63	1.71	.55	X
A15	I can get the information I need to delete my data stored in the cloud.	MX	4.50	1.93	.66	X
A16	I can delete the data on my device if I no longer want to use it.	AD	5.44	1.93		X
A17	I can manage open ports of the device.	X	3.59	1.92	.48	X
A18	I can get information on transfers of my data to third parties.	IN	3.29	1.72	.59	X: R
A19	I can figure out security default settings of my device.	IN	5.09	1.79		X
A20	I can follow the instructions of my device's setup assistant.	X	6.52	0.97		X
A21	I can set up multiple smart devices in a way that makes them safe to use.	MX	4.72	1.79	.63	X
A22	I can securely connect different devices with each other.	X	5.29	1.75		X
A23	I can detect when a user interface is designed to trick me into accepting disadvantageous privacy settings.	MX	4.32	1.92	.63	X
A24	I can notice when a user interface is designed to trick me into accepting disadvantageous privacy settings.	MX	4.22	1.90	.62	X
A25	I can motivate myself to keep my device's security software updated.	X	5.38	1.72		X
A26	I can update my device to the latest security standards.	IN	5.45	1.89		X
A27	I can protect the information on my device.	X	4.72	1.72	.74	X
A28	I can protect the data on my device against security attacks.	AD	3.91	1.89	.71	X: R
A29	I can reset my device if it gets compromised.	IN	5.78	1.66		X
A30	I can revert settings on my device if it gets compromised.	X	5.24	1.77		X
A31	I can use my device in a way that meets my own privacy expectations.	IN	5.09	1.69		X
A32	I can inform myself about the security of a device that I want to purchase.	MX	5.16	1.59		X
A33	I can find information about a device's security online.	AD	5.65	1.65		X
A34	I can detect bad settings in my device's security options.	X	4.15	1.85	.74	X
A35	I can set a secure password.	IN	6.24	1.39		X
A36	I feel confident that I understand the current privacy settings of my device.	IN	5.12	1.74		X
A37	I have the resources to read and understand privacy policies.	MX	5.87	1.53		X
A38	I know how to change default security options on my device.	IN	5.23	1.86		X
A39	I know how to navigate to the security settings menu of my device.	MX	5.71	1.54		X
A40	I know how to set up a secure password for my device.	X	6.30	1.28		X
A41	It is easy for me to get information on a device's security from trustworthy sources.	X	5.06	1.63		X
A42	It is hard for me to assess potential concerns regarding my device's security before set up.	IN	4.39	1.73	.06	X
A43	I am capable of disposing my device with none of my data left on it.	IN	5.48	1.86		X
A44	I am confident to find a person who installs my device securely.	IN	5.43	1.81		X
A45	I am confident to make the right decision when being informed about a threat to the security of my device.	IN	5.51	1.49		X
A46	I am confident that the security settings I choose will protect my device.	X	5.55	1.30		X

Note. The criteria for the exclusion of an item are shaded; item-total correlations were calculated only for items that passed the set mean and standard deviation inclusion criteria; IN = included in CySESH validation version; IN: B = included for balance; AD = adjusted; MX = marked for exclusion; X = excluded; X: R = excluded for redundancy.

Table 2: Psychometric Inclusion Criteria for Item Analyses

Analysis	Accepted Values	Interpretation
Mean item score	$3 - 5$	Moderate item difficulty
Standard deviation of item	> 1.4	Sufficient differentiation capability
Item-total correlation	> 0.3	Adequate homogeneity between item and scale

However, the remaining items only tangentially covered the forth stage of cybersecurity (maintenance tasks). To compensate this imbalance, we developed a new item about software updates (B13). Another imbalance was created by including A2 (deletion of cloud data) and the ineligibility of items on the deletion of local data. As a solution, we decided to include A3 and revise its qualitative issues by including more specific scale instructions. The final 13 items included in the validation version of CySESH are provided in Table 5.

5 SCALE EVALUATION

5.1 Methods

Two validation studies are described: (a) an initial validation study that focused on discriminant validity, and (b) a main validation study that examined reliability as well as additional validity characteristics.

5.1.1 Initial validation study. Data were collected in October 2021 via an online questionnaire designed in Qualtrics. An a priori power analysis (using the R package pwr) indicated a minimum sample size of 138 participants for an assumed correlation of $r = .30$, power of .95, and $\alpha = .05$. We recruited participants (at least 18 years old, native language English) from Prolific and paid them the recommended hourly rate of *7.5 GBP* (i.e., *1 GBP* for participation in this study).

The objective of the initial validation study was to estimate CySESH's discriminant validity from outcome expectation. The survey flow began with an informed consent form, followed by the validation version of CySESH (items B1 - B13 listed in Table 5) paired with outcome expectation items, and questions about participant demographics. We included two attention check items (see CySESH scale validation version and measures of initial validation study). In Table 3, we provide an overview of the measures used in the initial validation study.

The Outcome Expectation (OE) scale was adopted from Maddux, Norton and Stoltenberg [80] and consists of 2 generic items. Those two generic items are specified as regards content for each CySESH item (B1 - B13), making the scale a total of 26 OE items. This pairing of self-efficacy and outcome expectation is inevitable because CySESH items each describe a certain skill of which the effectiveness of doing so is being assessed by the OE items. Example pairings are for B1: "For those who can use devices' privacy policies for a risk assessment of their privacy, it is a very effective way to improve their IT security or privacy" (OE1-B1) and "If I were able to use devices' privacy policies for a risk assessment of my privacy, it would improve my IT security or privacy" (OE2-B1). The measure was primarily chosen due to the applicability of the generic items and its comprehensive discussion on the role of OE in SCT and self-efficacy theory.

A total of 229 people participated in the initial validation study. Participants were excluded from analysis if they stated to not own a smart home device or failed at least one attention check, which reduced the sample size to 166 participants (119 female, 45 male, and 2 non-binary). The mean age of participants was $M = 33$ ($SD = 11.54$) years. Participants had a relatively high level of education with 74 people reporting a Bachelor's degree, 46 having graduated from secondary school, 23 with a Master's degree, 19 trade school graduates, 2 PhD level participants, 1 primary school graduate, and 1 participant who preferred not to say. 84 participants stated being full-time and 23 part-time employees, 26 were students, and 33 declared other statuses (e.g., retirement).

5.1.2 Main validation study. The main validation study was preregistered on the OSF prior to data collection with detailed methodology records and an R script for statistical analyses plans (file link: CySESH preregistration). Data were collected in January 2022 using Qualtrics. Participants (at least 18 years old, native language English) were recruited from Prolific and received the recommended payment rate of *7.5 GBP* per hour as reward (i.e., *1.25 GBP* for participation in this study). We calculated an a priori power analysis in R (package: semPower) to plan the sample size needed to adequately power our registered structural equation model. Results indicated a minimal sample size of $N = 461$ with a power of .95, an $\alpha = .05$, and an expected model fit of $RMSEA = .04$ for a unidimensional model with 13 items. To account for dropouts, missingness, and additional exploratory analyses, we aimed to approximately double the required sample size.

Table 3: Measures Included in Initial Validation Study

	CySESH Scale	OE Scale
Construct	Self-efficacy	Outcome expectation
Item Count	13	26
Response Scale	7-point Likert	7-point Likert
Analysis Objective	-	Discriminant validity

The primary objectives of this study were to: (1) select items for a final CySESH version, (2) test the reliability of the CySESH items, (3) demonstrate its convergent validity to a closely related measure, and (4) verify its discriminant validity to theoretically related but distinct psychological traits. The survey began with an informed consent form, followed by four scales as well as three attention check items that were all presented in a randomized order, and ended with demographic questions (see Table 4). All items and scale instructions are accessible on the OSF (file links: CySESH scale validation version and measures of main validation study). First, cybersecurity self-efficacy in smart homes was measured with the validation version of CySESH (items B1 - B13 listed in Table 5). Second, for self-efficacy in information security, we used the SEIS scale [100]. It is a representative scale of its field and demonstrates acceptable scale quality criteria [100]. Third, self-esteem was measured with the Rosenberg Self-Esteem (RSE) scale [101]. We selected the RSE scale because of its long history and significant popularity in psychological self-esteem research [61], as well as its high quality [31, 34]. Lastly, we used the Life Orientation Test Revised (LOT-R) [107] to measure optimism. The LOT-R is a highly established measure of dispositional optimism [25, 109] that repeatedly demonstrated its psychometric quality [30, 51].

In the main study, a total of 1,068 survey responses were collected. We excluded 97 participants because they either indicated to not own a smart home device, failed at least one attention check, had missing responses on CySESH, or did not complete the demographic questions. The remaining 971 participants (613 female, 352 male, 4 non-binary, and 2 preferred not to say) had a mean age of $M = 37.9$ ($SD = 13$). 397 people had a Bachelor's degree, 357 graduated from secondary school, 125 had a Master's degree, 56 were trade school graduates, 22 had a PhD, 11 participants preferred not to say, and 3 were primary school graduates. For employment status, the sample included 532 full-time as well as 177 part-time employees, 83 students, and 179 people stated other statuses (e.g., retirement).

Table 4: Measures Included in Main Validation Study

	CySESH Scale	SEIS Scale	RSE Scale	LOT-R
Construct	Self-efficacy	Self-efficacy	Self-esteem	Optimism
Item Count	13	11	10	10
Response Scale	7-point Likert	7-point Likert	4-point Likert	5-point Likert
Reliability (original)	-	CR = .97 [100]	C_R = .92 [102]	α = .78 [107]
Analysis Objective	Item selection & reliability	Convergent validity	Discriminant validity	Discriminant validity

Note. CR: composite reliability; C_R: Guttman scale coefficient of reproducibility; α: Cronbach's alpha.

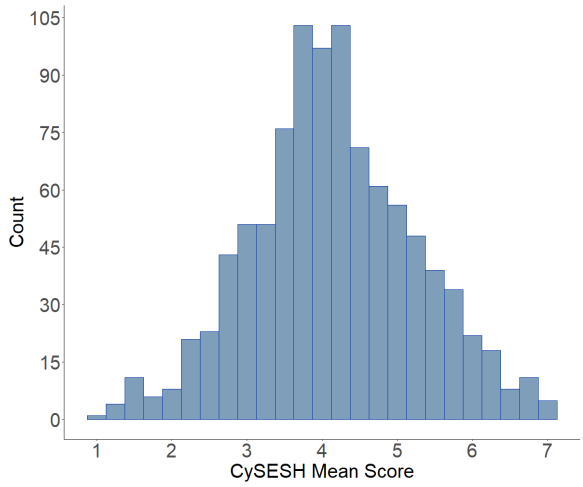

Figure 2: Distribution of CySESH Mean Scores

5.2 Results

The data and R scripts are publicly accessible on the OSF (file link: CySESH validation studies). Analysis results are primarily based on data from the main validation study (N = 971). Data from the initial validation study (N = 166) only find use in the discriminant validity analysis concerning outcome expectation.

5.2.1 Item analysis. We applied the same psychometric criteria (see Table 2) as in the pilot studies. The aim was to select items with moderate difficulty that sufficiently differentiate self-efficacy strengths and homogeneously align with the overall scale. Based on these criteria, one item (item number B13) was excluded, see Table 5. Its mean score was higher then the defined limit (M = 5.2 > $M_{\max}$ = 5). The exclusion of item B13 was calculated to affect reliability marginally; Cronbach's α dropped from 0.91 to 0.90. The other items (B1 - B12) performed well within the set thresholds (Table 5). Thus, the 12 final items of CySESH are: B1 - B12.

5.2.2 Scale analysis. The distribution of CySESH mean scores is shown in Figure 2. The overall mean scale score was M = 4.15 (SD = 1.10) with a skew of g_1 = 0.05 and a kurtosis of g_2 = −0.06. These values imply a symmetric distribution that is mesokurtic. No violation of normal distribution assumptions or biased tendencies of the sample to agreement or disagreement are indicated.

We conducted four analyses to test unidimensionality. Inter-item correlations indicated a moderate to strong relatedness, r ranging between .29 and .61 (OSF file link: scale analysis of main validation study). A confirmatory factor analysis (CFA) using a maximum likelihood (ML) estimator showed mixed results for a one factor solution (TLI = 0.938; CFI = 0.949; $RMSEA$ = 0.068; $\chi^2(54) = 295.84, p > 0.001$). We complemented this with another CFA model using a diagonally weighted least squares (WLSMV) estimator, which is specifically designed for ordinal data [74, 85]. This yielded better standard (TLI = 0.997; CFI = 0.998; $RMSEA$ = 0.020; $\chi^2(54) = 75.15, p < 0.05$) and robust fit indices (TLI = 0.944; CFI = 0.954; $RMSEA$ = 0.052; $\chi^2(54) = 194.61, p < 0.001$). Lastly, the path diagram presented in Figure 3 shows the standardized parameters of a unidimensional structural equation model. Loading coefficients of the 12 items reflect the assumed conceptually distant cybersecurity stages that were identified by experts in our first pilot study and still indicate strong relationships to a latent self-efficacy belief. A meta-analysis conducted by Peterson [93] found that for empirical factor loadings the average loading obtained is only λ = 0.32. The results from all four techniques support our validation approach.

For reliability analyses of CySESH items, we report two coefficients: Cronbach's α as the lower limit of reliability [53] and McDonald's ω as a better representation of the means [60, 119]. Both reliability estimates demonstrate excellent reliability; with α = 0.90 and ω = 0.90.

5.2.3 Convergent validity. To assess convergent construct validity of CySESH, we compared the correlation between CySESH and SEIS against a defined threshold. The said threshold represents a correlation limit that is aimed to be reached or surpassed and is determined by the attenuation correction formula [70]. As described by Kristof [70] the attenuation correction formula is given by:

$$\rho_{\text{limit},i} = \rho_{\text{ideal}} \cdot \sqrt{(rel_{\text{CySESH}} \cdot rel_i)} \tag{1}$$

For each comparison i, we compute the attenuation correction $\rho_{\text{limit},i}$ via Eq. 1 to obtain specific thresholds. Here, rel_{CySESH} and rel_i represent the reliability estimates of CySESH and SEIS respectively. For convergent validity, we set ρ_{ideal} = 0.8 because we assumed similarity but domain-specific differences between CySESH and SEIS beliefs. The reported ρ_{ideal} = 0.8 deviates from our preregistered ρ_{ideal} = 1 because a correlation of r = 1 would mean a perfect positive relatedness between CySESH and SEIS, which was not anticipated. A correlation of r = 0.8 still signifies a very large

Table 5: Selection Process of CySESH Validation Version

#	Item	*M*	*SD*	*ITC*	Result
B1	I can use devices' privacy policies for risk assessment of my privacy.	4.09	1.49	.65	IN
B2	I can delete the data stored in a cloud if I no longer want to use my devices.	4.59	1.69	.62	IN
B3	I can get the information I need to delete my data stored on my devices.	4.69	1.58	.67	IN
B4	I can find out which third parties have access to the data my devices collect.	3.58	1.61	.62	IN
B5	I can find out what data my devices collect.	4.31	1.56	.70	IN
B6	I can learn the technological know-how to understand my devices' technical data sheets.	4.31	1.73	.54	IN
B7	I can keep track of my privacy implications when I link multiple devices.	3.92	1.48	.73	IN
B8	I can detect when interfaces are designed to influence my decisions about security options.	3.80	1.60	.58	IN
B9	I can identify violations of my privacy rights by a device feature.	3.44	1.48	.67	IN
B10	I can get in touch with a manufacturer's data protection officer when necessary.	3.76	1.68	.48	IN
B11	I can find out about existing privacy implications before buying a new device.	4.52	1.54	.63	IN
B12	I can disable my devices in case of a security attack.	4.76	1.76	.53	IN
B13	I can find out about important security updates for my devices.	5.20	1.46		X

Note. The criteria for the exclusion of an item are shaded; IN = included in final CySESH scale; X = excluded.

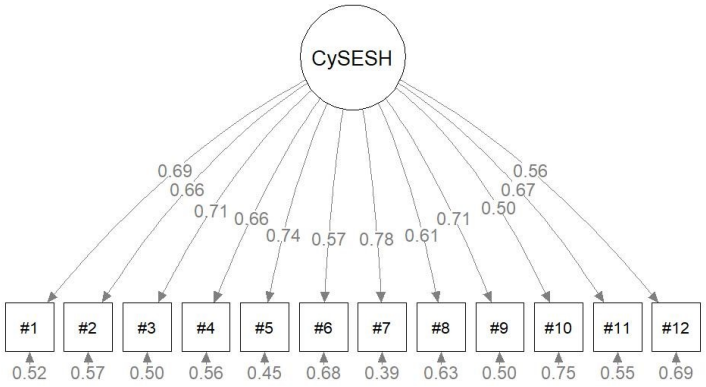

Figure 3: Path Diagram for Unidimensional SEM Model

effect size in human-related research [32, 45] and was therefore chosen as ρ_{ideal} for convergent validity.

The SEIS items had a reliability coefficient of $\alpha = 0.94$. This resulted in a convergent limit of $\rho_{\text{limit,SEIS}} = 0.73$. The correlation between CySESH and SEIS was $r = 0.64$, $p < .001$ and thus, fell below the defined threshold of convergent validity with a difference of $r_{\text{diff}} = -0.09$. However, the expected trend of the relationship between CySESH and SEIS was supported.

5.2.4 Discriminant validity. For each discriminant validity analysis, we also calculated the attenuation correction $\rho_{\text{limit},i}$ via Eq. 1. To reflect the theoretical differences and yet relatedness between CySESH, OE, RSE, and LOT-R, we set $\rho_{\text{ideal}} = 0.2$. In the preregistration, ρ_{ideal} for discriminant validity was set to 0.1. Taking into account the meta-scientific discussion that nonzero correlations are to be expected between any given variables [90], we heightened ρ_{ideal} to signify the theorized relationships between CySESH and its discriminant constructs. Discriminant validity between two scales is affirmed if their correlation coefficient is equal to or below the computed threshold.

Results of the three discriminant validity analyses are shown in Table 6. The comparison of the correlation coefficients with their specific discriminant limits substantiate the validity of CySESH. Overall, the assumed discrimination between CySESH and its discriminant constructs was sustained, but outcome expectation did not reach the defined threshold. The study demonstrated CySESH's excellent discriminant validity for self-esteem and optimism. Figure 4 shows the inter-construct correlations from both validation studies (left panel: main study; right panel: initial study) to illustrate the reported evidence.

5.2.5 Exploratory analysis. An exploratory multiple regression was estimated to determine whether the CySESH scores were systematically affected by demographic variables (i.e., age, gender, educational level, and employment status). The reference group of the regression analysis are female participants of average age, who are full-time employed and have a Bachelor's degree. Age and male gender significantly predicted CySESH scores; $\beta_{\text{age}} = -0.017, p < .01, 95\%$ CI $[-0.25, -0.09]$ and $\beta_{\text{male}} = 0.28, p < .01, 95\%$ CI $[0.15, 0.42]$. The other computed standardized regression coefficients were non-significant and ranged from $\beta = -0.01$ to $\beta = 1.12$ (OSF file link: exploratory analysis of main validation study). The demographic regression accounted together for $R^2 = .05, 95\%$ CI $[0.01, 0.06]$. Reasons for significant estimates could for instance be that our large sample size over-powers otherwise

Table 6: Results of Discriminant Validity Analysis

	OE Scale	RSE Scale	LOT-R
Reliability (this paper)	$\alpha = 0.80$	$\alpha = 0.93$	$\alpha = 0.90$
Discriminant Limit	$\rho_{\text{limit,OE}} = 0.17$	$\rho_{\text{limit,RSE}} = 0.18$	$\rho_{\text{limit,LOT-R}} = 0.18$
Correlation with CySESH Scale	$r = 0.26,\ p < .001$	$r = 0.17,\ p < .001$	$r = 0.18,\ p < .001$
Discriminant Difference	$r_{\text{diff}} = -0.09$	$r_{\text{diff}} = 0.01$	$r_{\text{diff}} = 0.00$

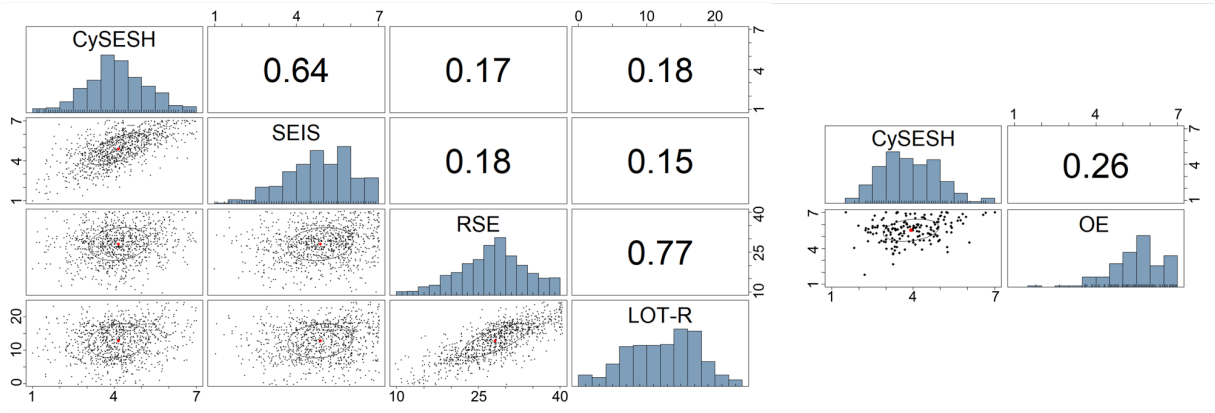

Figure 4: Scatter Plot Matrices of Validity Analyses

null effects or that there is an imbalance between the different manifestations of the demographic variables.

6 DISCUSSION

The final Cybersecurity Self-Efficacy in Smart Homes (CySESH) scale consists of 12 items that unidimensionally measure the domain-specific self-efficacy beliefs of users. Through five studies, we designed and validated its items to ensure the scale's psychodiagnostic quality. Reliability coefficients demonstrated excellent measurement precision. This was indicated by the consistent performance of smart home users across CySESH items and the extent to which all items reflected the same latent construct. In addition to the content validity offered by diverse experts in the pilot studies, construct validity was substantiated via correlational analyses. CySESH showed the expected relationship trends with other established psychological traits (i.e., self-efficacy in information security, outcome expectation, self-esteem, and optimism), which confirmed its distinctness. By following state-of-the-art test construction and open science processes, we hope that CySESH serves researchers and practitioners as a meaningful evaluation tool.

6.1 Limitations and Future Work

A valid instrument is merely the requirement for robust evidence. Long-term observations using CySESH, ideally including replications [46], will be needed to reach more clarity of other validity aspects. One validity aspect future work should consider are differences between self-efficacy beliefs and the corresponding skills or proficiency [125], which we did not assess. Meta-analytic work from other areas suggest a medium-sized correlation between job experience and self-efficacy [66]. Future validation should assess whether our validity benchmarks can be replicated with a population including highly experienced IT professionals. Depending on the view of the criterion [36], the question of domain-specificity [8, 9, 11] could be addressed within further content validation [2]. In our work, we did not emphasize domain-specific self-efficacy differences between smart home use and smart home cybersecurity use, given that there is related evidence on the difference in those two domains of functioning (for example, smartphones [20], online social networks [113], smart thermostats [73]). However, demonstrating CySESH's distinct applicability to smart home use activities would certainly provide additional content validation.

With regard to generalizability, we have to limit our conclusions to characteristics of a convenience sample recruited from Prolific as well as English speaking cultures. The sample was not strictly stratified analogous to representative census data, which would have prevented the gender imbalance. Our study only used participants that specifically reported owning a smart-home device as we could pre-select such persons through Prolific. At the time, Prolific had a user base that was skewed towards women [29], and so was our sample. This should not imply that women own more smart home devices than men. We do not expect the observed gender effects of $\beta_{\text{male}} = 0.28$ to impact the validity of our measurement as our study is in line with the current body of evidence for gender differences in self-efficacy, which further interact with cultural differences: Halevi et al. [56] reported a large gender difference in self-efficacy in the USA. Other single-region studies found varying results from zero difference between genders in a US-American study [84] over small differences in South Africa [123] to medium sized differences in an US-American [3] and Malay sample [43].

Across those studies, differences were always due to higher scores for males, despite some samples showing female overrepresentation. Considering these works, and since 87.5% of our participants reported their country of residence as UK or US, it is not surprising to detect a significant gender effect. However, it leaves future work to investigate gender-specific questions with more well-suited samples than ours.

We also do not expect our significant age effect of $\beta_{\text{age}} = -0.017$ to impact validity, as it is very small in magnitude. An effect should be evaluated not only by its significance (i.e. difference from zero) but also by its relevance in terms of magnitude, i.e. it should have a meaningful influence on the outcome [71]. We argue that our age effect does not satisfy the second criterion based on its magnitude. Keeping in mind that the underlying regression including every demographic factor only could explain 5.1% of variance in CySESH scores and our high sample size likely was the only factor that enabled detection of this very small effect, we argue that the age effect is not large enough to imply generalizability problems.

Given the promising results from our validation studies, it would also be informative to replicate findings of the current literature – which mostly rely on ad-hoc developed scales [18] – in order to introduce CySESH as a standardized measure among them. We plan to develop a short form (i.e., maximize the test economy) that benefits the large-scale CySESH use necessary for this endeavor. To allow for continued application, it will be inevitable to revise the items' wording as needed, e.g., in case of future changes in smart home security and privacy standards or user interactions.

6.2 Considerations for Using CySESH

To reinforce objectivity of research that uses CySESH, we give final guidelines for using CySESH. The instructions for participants, response format, and items are publicly accessible on the OSF (file link: CySESH scale final version). Middle values can be interpreted as the most frequent and - consequently - represent medium self-efficacy strength. Extreme values, on grounds of standard deviation, reflect participants who either have a strong (high values) or weak (low values) belief in their capability to control information processed by smart home devices and systems against unauthorized disclosure, modification, loss, or destruction.

Using CySESH can be a valuable method to evaluate human factors of cybersecurity in smart homes. First, we suggest CySESH be used in empirical HCI studies. Given its lightweight application, CySESH can be used to pre-screen study participants to ensure specific sampling distributions for self-efficacy manifestations. Researchers might take special interest in evaluating the time-stability of self-efficacy beliefs. Regardless, it will be crucial to assess CySESH's predictive validity for security behaviors. Practitioners can use this information to strategically support those users who are more likely to engage in future behaviors that compromise their smart home security and privacy. Here, we suggest to use CySESH as an assessment tool implemented in technological wizards or commissioning assistants. Developing interfaces that match the user's individual level of self-efficacy with the appropriate measures may contribute to genuine usable security. Lastly, CySESH can inform policy makers about the status quo of the people's digital sovereignty when included into census data surveys. Prevalent user profiles with a heightened risk of cybersecurity or privacy issues could be identified. Significant consumer protection measures may follow to acknowledge certain risk groups.

Understanding CySESH as a useful foundation of cybersecurity self-efficacy measurement can also inspire important methodological and substantive work in this domain. Children and adolescents, for example, increasingly become consumers of technologies with specific security and privacy vulnerabilities, either actively (e.g., by using personal devices such as as smart phones) or passively (e.g., by living in a smart home equipped with devices installed by their parents). Similarly, elderly people may face an increase of connected technology in their own homes or care facilities with supportive or medical functions. Understanding such population-specific use patterns and attitudes is important to predict security risks and implement measures to minimize them, but at the same time requires theory-driven modifications to validated measurements of relevant constructs, such as CySESH.

6.3 Conclusion

In this paper, we present the validation of the Cybersecurity Self-Efficacy in Smart Homes scale. Research and its practical implications for secure user behaviors are limited by the ability to measure important latent constructs, such as self-efficacy. Across five qualitative and quantitative studies, we developed a 12-item scale that measures cybersecurity self-efficacy in smart homes. The scale is a publicly accessible, lightweight, domain-specific assessment tool with use cases for researchers, HCI practitioners, and policy makers. An objective, reliable, and valid scale benefits the reduction of bias and error. Further, it facilitates replicability and generalizability of research. We provide a methodological contribution to the standardization of this emerging IT security and privacy research field that will allow for meaningful research consensus and the informed design of interfaces to support cybersecurity self-efficacy.

ACKNOWLEDGMENTS

This work was funded by the German Federal Ministry of Education and Research (BMBF) (grant no. V5DISO056-03) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (grant no. EXC 2092 CASA - 390781972). M.E. is supported by the Digital Society research program funded by the Ministry of Culture and Science of North Rhine-Westphalia, Germany (grant no. 1706dgn006).

REFERENCES

[1] Manfred Amelang and Lothar Schmidt-Atzert. 2006. *Psychologische Diagnostik und Intervention* (4 ed.). Springer, Heidelberg, Germany. https://doi.org/10.1007/3-540-28507-5

[2] American Psychological Association. 1954. Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin* 51, 2:2 (1954), 1–38. https://doi.org/10.1037/h0053479

[3] Mohd Anwar, Wu He, Ivan Ash, Xiaohong Yuan, Ling Li, and Li Xu. 2017. Gender difference and employees' cybersecurity behaviors. *Computers in Human Behavior* 69, 4 (2017), 437–443. https://doi.org/10.1016/j.chb.2016.12.040

[4] Noah Apthorpe, Dillon Reisman, Srikanth Sundaresan, Arvind Narayanan, and Nick Feamster. 2017. Spying on the Smart Home: Privacy Attacks and Defenses on Encrypted IoT Traffic: arXiv preprint. https://arxiv.org/abs/1708.05044

[5] Renana Arizon-Peretz, Irit Hadar, and Gil Luria. 2021. The Importance of Security is in the Eye of the Beholder: Cultural, Organizational, and Personal Factors Affecting the Implementation of Security by Design. *IEEE Transactions on Software Engineering* 48, 11 (2021), 4433–4446. https://doi.org/10.1109/TSE.2021.3119721

[6] Kregg Aytes and Terry Conolly. 2003. A Research Model for Investigating Human Behavior Related to Computer Security. In *AMCIS 2003 Proceedings*, Association for Information Systems (Ed.). AIS Electronic Library, Tampa, FL, United States, 206. https://aisel.aisnet.org/amcis2003/260

[7] Nazmiye Balta-Ozkan, Rosemary Davidson, Martha Bicket, and Lorraine Whitmarsh. 2013. Social barriers to the adoption of smart homes. *Energy Policy* 63, 12 (2013), 363–374. https://doi.org/10.1016/j.enpol.2013.08.043

[8] Albert Bandura. 1977. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review* 84, 2 (1977), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

[9] Albert Bandura. 1982. Self-efficacy mechanism in human agency. *American Psychologist* 37, 2 (1982), 122–147. https://doi.org/10.1037/0003-066X.37.2.122

[10] Albert Bandura. 1986. *Social foundations of thought and action: A social cognitive theory.* Prentice-Hall, Englewood Cliffs, NJ, USA.

[11] Albert Bandura. 1997. *Self-efficacy: The exercise of control.* W H Freeman/Times Books/ Henry Holt & Co, New York, NY, USA.

[12] Albert Bandura. 2006. Guide for constructing self-efficacy scales. In *Self-efficacy beliefs of adolescents*, Frank Pajares and Timothy C. Urdan (Eds.). IAP - Information Age Pub. Inc, Greenwich, CT, USA, 307–337.

[13] Albert Bandura. 2010. Self-Efficacy. In *The Corsini Encyclopedia of Psychology*, Irving B. Weiner and W. Edward Craighead (Eds.). John Wiley & Sons, Inc, Hoboken, NJ, USA. https://doi.org/10.1002/9780470479216.corpsy0836

[14] Albert Bandura. 2012. On the Functional Properties of Perceived Self-Efficacy Revisited. *Journal of Management* 38, 1 (2012), 9–44. https://doi.org/10.1177/0149206311410606

[15] Albert Bandura and Edwin A. Locke. 2003. Negative self-efficacy and goal effects revisited. *The Journal of applied psychology* 88, 1 (2003), 87–99. https://doi.org/10.1037/0021-9010.88.1.87

[16] Godfred O. Boateng, Torsten B. Neilands, Edward A. Frongillo, Hugo R. Melgar-Quiñonez, and Sera L. Young. 2018. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in public health* 6 (2018), 149. https://doi.org/10.3389/fpubh.2018.00149

[17] Andrew Booth, Diana Papaioannou, and Anthea Sutton. 2012. *Systematic approaches to a successful literature review.* Sage, Los Angeles, Calif.

[18] Nele Borgert, Jennifer Friedauer, Imke Böse, Angela M. Sasse, and Malte Elson. 2021. The Study of Cybersecurity Self-Efficacy: A Systematic Literature Review of Methodology. In *USENIX Symposium on Usable Privacy and Security (SOUPS) 2021*. USENIX Association, Virtual Conference, 1–4. https://www.usenix.org/system/files/soups21-abstract-poster56-borgert.pdf

[19] Nele Borgert, Luisa Jansen, Imke Böse, Jennifer Friedauer, M. Angela Sasse, and Malte Elson. 2023. A Decade of Dividedness: A Preregistered Systematic Review of the Cybersecurity Self-efficacy Methods. arXiv:10.31234/osf.io/ybec9

[20] Frank Breitinger, Ryan Tully-Doyle, and Courtney Hassenfeldt. 2020. A survey on smartphone user's security choices, awareness and education. *Computers & Security* 88 (2020), 101647. https://doi.org/10.1016/j.cose.2019.101647

[21] Robert L. Brennan. 2001. An Essay on the History and Future of Reliability from the Perspective of Replications. *Journal of Educational Measurement* 38, 4 (2001), 295–317. https://doi.org/10.1111/j.1745-3984.2001.tb01129.x

[22] Timothy A. Brown. 2015. *Confirmatory factor analysis for applied research.* The Guilford Press, New York and London. https://fid.fachportal-paedagogik.de/ebscoProxySearch/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=831411

[23] Burcu Bulgurcu, Hasan Cavusoglu, and Izak Benbasat. 2010. Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness. *MIS Quarterly* 34, 3 (2010), 523–548. https://doi.org/10.2307/25750690

[24] Matthias Burisch. 1984. Approaches to personality inventory construction: A comparison of merits. *American Psychologist* 39, 3 (1984), 214–227. https://doi.org/10.1037/0003-066X.39.3.214

[25] Kevin L. Burke, A. Barry Joyner, Daniel R. Czech, and Matthew J. Wilson. 2000. An investigation of concurrent validity between two optimism/pessimism questionnaires: The life orientation test-revised and the optimism/pessimism scale. *Current Psychology* 19, 2 (2000), 129–136. https://doi.org/10.1007/s12144-000-1009-5

[26] Jennifer D. Campbell and Loraine F. Lavallee. 1993. Who am I? The Role of Self-Concept Confusion in Understanding the Behavior of People with Low Self-Esteem. In *Self-Esteem*, Roy F. Baumeister (Ed.). Springer, Boston, MA, USA, 3–20. https://doi.org/10.1007/978-1-4684-8956-9{_}1

[27] George Chalhoub, Ivan Flechais, Norbert Nthala, and Ruba Abu-Salma. 2020. Innovation Inaction or In Action? The Role of User Experience in the Security and Privacy Design of Smart Home Cameras. In *USENIX Symposium on Usable Privacy and Security (SOUPS) 2020*. USENIX Association, Virtual Conference, 185–204.

[28] Mark Chan, Irene Woon, and Atreyi Kankanhalli. 2005. Perceptions of Information Security in the Workplace: Linking Information Security Climate to Compliant Behavior. *Journal of Information Privacy and Security* 1, 3 (2005), 18–41. https://doi.org/10.1080/15536548.2005.10855772

[29] Nick Charalambides. 2021. We recently went viral on TikTok - here's what we learned. https://www.prolific.co/blog/we-recently-went-viral-on-tiktok-heres-what-we-learned

[30] Francesca Chiesi, Silvia Galli, Caterina Primi, Paolo Innocenti Borgi, and Andrea Bonacchi. 2013. The accuracy of the Life Orientation Test-Revised (LOT-R) in measuring dispositional optimism: evidence from item response theory analyses. *Journal of personality assessment* 95, 5 (2013), 523–529. https://doi.org/10.1080/00223891.2013.781029

[31] Lian-Hwang Chiu. 1988. Measures of Self-Esteem for School-Age Children. *Journal of Counseling & Development* 66, 6 (1988), 298–301. https://doi.org/10.1002/j.1556-6676.1988.tb00874.x

[32] Jacob Cohen. 1992. Quantitative methods in psychology: A power primer. *Psychological Bulletin* 112, 1 (1992), 153–159.

[33] Consumers International and Internet Society. 2019. The Trust Opportunity: Exploring Consumer attitudes to the Internet of Things. https://www.internetsociety.org/resources/doc/2019/trust-opportunity-exploring-consumer-attitudes-to-iot/

[34] Rick Crandall. 1973. The measurement of self-esteem and related constructs. In *Measures of Social Psychological Attitudes*, John P. Robinson and Phillip R. Shaver (Eds.). The University of Michigan, Ann Arbor, MI, USA, 45–167.

[35] Lee J. Cronbach. 1971. Test Validation. In *Educational Measurement*, Robert L. Thorndike (Ed.). American Council on Education, Washington DC, USA, 443–507.

[36] Lee J. Cronbach and Paul E. Meehl. 1955. Construct validity in psychological tests. *Psychological Bulletin* 52 (1955), 281–302.

[37] Robert E. Crossler. 2010. Protection Motivation Theory: Understanding Determinants to Backing Up Personal Data. In *48th Hawaii International Conference on System Sciences (Hicss)*. IEEE, Kauai, HI, USA, 1–10. https://doi.org/10.1109/HICSS.2010.311

[38] Robert F. DeVellis. 2012. *Scale Development: Theory and Application.* SAGE Publications, Los Angeles, CA, USA.

[39] Charlette Donalds and Kweku-Muata Osei-Bryson. 2020. Cybersecurity compliance behavior: Exploring the influences of individual decision style and other antecedents. *International Journal of Information Management* 51 (2020), 102056. https://doi.org/10.1016/j.ijinfomgt.2019.102056

[40] Malte Elson. 2019. Examining Psychological Science Through Systematic Meta-Method Analysis: A Call for Research. *Advances in Methods and Practices in Psychological Science* 2, 4 (2019), 350–363. https://doi.org/10.1177/2515245919863296

[41] ENISA. 2019. Cybersecurity culture guidelines: Behavioural aspects of cybersecurity: European Union Agency For Network and Information Security Report 2019. https://doi.org/10.2824/324042

[42] Ulrike Fasbender. 2020. Outcome Expectancies. In *Encyclopedia of personality and individual differences*, Virgil Zeigler-Hill (Ed.). Springer International Publishing, Cham, 3377–3379. https://doi.org/10.1007/978-3-319-24612-3{_}1802

[43] Faith B. Fatokun, Suraya Hamid, Azah Anir Norman, and Johnson O. Fatokun. 2019. The Impact of Age, Gender, and Educational level on the Cybersecurity Behaviors of Tertiary Institution Students: An Empirical investigation on Malaysian Universities. *Journal of Physics: Conference Series* 1339, 1 (2019), 012098. https://doi.org/10.1088/1742-6596/1339/1/012098

[44] Massimo Ficco and Francesco Palmieri. 2019. Leaf: An open-source cybersecurity training platform for realistic edge-IoT scenarios. *Journal of Systems Architecture* 97 (2019), 107–129. https://doi.org/10.1016/j.sysarc.2019.04.004

[45] Andy P. Field. 2001. Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological methods* 6, 2 (2001), 161–180. https://doi.org/10.1037/1082-989X.6.2.161

[46] Jessica Kay Flake, Ian J. Davidson, Octavia Wong, and Jolynn Pek. 2022. Construct validity and the validity of replication studies: A systematic review. *The American psychologist* 77, 4 (2022), 576–588. https://doi.org/10.1037/amp0001006

[47] Jessica Kay Flake and Eiko I. Fried. 2020. Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Advances in Methods and Practices in Psychological Science* 3, 4 (2020), 456–465. https://doi.org/10.1177/2515245920952393

[48] Santos M. Galvez, Joshua D. Shackman, Indira R. Guzman, and Shuyuan M. Ho. 2015. Factors Affecting Individual Information Security Practices. In *Proceedings of the 2015 ACM SIGMIS Conference on Computers and People Research (ACM Digital Library)*, Diana Burley (Ed.). ACM, New York, NY, USA, 135–144. https://doi.org/10.1145/2751957.2751966

[49] Radhika Garg, Hua Cui, Spencer Seligson, Bo Zhang, Martin Porcheron, Leigh Clark, Benjamin R. Cowan, and Erin Beneteau. 2022. The Last Decade of HCI Research on Children and Voice-based Conversational Agents. In *CHI Conference on Human Factors in Computing Systems (ACM Digital Library)*, Simone Barbosa (Ed.). Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3491102.3502016

[50] Stacey George, Michael Clark, and Maria Crotty. 2007. Development of the Adelaide driving self-efficacy scale. *Clinical rehabilitation* 21, 1 (2007), 56–61. https://doi.org/10.1177/0269215506071284

[51] Heide Glaesmer, Winfried Rief, Alexandra Martin, Ricarda Mewes, Elmar Brähler, Markus Zenger, and Andreas Hinz. 2012. Psychometric properties and population-based norms of the Life Orientation Test Revised (LOT-R). *British journal of health psychology* 17, 2 (2012), 432–445. https://doi.org/10.1111/j.2044-8287.2011.02046.x

[52] Dan Goodin. 2021. Home alarm tech backdoored security cameras to spy on customers having sex: Employee for ADT accessed ˜200 customer cams on more than 9,600 occasions. https://arstechnica.com/information-technology/2021/01/home-alarm-tech-backdoored-security-cameras-to-spy-on-customers-having-sex/

[53] James M. Graham. 2006. Congeneric and (Essentially) Tau-Equivalent Estimates of Score Reliability. *Educational and Psychological Measurement* 66, 6 (2006), 930–944. https://doi.org/10.1177/0013164406288165

[54] Eliza M. Grames, Andrew N. Stillman, Morgan W. Tingley, and Chris S. Elphick. 2019. An automated approach to identifying search terms for systematic reviews using keyword co–occurrence networks. *Methods in Ecology and Evolution* 10, 10 (2019), 1645–1654. https://doi.org/10.1111/2041-210X.13268

[55] Nadine Guhr, Oliver Werth, Philip Peter Hermann Blacha, and Michael H. Breitner. 2020. Privacy concerns in the smart home context. *SN Applied Sciences* 2, 2 (2020), 247. https://doi.org/10.1007/s42452-020-2025-8

[56] Tzipora Halevi, Nasir Memon, James Lewis, Ponnurangam Kumaraguru, Sumit Arora, Nikita Dagar, Fadi Aloul, and Jay Chen. 2016. Cultural and psychological factors in cyber-security. In *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services (ACM Digital Library)*, Gabriele Anderst-Kotsis (Ed.). ACM, New York, NY, USA, 318–324. https://doi.org/10.1145/3011141.3011165

[57] Ronald K. Hambleton and Daniel R. Eignor. 1979. *A Practitioner's Guide to Criterion-Referenced Test Development, Validation, and Test Score Usage* (2 ed.). University of Massachusetts, School of Educacation, Laboratory of Psychometric and Evaluative Research, Amherst, MA, USA.

[58] Bartlomiej Hanus and Yu "Andy" Wu. 2016. Impact of Users' Security Awareness on Desktop Security Behavior: A Protection Motivation Theory Perspective. *Information Systems Management* 33, 1 (2016), 2–16. https://doi.org/10.1080/10580530.2015.1117842

[59] Tom Hargreaves, Charlie Wilson, and Richard Hauxwell-Baldwin. 2018. Learning to live in a smart home. *Building Research & Information* 46, 1 (2018), 127–139. https://doi.org/10.1080/09613218.2017.1286882

[60] Andrew F. Hayes and Jacob J. Coutts. 2020. Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But. . .. *Communication Methods and Measures* 14, 1 (2020), 1–24. https://doi.org/10.1080/19312458.2020.1718629

[61] Todd F. Heatherton and Carrie L. Wyland. 2009. Assessing self-esteem. In *Positive psychological assessment*, Shane J. Lopez and C. R. Snyder (Eds.). American Psychological Assoc, Washington, DC, 219–233. https://doi.org/10.1037/10612-014

[62] Gordon Hodson. 2021. Construct jangle or construct mangle? Thinking straight about (nonredundant) psychological constructs. *Journal of Theoretical Social Psychology* 592, 4 (2021), 258. https://doi.org/10.1002/jts5.120

[63] Val Hooper and Chris Blunt. 2020. Factors influencing the information security behaviour of IT employees. *Behaviour & Information Technology* 39, 8 (2020), 862–874. https://doi.org/10.1080/0144929X.2019.1623322

[64] Angel Hsing-Chi Hwang and Andrea Stevenson Won. 2022. AI in Your Mind: Counterbalancing Perceived Agency and Experience in Human-AI Interaction. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (ACM Digital Library)*, Simone Barbosa (Ed.). Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3491101.3519833

[65] Ge Jin, Manghui Tu, Tae-Hoon Kim, Justin Heffron, and Jonathan White. 2018. Game based Cybersecurity Training for High School Students. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education (ACM Conferences)*, Tiffany Barnes (Ed.). ACM, New York, NY, USA, 68–73. https://doi.org/10.1145/3159450.3159591

[66] Timothy A. Judge, Christine L. Jackson, John C. Shaw, Brent A. Scott, and Bruce L. Rich. 2007. Self-efficacy and work-related performance: The integral role of individual differences. *Journal of Applied Psychology* 92, 1 (2007), 107–127. https://doi.org/10.1037/0021-9010.92.1.107

[67] Kyongseok Kim and Jooyoung Kim. 2011. Third-party Privacy Certification as an Online Advertising Strategy: An Investigation of the Factors Affecting the Relationship between Third-party Certification and Initial Trust. *Journal of Interactive Marketing* 25, 3 (2011), 145–158. https://doi.org/10.1016/j.intmar.2010.09.003

[68] Spyros Kokolakis. 2017. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security* 64 (2017), 122–134. https://doi.org/10.1016/j.cose.2015.07.002

[69] Daniel Koloseni, Chong Yee Lee, and Gan Ming Lee. 2018. Security policy compliance in public institutions: An integrative approach. *Journal of Applied Structural Equation Modeling* 2, 1 (2018), 13–28. https://doi.org/10.47263/jasem.2(1)03

[70] W. Kristof. 1983. Klassische Testtheorie und Testkonstruktion. In *Messen und Testen*, Hubert Feger and Jürgen Bredenkamp (Eds.). Springer, Göttingen, Germany, 544–603.

[71] Daniël Lakens. 2014. Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology* 44, 7 (2014), 701–710. https://doi.org/10.1002/ejsp.2023 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ejsp.2023.

[72] Jeremiah Lasquety-Reyes. 2021. Number of Smart Homes forecast in the World from 2017 to 2025. https://www.statista.com/forecasts/887613/number-of-smart-homes-in-the-smart-home-market-in-the-world

[73] Zachary E. Lee and K. Max Zhang. 2022. Unintended consequences of smart thermostats in the transition to electrified heating. *Applied Energy* 322 (2022), 119384. https://doi.org/10.1016/j.apenergy.2022.119384

[74] Cheng-Hsien Li. 2016. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior research methods* 48, 3 (2016), 936–949. https://doi.org/10.3758/s13428-015-0619-7

[75] Jiunn-Woei Lian. 2021. Understanding cloud-based BYOD information security protection behaviour in smart business: in perspective of perceived value. *Enterprise Information Systems* 15, 9 (2021), 1216–1237. https://doi.org/10.1080/17517575.2020.1791966

[76] Scott O. Lilienfeld and Adele N. Strother. 2020. Psychological measurement and the replication crisis: Four sacred cows. *Canadian Psychology/Psychologie canadienne* 61, 4 (2020), 281–288. https://doi.org/10.1037/cap0000236

[77] Sonia Lippke. 2020. Outcome Expectation. In *Encyclopedia of personality and individual differences*, Virgil Zeigler-Hill (Ed.). Springer International Publishing, Cham, 3379–3381. https://doi.org/10.1007/978-3-319-24612-3{_}1145

[78] Jean M. Lown. 2011. Development and Validation of a Financial Self-Efficacy Scale. *Journal of Financial Counseling and Planning* 22, 2 (2011), 54–63. https://ssrn.com/abstract=2006665

[79] J. E. Maddux and J. T. Gosselin. 2012. Self-efficacy. In *Handbook of self and identity*, M. R. Leary and J. P. Tangney (Eds.). The Guilford Press, New York, NY, USA, 198–224.

[80] James E. Maddux, Larry W. Norton, and Cal D. Stoltenberg. 1986. Self-efficacy expectancy, outcome expectancy, and outcome value: Relative effects on behavioral intentions. *Journal of Personality and Social Psychology* 51, 4 (1986), 783–789. https://doi.org/10.1037/0022-3514.51.4.783

[81] Philip R. Magaletta and Joan M. Oliver. 1999. The hope construct, will, and ways: Their relations with self-efficacy, optimism, and general well-being. *Journal of Clinical Psychology* 55, 5 (1999), 539–551. https://doi.org/10.1002/(SICI)1097-4679(199905)55:5<539::AID-JCLP2>3.0.CO;2-G

[82] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–32. https://doi.org/10.1145/3359183

[83] Jeffrey E. McGee, Mark Peterson, Stephen L. Mueller, and Jennifer M. Sequeira. 2009. Entrepreneurial Self–Efficacy: Refining the Measure. *Entrepreneurship Theory and Practice* 33, 4 (2009), 965–988. https://doi.org/10.1111/j.1540-6520.2009.00304.x

[84] Tanya McGill and Nik Thompson. 2018. Gender Differences in Information Security Perceptions and Behaviour. In *Australasian Conference on Information Systems 2018*. University of Technology, Sydney, Australia, Sydney, Australia. https://doi.org/10.5130/acis2018.co

[85] Diana Mîndrilă. 2010. Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) Estimation Procedures: A Comparison of Estimation Bias with Ordinal and Multivariate Non-Normal Data. *International Journal of Digital Society* 1, 1 (2010), 60–66.

[86] Erica M. Mitchell. 2020. *Cyber Security at Home: The Effect of Home User Perceptions of Personal Security Performance on Household IoT Security Intentions: Dissertation.* ProQuest LLC, Ann Arbor, MI, USA.

[87] Helfried Moosbrugger and Augustin Kelava (Eds.). 2020. *Testtheorie und Fragebogenkonstruktion*. Springer, Berlin, Heidelberg, Germany. https://doi.org/10.1007/978-3-662-61532-4

[88] Lily Hay Newman. 2016. The Botnet That Broke the Internet Isn't Going Away. https://www.wired.com/2016/12/botnet-broke-internet-isnt-going-away/

[89] Patricia A. Norberg, Daniel R. Horne, and David A. Horne. 2007. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs* 41, 1 (2007), 100–126. https://doi.org/10.1111/j.1745-6606.2006.00070.x

[90] Amy Orben and Daniël Lakens. 2020. Crud (Re)Defined. *Advances in Methods and Practices in Psychological Science* 3, 2 (2020), 238–247. https://doi.org/10.1177/2515245920917961 arXiv:https://doi.org/10.1177/2515245920917961 PMID: 33103054.

[91] Dorottya Papp, Kristóf Tamás, and Levente Buttyán. 2019. IoT Hacking – A Primer. *Infocommunications journal* 6, 2 (2019), 2–13. https://doi.org/10.36244/ICJ.2019.2.1

[92] Delroy L. Paulhus. 1991. Measurement and Control of Response Bias. In *Measures of personality and social psychological attitudes*, John P. Robinson, Phillip R. Shaver, and Lawrence S. Wrightsman (Eds.). Academic Press, San Diego, CA, USA, 17–59. https://doi.org/10.1016/B978-0-12-590241-0.50006-X

[93] Robert A. Peterson. 2000. A Meta-Analysis of Variance Accounted for and Factor Loadings in Exploratory Factor Analysis. *Marketing Letters* 11, 3 (2000), 261–275.

[94] Hiep Cong Pham, Linda Brennan, and Steven Furnell. 2019. Information security burnout: Identification of sources and mitigating factors from security demands and resources. *Journal of Information Security and Applications* 46 (2019), 96–107. https://doi.org/10.1016/j.jisa.2019.03.012

[95] Robert Pinka. 2021. Synthetic Deliberation: Can Emulated Imagination Enhance Machine Ethics? *Minds and Machines* 31, 1 (2021), 121–136. https://doi.org/10.1007/s11023-020-09531-w

[96] William James Popham. 1978. *Criterion-referenced measurement.* Prentice-Hall, Englewood Cliffs, NJ, USA.

[97] William T. Powers. 1991. Comment on Bandura's "human agency.". *American Psychologist* 46, 2 (1991), 151–153. https://doi.org/10.1037/0003-066X.46.2.151.b

[98] Tenko. Raykov and George A. Marcoulides. 2011. *Introduction to psychometric theory.* Routledge, New York.

[99] Byron Reeves and Clifford Nass. 1996. *The media equation: How people treat computers, television, and new media like real people.* Cambridge University Press, New York, NY, USA.

[100] Hyeun-Suk Rhee, Cheongtag Kim, and Young U. Ryu. 2009. Self-efficacy in information security: Its influence on end users' information security practice behavior. *Computers & Security* 28, 8 (2009), 816–826. https://doi.org/10.1016/j.cose.2009.05.008

[101] Morris Rosenberg. 1965. Rosenberg self-esteem scale (RSE). *Acceptance and commitment therapy. Measures package* 61, 52 (1965), 18.

[102] Morris Rosenberg. 1979. *Conceiving the Self.* Basic Books, New York, NY, USA.

[103] Regner Sabillon, Jordi Serra-Ruiz, Victor Cavaller, and Jeimy J. Cano M. 2021. An Effective Cybersecurity Training Model to Support an Organizational Awareness Program. In *Research anthology on artificial intelligence applications in security*, Information Resources Management Association (Ed.). IGI Global, Hershey, PA, USA, 174–188. https://doi.org/10.4018/978-1-7998-7705-9.ch008

[104] M. Angela Sasse, Jonas Hielscher, Jennifer Friedauer, and Annalina Buckmann. 2022. Rebooting IT-Security: How Organisations Can Encourage and Sustain Secure Behaviours. In *27th European Symposium on Research in Computer Security*. Springer, Copenhagen, Denmark, 1–18.

[105] Sergio Sayago, Barbara Barbosa Neves, and Benjamin R. Cowan. 2019. Voice assistants and older people. In *Proceedings of the 1st International Conference on Conversational User Interfaces (ACM Digital Library)*, Benjamin R. Cowan (Ed.). Association for Computing Machinery, New York, NY, USA, 1–3. https://doi.org/10.1145/3342775.3342803

[106] Michael F. Scheier and Charles S. Carver. 1985. Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology* 4, 3 (1985), 219–247. https://doi.org/10.1037/0278-6133.4.3.219

[107] Michael F. Scheier, Charles S. Carver, and Michael W. Bridges. 1994. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology* 67, 6 (1994), 1063–1078. https://doi.org/10.1037/0022-3514.67.6.1063

[108] Michael F. Scheier, Charles S. Carver, and Michael W. Bridges. 2002. Optimism, pessimism, and psychological well-being. In *Optimism & pessimism*, Edward C. Chang and Edward Chin-Ho Chang (Eds.). American Psychological Association, Washington, D.C., USA, 189–216. https://doi.org/10.1037/10385-009

[109] Inger Schou-Bredal, Trond Heir, Laila Skogstad, Tore Bonsaksen, Anners Lerdal, Tine Grimholt, and Øivind Ekeberg. 2017. Population-based norms of the Life Orientation Test-Revised (LOT-R). *International journal of clinical and health psychology : IJCHP* 17, 3 (2017), 216–224. https://doi.org/10.1016/j.ijchp.2017.07.005

[110] Ahmad Bakhtiyari Shahri, Zuraini Ismail, and Shahram Mohanna. 2016. The Impact of the Security Competency on "Self-Efficacy in Information Security" for Effective Health Information Security in Iran. *Journal of medical systems* 40, 11 (2016), 241. https://doi.org/10.1007/s10916-016-0591-5

[111] Hawal Shamon and Carl Clemens Berning. 2020. Attention Check Items and Instructions in Online Surveys: Boon or Bane for Data Quality? *Survey Research Methods* 14, 1 (2020), 55–77. https://doi.org/10.18148/srm/2020.v14i1.7374

[112] Mark Sherer, James E. Maddux, Blaise Mercandante, Steven Prentice-Dunn, Beth Jacobs, and Ronald W. Rogers. 1982. The Self-Efficacy Scale: Construction and Validation. *Psychological Reports* 51, 2 (1982), 663–671. https://doi.org/10.2466/pr0.1982.51.2.663

[113] Katherine P. Strater and Heather Richter Lipford. 2008. Strategies and Struggles with Privacy in an Online Social Networking Community. In *People and Computers XXII Culture, Creativity, Interaction (HCI) (Electronic Workshops in Computing)*. BCS Learning & Development, Liverpool, UK, 111–119. https://doi.org/10.14236/ewic/HCI2008.11

[114] Hyewon Suh, Nina Shahriaree, Eric B. Hekler, and Julie A. Kientz. 2016. Developing and Validating the User Burden Scale. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (ACM Digital Library)*, Jofish Kaye (Ed.). ACM, New York, NY, USA, 3988–3999. https://doi.org/10.1145/2858036.2858448

[115] Yan Sun and Reenay Rogers. 2021. Development and validation of the Online Learning Self-efficacy Scale (OLSS): A structural equation modeling approach. *American Journal of Distance Education* 35, 3 (2021), 184–199. https://doi.org/10.1080/08923647.2020.1831357

[116] Ahn Sungyong. 2021. Symmetrifying Smart Home: Topological Power and the New Governmentality of the Internet of Things. *Media Theory* 5, 1 (2021), 89–114.

[117] Wei Tang, Ying Cui, and Oksana Babenko. 2014. Internal Consistency: Do We Really Know What It Is and How to Assess It? *Journal of Psychology and Behavioral Science* 2, 2 (2014), 205–220.

[118] Robert L. Thorndike and Elizabeth P. Hagen. 1977. *Measurement and evaluation in education and psychology* (4 ed.). Wiley, New York, NY, USA.

[119] Italo Trizano-Hermosilla and Jesús M. Alvarado. 2016. Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements. *Frontiers in psychology* 7 (2016), 769. https://doi.org/10.3389/fpsyg.2016.00769

[120] Emily L. Tuthill, Jacqueline M. McGrath, Melanie Graber, Regina M. Cusson, and Sera L. Young. 2016. Breastfeeding Self-efficacy: A Critical Review of Available Instruments. *Journal of human lactation* 32, 1 (2016), 35–45. https://doi.org/10.1177/0890334415599533

[121] Anna-Sophie Ulfert-Blank and Isabelle Schmidt. 2022. Assessing digital self-efficacy: Review and scale development. *Computers & Education* 191, 1 (2022), 104626. https://doi.org/10.1016/j.compedu.2022.104626

[122] Tommy van Steen and Julia R. A. Deeleman. 2021. Successful Gamification of Cybersecurity Training. *Cyberpsychology, behavior and social networking* 24, 9 (2021), 593–598. https://doi.org/10.1089/cyber.2020.0526

[123] Silas Formunyuy Verkijika. 2019. "If you know what to do, will you take action to avoid mobile phishing attacks": Self-efficacy, anticipated regret, and gender. *Computers in Human Behavior* 101 (2019), 286–296. https://doi.org/10.1016/j.chb.2019.07.034

[124] Neil Vigdor. 2019. Somebody's Watching: Hackers Breach Ring Home Security Cameras. https://www.nytimes.com/2019/12/15/us/Hacked-ring-home-security-cameras.html

[125] Daniel Votipka, Desiree Abrokwa, and Michelle L. Mazurek. 2020. Building and Validating a Scale for Secure Software Development Self-Efficacy. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (ACM Digital Library)*, Regina Bernhaupt (Ed.). Association for Computing Machinery, New York, NY, USA, 1–20. https://doi.org/10.1145/3313831.3376754

[126] Aaron C. Weidman, Conor M. Steckler, and Jessica L. Tracy. 2017. The jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion* 17, 2 (2017), 267–295. https://doi.org/10.1037/emo0000226

[127] Maor Weinberger, Maayan Zhitomirsky-Geffet, and Dan Bouhnik. 2017. Factors affecting users' online privacy literacy among students in Israel. *Online Information Review* 41, 5 (2017), 655–671. https://doi.org/10.1108/OIR-05-2016-0127

[128] Yaxing Yao, Justin Reed Basdeo, Oriana Rosata Mcdonough, and Yang Wang. 2019. Privacy Perceptions and Designs of Bystanders in Smart Homes. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–24. https://doi.org/10.1145/3359161

[129] Seounmi Youn. 2009. Determinants of Online Privacy Concern and Its Influence on Privacy Protection Behaviors Among Young Adolescents. *Journal of Consumer Affairs* 43, 3 (2009), 389–418. https://doi.org/10.1111/j.1745-6606.2009.01146.x

[130] Ross McD Young, Tian P. S. Oei, and Gabrielle M. Crook. 1991. Development of a drinking self-efficacy questionnaire. *Journal of Psychopathology and Behavioral Assessment* 13, 1 (1991), 1–15. https://doi.org/10.1007/BF00960735

[131] Brahim Zarouali, Karolien Poels, Koen Ponnet, and Michel Walrave. 2018. "Everything under control?": Privacy control salience influences both critical processing and perceived persuasiveness of targeted advertising among adolescents. *Cyberpsychology Journal of Psychosocial Research on Cyberspace* 12, 1 (2018), 5. https://doi.org/10.5817/cp2018-1-5

[132] Eric Zeng, Shrirang Mare, and Franziska Roesner. 2017. End User Security & Privacy Concerns with Smart Homes. In *USENIX Symposium on Usable Privacy and Security (SOUPS) 2017 (SOUPS '17)*. USENIX Association, USA, 65–80.

[133] Serena Zheng, Noah Apthorpe, Marshini Chetty, and Nick Feamster. 2018. User Perceptions of Smart Home IoT Privacy. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–20. https://doi.org/10.1145/3274469