

OSF REGISTRATION FORM: Cybersecurity Self-Efficacy

Study Information

1. Title (required)

The Study of Cybersecurity Self-Efficacy: A Systematic Literature Review of Methodology and Application

2. Authors (required)

Nele Borgert^{1,3}

Jennifer Friedauer^{2,3}

Imke Böse^{1,3}

Martina Angela Sasse^{2,3}

Malte Elson^{1,3}

¹ Psychology of Human Technology Interaction, Faculty of Psychology, Ruhr University Bochum

² Human Centered Security, Faculty of Electrical Engineering and Information Technology, Ruhr University Bochum

³ Horst Görtz Institute for IT Security, Ruhr University Bochum

3. Description (optional)

Overall Project Description:

The goal of this project is to obtain a systematic assessment of how cybersecurity self-efficacy research is conducted. We will work on two topics: (1) metrics that exist to measure and influence self-efficacy in IT-security/-privacy contexts and (2) messages that practitioners can take away from current research. We will perform a literature review on both topics, working with published data from 2010 to 2021.

Review Description:

Cybersecurity self-efficacy is the belief about one's own ability to enact skills related to IT-security/-privacy. It is an important construct that inherently affects individual security behaviors. Implementing effective security behaviors is key to maintaining safe use of smart devices and ensuring one's own data is protected. Given the sensitivity of data collected by e.g., smart home appliances, its protection is of high relevance. Still, there remains a lack of systematic research regarding the psychological construct of cybersecurity self-efficacy. The review aims to systematically investigate published empirical research on cybersecurity self-efficacy.

For this study, a PROSPERO registration form was completed to maximize the transparency of our systematic review (see attachment "Prospero Registration Form"). The literature review is preregistered on the Open Science Framework.

4. Hypotheses (required)

Instead of including hypotheses, we formulated specific research questions.

Review A:

1. What are the demographics of studies of cybersecurity self-efficacy? (Including year of study, study type, sample groups, sample size, countries)
2. What measures are used to assess cybersecurity self-efficacy? What are the scale characteristics and reported psychometrics?
3. What role does cybersecurity self-efficacy play in the theoretical or research models of the studies? (Including cause or outcome)
4. Do the studies report interventions that have been carried out to manipulate cybersecurity self-efficacy? If so, what was their approach?
5. Are there studies of cybersecurity self-efficacy regarding smart home devices? If so:
 - a. What are the demographics of those studies? (Including year of study, study type, sample groups, sample size, and countries)
 - b. What measures are used to assess cybersecurity self-efficacy in those studies? What are the scale characteristics and reported psychometrics?
 - c. What role does cybersecurity self-efficacy play in the theoretical or research models of those studies? (Including cause or outcome)
 - d. Do those studies report interventions that have been carried out to manipulate cybersecurity self-efficacy? If so, what was their approach?

Review B:

[Research Question 1 – to describe study sample]

6. Do the used measures take the context in which cybersecurity self-efficacy is needed into account? (Impact of context)
 - a. Is the cybersecurity self-efficacy instrument adapted to a specific situation? (e.g. item text adapted to a scenario)
 - b. Is environmental impact measured in the study? (e.g. questions that ask for working conditions / conditions of use; behavior of colleagues or others who share the environment)
 - c. Are there items in the cybersecurity self-efficacy instrument that take emotional states into account? Or are emotions measured in an additional metric?
7. Is the security task in the conducted studies a primary or secondary task?
8. Do the conducted studies give advice to practitioners on how to improve cybersecurity self-efficacy? Who is addressed? (Organization, single position in organization, end user)
9. Do the addressed practitioners (organization, single position in organization, end user) need to be active? Do they need abilities or characteristics to implement the given advice? (User / practitioner model)
10. What behavior or what actions are suggested as an outcome of the conducted studies to improve cybersecurity self-efficacy in practice? (Types of training)

Design Plan

5. Study type (required)
 - 5.1 A systematic review of published studies.

6. Blinding (required)

6.1 Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply.

None apply

7. Is there any additional blinding in this study?

Reviewers will be blinded to each other's decisions during data extraction (each reviewer extracts data from two-thirds of the papers).

8. Study design (required)

This study is a systematic literature review. Three reviewers (NB, JF, IB) will be responsible for study selection and data extraction.

Prior to the process of selection and extraction, reviewers will be trained with studies excluded from the review due to their publication year (2009). If necessary, codebooks will be adjusted until interrater agreement reaches a satisfactory level (iota coefficient > 0.6 for the training data set). If codebook adjustments are made, we will make those changes transparent again in a supplementary pre-registration.

The general process will be as follows: 1. title sift (exclude nonrelevant literature based on title), 2. abstract sift (exclude nonrelevant literature based on abstract), 3. full-text sift (exclude nonrelevant literature based on full-text). Reasons will always be documented for why studies do not meet inclusion criteria. Disagreements of selection or extraction in the full-text sift stage will be discussed in group (2/3 vote wins).

9. Randomization (optional)

Identified studies for the review will be randomly split into three separate, identically sized blocks for the title and abstract sift. A second randomization takes place before data will be extracted in the full-text sift (three blocks à two-thirds of the remaining yet included studies). To create randomized IDs for the reviewers, we will use randomizer.org.

Sampling Plan

10. Existing data (required)

10.1.1. Registration prior to creation of data: As of the date of submission of this research plan for preregistration, the data have not yet been collected, created, or realized.

11. Explanation of existing data (optional)

-

12. Data collection procedures (required)

Selection Criteria:

We will apply the following selection criteria for all studies:

- It should be an empirical study.

- It should examine the relationship between self-efficacy and IT-security/-privacy.
- It should cover security or privacy in the IT/cyber context.
- It should include a measure of self-efficacy.
- It should be published ≥ 2010 .
- Inclusion of: all populations, all interventions or exposures, qualitative and quantitative study types, all cultures and countries.
- Exclusion of: individual case studies ($N = 1$), studies not published in English, studies where self-efficacy is not measured regarding IT-security/-privacy

Databases:

The search strategy will cover the following search sources (databases): EBSCOhost, IEEE Xplore, ACM Digital Library, Science Direct, dimensions.ai, arXiv, Scopus, Web of Science, and Wiley Online Library. No other search methods will be used.

Search Terms:

Our search string will be: ("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"). These words must appear in title and/or abstract (to ensure this requirement, we will use a supplementary self-written code [see attachment "Search Improvement"]).

For the specific strings applied in the respective database see attachment "Research Terms".

13. Sample size (required)

We will export all hits in each database for our search string. The minimum number of studies for synthesis is two.

14. Sample size rationale (optional)

Synthesis with less than two studies is not possible.

15. Stopping rule (optional)

We will consider papers published in the databases from 01/01/2010 to our search date.

Variables

16. Manipulated variables (optional)

-

17. Measured variables (required)

See attachments “Codebooks 1-3” for all measured variables and their detailed description. Codebook 1 covers overall information, codebook 2 is a detailing of general methodology as well as psychometric variables, and codebook 3 includes in-depth variables about practical implications.

Variables will be studied in the following categories:

- publication information,
- study information,
- self-efficacy scale characteristics,
- self-efficacy scale psychometrics,
- self-efficacy research models,
- sample description,
- context factors,
- security task,
- advice for others,
- practitioner models,
- user models,
- secure behavior suggestions, and
- trainings suggestions.

18. Indices (optional)

-

Analysis Plan

19. Statistical models (required)

See attachment “Synthesis Scheme” for a detailed assignment of variables to their research question and corresponding synthesis methods. We will use a specific set of synthesis methods to answer our research questions:

- Narrative synthesis of data
- Thematic synthesis of reported interventions
- Descriptive grouping of open-ended variables
- Numerical presentation of data
- Tabular presentation of data
- Graphing data
- Categorization of articles by type of IT device: smart home / not

To account for interrater agreement, we will use kappa coefficients (two indices: one for nominal and one for continuous variables). One key variable for each research question will be accessed to determine the level of agreement:

Research Question	Variable from Codebooks
1	sample_size
2	reliability_alpha_studypaper or reliability_alpha
3	as_outcome_variable
4	intervention

5	smarthome
6	context_adaption
7	task_type
8	advice_type
9	behavior_suggested
10	training_suggested

20. Transformations (optional)

For our coding scheme and description of each variable see attachments “Codebooks 1-3”.

21. Inference criteria (optional)

We intent to report all tests conducted. To make inferences regarding the iota coefficients, we will use the interpretation of Janson and Olsson (2001).

22. Data exclusion (optional)

In- or exclusion of studies will be determined by our selection criteria (see 12. Data collection procedures).

23. Missing data (optional)

Missing data will be marked with NA. Authors will not be contacted for unreported data.

24. Exploratory analysis (optional)

We expect there will be differences between studies focusing on smart home devices and those that inspect other smart information technologies. Therefore, we will look for differences and similarities between these two categories of studies on a synthesis level.

We are planning to do future work (using the same selection criteria) on whether there is consensus in IT-security research on the construct of cybersecurity self-efficacy. This project will be preregistered separately.

Other

25. Other (Optional)

The data of this review will be utilized in two reviews, each focusing on different research questions (all of which are stated in 4. Hypotheses), as well as another potential review regarding our exploratory question (see 24. Exploratory analysis).

Janson, H., & Olsson, U. (2001). A Measure of Agreement for Interval or Nominal Multivariate Observations. *Educational and Psychological Measurement*, 61(2), 277-289.

<https://doi.org/10.1177/00131640121971239>

PROSPERO REGISTRATION FORM: Cybersecurity Self-Efficacy

1. Review title:

The Study of Cybersecurity Self-Efficacy: A Systematic Literature Review of Methodology and Application

2. Original language title:

-

3. Anticipated or actual start date:

01 February 2021

4. Anticipated completion date:

01 June 2021

5. Stage of review at time of this submission:

Review Stage	Started	Completed
Preliminary searches	x	
Piloting of the study selection process		
Formal screening of search results against eligibility criteria		
Data extraction		
Risk of bias (quality) assessment		
Data analysis		

6. Named contact:

Mrs. Nele Borgert

7. Named contact email:

Nele.borgert@rub.de

8. Named contact address:

Ruhr-Universität Bochum
Fak. f. Psychologie
Mensch-Technik-Interaktion, GAFO 04/252
Universitätsstraße 150
44801 Bochum
Germany

9. Named contact phone number:

+49 234 32-24082

10. Organizational affiliation of the review:

¹ Psychology of Human Technology Interaction, Faculty of Psychology, Ruhr University Bochum

hti.ruhr-uni-bochum.de/

² Human Centered Security, Faculty of Electrical Engineering and Information Technology, Ruhr University Bochum

<https://www.hcs.ruhr-uni-bochum.de>

³ Horst Görtz Institute for IT Security, Ruhr University Bochum

<https://hgi.rub.de>

11. Review team members and their organizational affiliations:

Nele Borgert^{1,3}

Jennifer Friedauer^{2,3}

Imke Böse^{1,3}

Martina Angela Sasse^{2,3}

Malte Elson^{1,3}

12. Funding sources/sponsors

This work was supported by the German Federal Ministry of Education and Research (BMBF) under the grant 16SV8505 (UsableSec@Home project).

13. Conflict of interest:

None

14. Collaborators:

-

15. Review Questions:

Review A:

1. What are the demographics of studies of cybersecurity self-efficacy? (Including year of study, study type, sample groups, sample size, countries)
2. What measures are used to assess cybersecurity self-efficacy? What are the scale characteristics and reported psychometrics?
3. What role does cybersecurity self-efficacy play in the theoretical or research models of the studies? (Including cause or outcome)
4. Do the studies report interventions that have been carried out to manipulate cybersecurity self-efficacy? If so, what was their approach?
5. Are there studies of cybersecurity self-efficacy regarding smart home devices? If so:
 - a. What are the demographics of those studies? (Including year of study, study type, sample groups, sample size, and countries)
 - b. What measures are used to assess cybersecurity self-efficacy in those studies? What are the scale characteristics and reported psychometrics?
 - c. What role does cybersecurity self-efficacy play in the theoretical or research models of those studies? (Including cause or outcome)

d. Do those studies report interventions that have been carried out to manipulate cybersecurity self-efficacy? If so, what was their approach?

Review B:

[Research Question 1 – to describe study sample]

6. Do the used measures take the context in which cybersecurity self-efficacy is needed into account? (Impact of context)

a. Is the cybersecurity self-efficacy instrument adapted to a specific situation? (e.g. item text adapted to a scenario)

b. Is environmental impact measured in the study? (e.g. questions that ask for working conditions / conditions of use; behavior of colleagues or others who share the environment)

c. Are there items in the cybersecurity self-efficacy instrument that take emotional states into account? Or are emotions measured in an additional metric?

7. Is the security task in the conducted studies a primary or secondary task?

8. Do the conducted studies give advice to practitioners on how to improve cybersecurity self-efficacy? Who is addressed? (Organization, single position in organization, end user)

9. Do the addressed practitioners (organization, single position in organization, end user) need to be active? Do they need abilities or characteristics to implement the given advice? (User / practitioner model)

10. What behavior or what actions are suggested as an outcome of the conducted studies to improve cybersecurity self-efficacy in practice? (Types of training)

16. Searches:

Sources: EBSCOhost, IEEE Xplore, ACM Digital Library, Science Direct, dimensions.ai, arXiv, Scopus, Web of Science, Wiley Online Library

Search Dates: after preregistration

Restrictions: English

17. Search strategy:

See research terms document.

18. Condition or domain being studied:

Cybersecurity self-efficacy

19. Participants/population:

Inclusion: all

20. Interventions, exposures:

Inclusion: all (population, group, or individual level interventions)

21. Comparators/control:

-

22. Types of study to be included:

Inclusion: qualitative / quantitative

Exclusion: individual case studies ($N = 1$), none-empirical studies

23. Context:

Inclusion: all cultures/countries, setting must be cybersecurity ≥ 2010

Exclusion: if self-efficacy is not measured regarding IT security or privacy, the study will be excluded

24. Main outcomes:

All reported publication information, study information, self-efficacy scale characteristics, self-efficacy scale psychometrics, self-efficacy research models, sample description, context factors, security task, advice for others, practitioner models, user models, secure behavior suggestions, and training suggestions

Measures of effect: specified in codebooks 1-3

25. Additional outcomes:

See codebooks 1-3 for a detailed description of outcome measures.

26. Data extraction (selection and coding)

Study selection:

- Number of reviewers: 3
- Independently screening records for inclusion (each reviewer sifts one-third of the papers for inclusion [title and abstract sift])
- Disagreements of selection in all-text sift stage will be discussed in group (2/3 vote wins)
- Software system for recording decisions: Citavi and Excel

Data extraction:

- List of data that will be extracted is attached in codebook 1-3
- Number of people extracting: 3
- Blinded to each other's decisions (each reviewer extracts data from two-thirds of the papers)
- Disagreements will be discussed in group (2/3 vote wins)
- Missing data will be marked with NA, authors will not be contacted for unreported data
- Means of recording data: Citavi and Excel
- Software system for extracting data or data management: Citavi and Excel

27. Risk of bias (quality) assessment:

- In addition to the quality assessment inherent to this study no further characteristics will be assessed
- Codebooks 1-3 cover the assessment of information about study type and setting as well as reliability and validity of the scale
- Level of assessment: study (e.g., type) and results (e.g., reliability)
- Number of reviewers: 3
- Disagreements will be discussed in group (2/3 vote wins)

28. Strategy of data synthesis:

- Criteria under which the data will be synthesized:
 - minimum number of studies: 2
- The following data will be synthesized: publication information, study information, self-efficacy scale characteristics, self-efficacy scale psychometrics, self-efficacy research models, and sample description
- Methods for synthesizing data will be within a descriptive data synthesis (see synthesis scheme document):
 - Narrative synthesis of data
 - Thematic synthesis of reported interventions
 - Descriptive grouping of open-ended variables
 - Numerical presentation of data
 - Tabular presentation of data
 - Graphing data
 - Categorization of articles by type of IT device: smart home / not

29. Analysis of subgroups or subsets:

A subgroup synthesis is inherent to research question number 5.a. - 5.d. (in contrast to 1. - 4.). Users of smart home appliances might differ from users of other smart IT devices due to the technology's e.g., prevalence and primary tasks.

30. Type and method of review:

Type of review

- Intervention
- Methodology
- Narrative synthesis
- Systematic review

Area of review: Human-Technology Interaction

31. Language:

Englisch

32. Country:

Germany

33. Other registration details:

Other organizations where protocol is registered: OSF

Identification number: tba

Link to repository: extracted data will be stored and made available through OSF

34. Reference for published protocol:

If protocol is published provide details, add Link or upload protocol

35. Dissemination plans:

Yes, I intend to publish the review on completion.

Plans for communication review findings: research group website, project website, conferences, journal publication

36. Keywords:

Review; Self-Efficacy; Smart Home; Cybersecurity

37. Details of any existing review of the same topic by the same authors:

-

38. Current review status:

Ongoing

39. Any additional information:

This review is being undertaken as part of the planning for studies investigating cybersecurity self-efficacy and its impact on data security decisions regarding privacy settings of smart home devices.

40. Details of final report/publication or reprints if available:


Leave empty until publication details are available

RESEARCH TERMS: Cybersecurity Self-Efficacy

	Search term
EBSCOhost	AB "self-efficacy" AND AB ("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")
	Additional instruction not included in search term: Results filtered for publication date: 2010 - 2021
	Additional instruction not included in search term: choose databases = Academic Search Premier, APA PsycArticles, APA PsycInfo, Historical Abstracts, OpenDissertations, PSYNDEX Literature with PSYNDEX Tests
IEEE Xplore	("Abstract": "self-efficacy") AND ("Abstract": "cybersecurity" OR "Abstract": "cyber security" OR "Abstract": "information security" OR "Abstract": "IT security" OR "Abstract": "information technology security" OR "Abstract": "IS security" OR "Abstract": "information system security" OR "Abstract": "wireless security" OR "Abstract": "home wireless security" OR "Abstract": "usable security" OR "Abstract": "computer security" OR "Abstract": "data protection" OR "Abstract": "data security" OR "Abstract": "personal data" OR "Abstract": "privacy" OR "Abstract": "security threat" OR "Abstract": "wireless network" OR "Abstract": "device security")
	Additional instruction not included in search term: Results filtered for publication date: 2010 - 2021
ACM Digital Library	[Abstract: "self-efficacy"] AND [[Abstract: "cybersecurity"] OR [Abstract: "cyber security"] OR [Abstract: "information security"] OR [Abstract: "it security"] OR [Abstract: "information technology security"] OR [Abstract: "is security"] OR [Abstract: "information system security"] OR [Abstract: "wireless security"] OR [Abstract: "home wireless security"] OR [Abstract: "usable security"] OR [Abstract: "computer security"] OR [Abstract: "data protection"] OR [Abstract: "data security"] OR [Abstract: "personal data"] OR [Abstract: "privacy"] OR [Abstract: "security threat"] OR [Abstract: "wireless network"] OR [Abstract: "device security"]] AND [Publication Date: (01/01/2010 TO 12/31/2021)]
Science Direct	Title, abstract, keywords: "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")
	Title, abstract, keywords: "self-efficacy" AND ("home wireless security" OR "usable security" OR "computer security" OR "data protection" OR "data security" OR "personal data" OR "privacy")
	Title, abstract, keywords: "self-efficacy" AND ("security threat" OR "wireless network" OR "device security")
	Additional instruction not included in search term: Results filtered for publication date: 2010 - 2021
dimensions.ai	"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")
	Additional instruction not included in search term: Results filtered for publication date: 2010 - 2021
	Additional instruction not included in search term: Keyword search in title and abstract
arXiv	abstract="self-efficacy"; AND abstract="cybersecurity"
	abstract="self-efficacy"; AND abstract="cyber security"
	abstract="self-efficacy"; AND abstract="information security"
	abstract="self-efficacy"; AND abstract="IT security"
	abstract="self-efficacy"; AND abstract="information technology security"
	abstract="self-efficacy"; AND abstract="IS security"
	abstract="self-efficacy"; AND abstract="information system security"
	abstract="self-efficacy"; AND abstract="wireless security"
	abstract="self-efficacy"; AND abstract="home wireless security"
	abstract="self-efficacy"; AND abstract="usable security"
	abstract="self-efficacy"; AND abstract="computer security"

	abstract="self-efficacy"; AND abstract="data protection"
	abstract="self-efficacy"; AND abstract="data security"
	abstract="self-efficacy"; AND abstract="personal data"
	abstract="self-efficacy"; AND abstract="privacy"
	abstract="self-efficacy"; AND abstract="security threat"
	abstract="self-efficacy"; AND abstract="wireless network"
	abstract="self-efficacy"; AND abstract="device security"
	Additional instruction not included in search term: Results filtered for publication date: 2010 - 2021
Scopus	(TITLE-ABS-KEY ("self-efficacy") AND TITLE-ABS-KEY ("cybersecurity" OR "cyber AND security" OR "information AND security" OR "it AND security" OR "information AND technology AND security" OR "is AND security" OR "information AND system AND security" OR "wireless AND security" OR "home AND wireless AND security" OR "usable AND security" OR "computer AND security" OR "data AND protection" OR "data AND security" OR "personal AND data" OR "privacy" OR "security AND threat" OR "wireless AND network" OR "device AND security")) AND PUBYEAR > 2009
Web of Science	AB=("self-efficacy") AND AB=("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")
	Additional instruction not included in search term: Databases= WOS, KJD, MEDLINE, RSCI, SCIELO Timespan=2010-2021 Search language=Auto
Wiley Online Library	("self-efficacy") in Abstract 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") in Abstract



Additional instruction not included in search term: Results filtered for publication date: 2010 - 2021


```

# -*- coding: utf-8 -*-
"""ebshost_search_improvement.ipynb"""

import pandas as pd
import os

data_dataset = pd.read_csv(r'dataset.csv')
display(data_dataset)

new_dataframe = data_dataset.copy()
var1 = "cybersecurity"
var2 = "cyber security"
var3 = "information security"
var4 = "IT security"
var5 = "information technology security"
var6 = "IS security"
var7 = "information system security"
var8 = "wireless security"
var9 = "home wireless security"
var10 = "usable security"
var11 = "computer security"
var12 = "data protection"
var13 = "data security"
var14 = "personal data"
var15 = "privacy"
var16 = "security threat"
var17 = "wireless network"
var18 = "device security"
removed = 0
removed_index = []
for index, row in data_dataset.iterrows():
    #print(row["Abstract"])
    abstract = row["Abstract"]
    abstract = abstract.lower() #to also find research terms written with
capital letters
    abstract = abstract.replace("-", " ") #to also find research terms written
with hyphens instead of spaces
    final = False
    firsttest = abstract.find("self efficacy")
    #print(firsttest)
    if firsttest != -1:
        print("self-efficacy in Text")
        secondtest1 = abstract.find(var1)
        secondtest2 = abstract.find(var2)
        secondtest3 = abstract.find(var3)
        secondtest4 = abstract.find(var4)
        secondtest5 = abstract.find(var5)
        secondtest6 = abstract.find(var6)
        secondtest7 = abstract.find(var7)
        secondtest8 = abstract.find(var8)
        secondtest9 = abstract.find(var9)
        secondtest10 = abstract.find(var10)
        secondtest11 = abstract.find(var11)
        secondtest12 = abstract.find(var12)

```

```

secondtest13 = abstract.find(var13)
secondtest14 = abstract.find(var14)
secondtest15 = abstract.find(var15)
secondtest16 = abstract.find(var16)
secondtest17 = abstract.find(var17)
secondtest18 = abstract.find(var18)
if secondtest1 !=-1 or secondtest2 !=-1 or secondtest3 !=-1 or
secondtest4 !=-1 or secondtest5 !=-1 or secondtest6 !=-1 or secondtest7 !=-1 or
secondtest8 !=-1 or secondtest9 !=-1 or secondtest10 !=-1 or secondtest11 !=-1
or secondtest12 !=-1 or secondtest13 !=-1 or secondtest14 !=-1 or secondtest15
!=-1 or secondtest16 !=-1 or secondtest17 !=-1 or secondtest18 !=-1:
    print("OR Abfrage bestanden")
    print(secondtest1, secondtest2, secondtest3, secondtest4,
secondtest5, secondtest6, secondtest7, secondtest8, secondtest9, secondtest10,
secondtest11, secondtest12, secondtest13, secondtest14, secondtest15,
secondtest16, secondtest17, secondtest18)
    final = True
else:
    print("OR Abfrage NICHT bestanden")
    final = False
else:
    print("self-efficacy NICHT im Text")
    final = False
print(final)
if not final:
    new_dataframe.drop([index], inplace=True)
    removed = removed + 1
    removed_index.append(index)
    print('Remove row #', index)
print('# Articles removed:', removed)
print('The following articles have been removed:', removed_index)
#new_dataframe.to_csv('cleaned_articles.csv')
new_dataframe.to_excel('cleaned_articles.xlsx')

display(new_dataframe)

```

CODEBOOK 1 – Overall Information: Cybersecurity Self-Efficacy

VARIABLE NAME	VALUES	DESCRIPTION
for all variables	NA = Missing / not reported / not applicable	Coding of missing, not reported or not applicable data
PUBLICATION INFORMATION		
id	Consecutive from 1	Numerical identifier for each publication. Matches variable id in all other datasets
publication_type	0 = Journal article 1 = Conference proceedings 2 = Conference paper 3 = Conference poster 4 = Book chapter 5 = Dissertation 6 = Report	Type of publication. Type of dissertation specified in <journal_book_conf>
authors	format: [Last name] et al.	Authors of publication
year	2010 - 2021	Year of publication
title		Title of publication
editors	format: [Last name] et al.	Names of editors of book or proceedings study was published in
journal_book_conf		Full title of journal, book, conference, proceedings, URL study was published in. Specifies type of dissertation when <pub_type> = 5
volume		Volume of journal (only if <pub_type> = 0)
issue		Issue of volume (only if <pub_type> = 0)
pages		Pages of publication
publisher		If <pub_type> = 1 4: Location and name of publisher If <pub_type> = 2 3: Location of conference If <pub_type> = 5: Name and location of university If <pub_type> = 6: Name and location of issuing institution
doi		Digital Object Identifier of publication, where available
isbn		International Standard Book Number, where available
citations	format: [number (source)]	Number of citations and the source for that information. For source try google scholar first.
notes		Additional relevant information about the test or publication. Always note reasons why studies were not included.
STUDY INFORMATION		
exclude	0 = excluded from synthesis 1 = included in synthesis	Studies marked with <exclude> = 0 were not considered for any of our synthesis (see notes for reasons).

study_samples		Numeration for studies (as in the original publication) in multi-study papers; alphabetical numeration for samples (assigned by us, see <notes>) in publications that report separate reliability scores for multiple (sub-)samples
study_type	0 = experiment 1 = quasi-experiment 2 = survey 3 = other	Type of study: experiment (= randomization & manipulation, including interventions with pre/post design); quasi-experiment (at least one experimental condition with nonrandom assignment); survey; other (e.g. longitudinal study (survey or experiment), content analysis, observational study, interview study)
study_setting	0 = online 1 = physical	Was the study conducted online or on-site (i.e. physical)?
smarthome	0 = no 1 = yes	Is the information technology used / focused in the study a smart home device? Smart home devices = technical processes and systems in living spaces and houses, which focus on increasing the quality of living and life, safety and efficient energy use on the basis of networked and remotely controllable devices and installations as well as automatable processes (e.g. automatically controlled heating, ventilation, doors, windows, awnings, blinds and lamps (building or home automation) as well as manually via mobile devices / smartphones controllable and manipulable systems, intelligent refrigerators and coffee machines (household appliance automation))
SCALE CHARACTERISTICS		
scale_origin	0 = developed by others 1 = self-developed	Is the scale developed by other authors (referenced original scale with prior development) or is it self-developed within that publication (new development / modifications/translations of original scale)?
scale_number	Consecutive from 1[letter in alphabetical order if modified version]	Numeration for original scales in the scale paper codebook (2. codebook). We refer to the other codebook, where each original scale is coded. If a scale is modified (e.g. wording is different), the scale number will be followed by an alphabetical numeration for the modification.
SCALE PSYCHOMETRICS (only if <scale_origin> = 0; for <scale_origin> = 1 use 2. codebook)		
reliability_alpha_studypaper	0 - 1	The reported Cronbach's alpha
reliability_composite_studypaper	0 - 1	The reported composite reliability
reliability_other_studypaper	format: [measure]: [symbol] = [result]	The reported type of reliability measure (e.g. test-retest with Pearson correlation coefficient r (type) or split-half reliability) and its result
validity_information_studypaper	format: [validity type]: [variable, used statistical measure/test, and findings];	If reported (i.e. it is sufficient when they themselves speak of validity), we describe supporting validity information. Validity types: content, construct factor, construct discriminant, construct convergent, criterion, incremental.
SE RESEARCH MODEL		

as_outcome_variable	0 = no 1 = yes	Is self-efficacy measured / constructed as an outcome variable in the research model?
cause_variables	format: [variable];	If <as_outcome_variable> = 1: We list all measured cause variables to self-efficacy.
as_cause_variable	0 = no 1 = yes	Is self-efficacy measured / constructed as a cause variable in the research model?
outcome_variables	format: [variable];	If <as_cause_variable> = 1: We list all measured outcome variables of self-efficacy.
intervention	0 = no 1 = yes	Was there a use of an intervention that was designed to explicitly influence self-efficacy?
intervention_description		If <intervention> = 1: We describe the intervention.
SAMPLE DESCRIPTION		
sample_size	> 0	Reported sample size
sample_age	format: [mean]([SD])	Reported mean age and its SD of the sample or the reported age range
sample_sex		If reported female, male, and other percentages
sample_profession	format: [variable];	Reported profession of the sample (e.g. students; software developers; ...)
sample_recruitment	0 = online panels 1 = ad-hoc 2 = organizations 3 = mixed	Was the sample acquired via online panels (e.g. MTurk), ad-hoc (population is unknown), organizations (e.g. a certain business, university) or was the acquisition a mix of those strategies?
sample_country		Full name of the country of origin of the sample
CONTEXT FACTORS (in statistic measures)		
context_adaption	0 = no 1 = yes	Is the self-efficacy instrument adapted to a special context? (The item text refers to a specific situation, e.g. all items adapted to one scenario)
context_adaption_description		If <context_situation> = 1: We shortly describe the situation. e.g. "smart phone use for observing the baby's room" or e.g. "receiving phishing mail at work"
context_environment	0 = no 1 = yes	Is environmental impact measured in the study? (Questions that ask for working conditions / conditions of use; behaviour of colleagues or others who share the environment) e.g. "How many emails do you receive every day?" or "Can the others do it?"
context_environment_type	0 = conditions 1 = behaviour of others 2 = other	If <context_environment> = 1: What does environment mean? Conditions (influencing factors that are non-human): e.g. structure, processes, goals, time for task Behaviour of others: e.g. colleagues, family members

context_emotions	0 = no, not at all 1 = yes, in the self-efficacy instrument 2 = yes, in (an) additional instrument(s)	Are there items that take emotional states into account? Emotional states are for example: affective states, basic emotions, and stress. e.g. in self-efficacy instrument: “Even if I have a bad day, I know how to handle this” e.g. additional instrument to measure stress / PANAS
context_emotions_instrument		If <context_emotion> = 2 : What emotion is measured additionally? (e.g. stress) Is there a correlation between self-efficacy and the measured emotion?
SECURITY TASK		
task_type	0 = primary 1 = secondary	Is the security task the primary task? e.g. primary: the task is to encrypt an email e.g. secondary: the task is to send an email (therefore encryption is only a secondary task)
task_description		If <task_type> = 0/1 : What is the task (if reported both, primary and secondary).
ADVICE FOR OTHERS (for <advice_type> = 2 or 3 use 3. codebook)		
advice_type	0 = no advice 1 = future work 2 = for practitioners 3 = for end users 4 = others 5 = not clear	Is there any advice for others as an outcome of the presented work? For practitioners, for future research or others. Practitioner implements a mechanism to make processes secure (Organizations, e.g. Bank uses special authentication mechanism) End User uses the mechanism to reach primary goal (e.g. has to authenticate to get access to bank account). It is not clear if it fits more than one option, e.g. a practitioner can be a user (e.g. implementing for own safety) and a user can be a practitioner (e.g. parents implementing a mechanism on their children’s device)
advice_SE	0 = no 1 = yes	Is self-efficacy mentioned in this advice?

CODEBOOK 2 – Methodology & Psychometrics: Cybersecurity Self-Efficacy

VARIABLE NAME	VALUES	DESCRIPTION
for all variables	NA = Missing / not reported / not applicable	Coding of missing, not reported or not applicable data
PUBLICATION INFORMATION		
id	Consecutive from 1	Numerical identifier for each publication, which is coded in the study codebook. Matches variable in 1. codebook.
authors	format: [Last name] et al.	Authors of publication. Matches variable in 1. codebook.
year	2010 - 2021	Year of publication. Matches variable in 1. codebook.
SCALE CHARACTERISTICS		
scale_number	Consecutive from 1 [letter in alphabetical order if modified/translated version]	Numeration for original scales. If a scale is modified (e.g. wording/language is different), the scale number will be followed by an alphabetical numeration for the modification. Matches variable in 1. codebook.
scale_development	0 = ad-hoc test development 1 = modified test version 2 = translated test version 3 = validating test development	Is the scale/test developed on the spot as part of an empirical paper (ad-hoc); is it a modified version of an original scale; is it a translation of an original scale; is it rather developed in a psychometrical fashion for its own purpose of validation research?
scale_authors	format: [Last name] et al. ([year])	If <scale_development> = 1 2: We state the authors and year of the referenced scale source
scale_changes		If <scale_development> = 1 2: We describe the changes made to the scale
scale_name		Name of the scale / test used to assess self-efficacy
scale_language		Language of the scale
item_number	> 0	Number of items
factors	> 0	Number of factors
facets	format: [facet];	Names of the facets
item_list	format: [consecutive number]. [item text];	If reported, list of the actual items
SCALE PSYCHOMETRICS		
reliability_alpha	0 - 1	The reported Cronbach's alpha
reliability_composite	0 - 1	The reported composite reliability
reliability_testretest	-1 - 1	The reported test-retest reliability
reliability_splithalf	-1 - 1	The reported split-half reliability
validity_content	0 = not reported 1 = reported	Whether items are really suitable to capture the self-efficacy construct. As the content validity of a scale is determined by its final list of items, performing

		good practices during the construction of this list is essential. We report whether there is reporting of the construction process.
validity_content_description		To evaluate content validity, we summarize the item construction process of each scale.
validity_construct_factor	0 = EFA 1 = PCA 2 = CFA 3 = SEM	Whether the self-efficacy construct is captured and no other. The internal structure / factor validity can be measured via EFA, PCA, CFA, or more generally SEM (reliability tests cannot be considered a substitute for factor analyses in evaluating structural validity). We report the type of factor analysis that has been used in exploring the factor structure of a scale.
validity_construct_factor_description		We describe the findings from factor analyses that were conducted with the scale.
validity_construct_factor_fitindices	format: <code>[[fit index]] = [[value]];</code>	If <validity_construct_factor> = 2: We include any fit indices (if CFA: RMSEA, SRMR, CFI, TLI) reported for the final structural model of the scale.
validity_construct_discriminant_type	0 = correlation 1 = MTMM 2 = CFA	Whether the self-efficacy construct is captured and no other. Discriminant validity with unrelated tests (correlations, multi-trait-multimethod-matrix, CFA).
validity_construct_discriminant_description	format: <code>[[variable]]: [[findings]];</code>	We summarize the findings of the discriminant validity.
validity_construct_convergent_type	0 = correlation 1 = MTMM 2 = CFA	Whether the self-efficacy construct is captured and no other. Convergent validity with related tests (correlations, multi-trait-multimethod-matrix, CFA).
validity_construct_convergent_description	format: <code>[[variable]]: [[findings]];</code>	We summarize the findings of the convergent validity.
validity_criterion_type	0 = retrospective 1 = competitive 2 = predictive	Retrospective: current measurement of self-efficacy correlating with measurements in the past (of the outcome variable); Competitive: current measurement of self-efficacy correlating with other current measurements (of the outcome variable); Predictive: current measurement of self-efficacy correlating with measurements made later (of the outcome variable).
validity_criterion_results	format: <code>[[variable]]: r = [[value from -1 - 1[[*]]]];</code>	How well results of other tests / behaviors can be predicted by the test result (correlation with the external criterion).
validity_incremental	format: <code>[[variable]]: r = [[value from -1 - 1[[*]]]];</code>	The incremental validity of the new scale is supported by a significant association between a new scale and its related outcomes while statistically controlling for scores on another well-validated measure of self-efficacy.

CODEBOOK 3 – Practical Implications: Cybersecurity Self-Efficacy

VARIABLE NAME	VALUES	DESCRIPTION
for all variables	NA = Missing / not reported / not applicable	Coding of missing, not reported or not applicable data
PUBLICATION INFORMATION		
id	Consecutive from 1	Numerical identifier for each publication. Matches variable id in all other datasets
ADVICE FOR OTHERS		
advice_type	0 = no advice 1 = future work 2 = for practitioners 3 = for end users 4 = others 5 = not clear	<p>Is there any advice for others as an outcome of the presented work? For practitioners, for future research or others.</p> <p>Practitioner implements a mechanism to make processes secure (Organizations, e.g. Bank uses special authentication mechanism) End User uses the mechanism to reach primary goal (e.g. has to authenticate to get access to bank account).</p> <p>It is not clear if it fits more than one option, e.g. a practitioner can be a user (e.g. implementing for own safety) and a user can be a practitioner (e.g. parents implementing a mechanism on their children's device)</p> <p>Matches variable in 1. codebook.</p>
advice_SE	0 = no 1 = yes	Is self-efficacy mentioned in this advice? Matches variable in 1. codebook.
PRACTITIONER MODEL (If <advice_type> = 2)		
practitioner_active	0 = no 1 = yes	<p>The practitioners (organizations, groups in organizations, or person in organization) have to actively change their behavior = change processes / organizational learning</p> <p>Examples:</p> <ul style="list-style-type: none"> - implement trainings - invest into feedback and support services - change organizational structure - make processes transparent - learning about the user / group of users

practitioner_type	0 = organization 1 = group in Organization 2 = person in Organization 3 = not clear	Who is addressed? It is not clear if it fits more than one option, e.g. a practitioner can be a user (e.g. implementing for own safety) and a user can be a practitioner (e.g. parents implementing a mechanism on their children's device)
practitioner_characteristics	0 = no 1 = yes	Are there practitioner characteristics mentioned which are necessary for behavior change? Depends on practitioner_type e.g. [2] needs knowledge, abilities e.g. [0 or 1] needs culture, structure, special people
USER MODEL (If <advice_type> = 3)		
user_active	0 = no 1 = yes	The end user is the one who has to change their behavior = change habits / learning Examples: - learn about functions - learn about who to ask - get to know about what data is collected
user_characteristics	0 = no 1 = yes	Are there user characteristics mentioned which are necessary for behavior change? (e.g. knowledge, abilities)
SECURE BEHAVIOUR SUGGESTIONS		
behavior_suggested	0 = no 1 = yes	Is there any behavior suggestion as an outcome of the advice?
behavior_description	0 = no 1 = yes	If behaviour_suggested = 1: Is there a description of how practitioners or users can implement that behavior? E.g. not only does the study state to improve one's knowledge but also describes how (read the manual, go to class etc.)
behavior_goal	0 = no 1 = yes	If behaviour_suggested = 1: Has the suggested behavior a particular goal that is mentioned? E.g. raising self-efficacy, raising awareness, increase knowledge.
TRAINING SUGGESTIONS		
training_suggested	0 = no 1 = yes	Is there any training suggested?
training_goal	0 = train individual	If training_suggested = 1: Who should be trained?

	1 = train group in organization 2 = train organization 3 = other 4 = not clear	It is not clear if it fits more than one option.
training_type_awareness	0 = no 1 = yes	If <training_suggested> = 1: Is the suggested mechanism a training to raise awareness?
training_type_knowledge	0 = no 1 = yes	... or is it to increase knowledge?
training_type_behavior	0 = no 1 = yes	... or is it to practice and actively change behavior?
training_type_other		If not awareness, knowledge, or behavior, we describe the training type.
training_description	0 = no 1 = yes	If <training_suggested> = 1: Is there a description of how this training works? E.g. process description, what steps it takes to reach a goal

Synthesis Scheme for Codebooks 1-3 / Research Questions 1-10

Research Questions	Variables from Codebooks	Analyses / Methods
Differentiation for RQ 1.-4. and 5.	smarthome	Categorization
1. What are the demographics of studies of cybersecurity self-efficacy? (Including year of study, study type, sample groups, sample size, countries)	year study_type study_setting sample_size sample_age sample_sex sample_profession sample_recruitment sample_country	Narrative synthesis of data Descriptive grouping of open-ended variables Numerical presentation of data Tabular presentation of data Graphing data
2. What measures are used to assess cybersecurity self-efficacy? What are the scale characteristics and reported psychometrics?	scale_origin scale_number reliability_alpha_studypaper reliability_composite_studypaper reliability_other_studypaper validity_information_studypaper authors year citations scale_development scale_authors scale_changes scale_name scale_language item_number factors facets item_list reliability_alpha reliability_composite reliability_testretest reliability_splithalf validity_content validity_content_description validity_construct_factor validity_construct_factor_description validity_construct_factor_fitindices validity_construct_discriminant_type validity_construct_discriminant_description validity_construct_convergent_type validity_construct_convergent_description validity_criterion_type validity_criterion_results validity_incremental	Narrative synthesis of data Descriptive grouping of open-ended variables Numerical presentation of data Tabular presentation of data Graphing data
3. What role does cybersecurity self-efficacy play in the theoretical or research models of the studies? (Including cause or outcome)	as_outcome_variable cause_variables as_cause_variable outcome_variables	Narrative synthesis of data Descriptive grouping of open-ended variables Numerical presentation of data Tabular presentation of data Graphing data
4. Do the studies report interventions that have been carried out to manipulate cybersecurity self-efficacy? If so, what was their approach?	intervention intervention_description	Narrative synthesis of data Numerical presentation of data Narrative synthesis of data Thematic synthesis of interventions
5.a. Are there studies of cybersecurity self-efficacy regarding smart home devices? If so: What are the demographics of those studies? (Including year of study, study type, sample groups, sample size, and countries)	year study_type study_setting sample_size sample_age sample_sex sample_profession sample_recruitment sample_country	Narrative synthesis of data Descriptive grouping of open-ended variables Numerical presentation of data Tabular presentation of data Graphing data
5.b. Are there studies of cybersecurity self-efficacy regarding smart home devices? If so: What measures are used to assess cybersecurity self-efficacy in those studies?	scale_origin scale_number reliability_alpha_studypaper	

What are the scale characteristics and reported psychometrics?	reliability_composite_studypaper reliability_other_studypaper validity_information_studypaper authors year citations scale_development scale_authors scale_changes scale_name scale_language item_number factors facets item_list reliability_alpha reliability_composite reliability_testretest reliability_splithalf validity_content validity_content_description validity_construct_factor validity_construct_factor_description validity_construct_factor_fitindices validity_construct_discriminant_type validity_construct_discriminant_description validity_construct_convergent_type validity_construct_convergent_description validity_criterion_type validity_criterion_results validity_incremental	Narrative synthesis of data Descriptive grouping of open-ended variables Numerical presentation of data Tabular presentation of data Graphing data
5.c. Are there studies of cybersecurity self-efficacy regarding smart home devices? If so: What role does cybersecurity self-efficacy play in the theoretical or research models of those studies? (Including cause or outcome)	as_outcome_variable cause_variables as_cause_variable outcome_variables	Narrative synthesis of data Descriptive grouping of open-ended variables Numerical presentation of data Tabular presentation of data Graphing data
5.d. Are there studies of cybersecurity self-efficacy regarding smart home devices? If so: Do those studies report interventions that have been carried out to manipulate cybersecurity self-efficacy? If so, what was their approach?	intervention intervention_description	Narrative synthesis of data Numerical presentation of data Narrative synthesis of data Thematic synthesis of interventions
6. Do the used measures take the context in which cybersecurity self-efficacy is needed into account? (Impact of context) a. Is the cybersecurity self-efficacy instrument adapted to a specific situation? (e.g. item text adapted to a scenario)	context_adaption context_adaption_description	Numerical presentation of data Tabular presentation of data Graphing data Narrative synthesis of data Thematic synthesis
6. Do the used measures take the context in which cybersecurity self-efficacy is needed into account? (Impact of context) b. Is environmental impact measured in the study? (e.g. questions that ask for working conditions / conditions of use; behavior of colleagues or others who share the environment)	context_environment context_environment_type	Numerical presentation of data Tabular presentation of data Graphing data Narrative synthesis of data
6. Do the used measures take the context in which cybersecurity self-efficacy is needed into account? (Impact of context) c. Are there items in the cybersecurity self-efficacy instrument that take emotional states into account? Or are emotions measured in an additional metric?	context_emotions context_emotions_instrument	Numerical presentation of data Tabular presentation of data Graphing data Narrative synthesis of data
7. Is the security task in the conducted studies a primary or secondary task?	task_type task_description	Numerical presentation of data Tabular presentation of data Graphing data Narrative synthesis of data

8. Do the conducted studies give advice to practitioners on how to improve cybersecurity self-efficacy? Who is addressed? (Organization, single position in organization, end user)	<div>advice_type</div> <div></div> <div>advice_SE</div>	Numerical presentation of data Tabular presentation of data Graphing data Narrative synthesis of data
9. Do the addressed practitioners (organization, single position in organization, end user) need to be active? Do they need abilities or characteristics to implement the given advice? (User / practitioner model)	<div>practitioner_active</div> <div>practitioner_type</div> <div>practitioner_characteristics</div> <div>user_active</div> <div>user_characteristics</div> <div>behavior_suggested</div>	Numerical presentation of data Tabular presentation of data Graphing data Narrative synthesis of data Thematic synthesis
10. What behavior or what actions are suggested as an outcome of the conducted studies to improve cybersecurity self-efficacy in practice? (Types of training)	<div>behavior_suggested</div> <div>behavior_description</div> <div>behavior_goal</div> <div>training_suggested</div> <div>training_goal</div> <div>training_type_awareness</div> <div>training_type_knowledge</div> <div>training_type_behavior</div> <div>training_type_other</div> <div>training_description</div>	Numerical presentation of data Tabular presentation of data Graphing data Narrative synthesis of data Thematic synthesis