Computers & Security 150 (2025) 104206

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

Full Length Article

Assessing the effect of cybersecurity training on End-users: A Meta-analysis

Julia Prümmer *, Tommy van Steen, Bibi van den Berg

Institute of Security and Global Affairs, Faculty of Governance and Global Affairs, Leiden University, The Netherlands

ARTICLE INFO

Keywords:
Meta-Analysis
Cybersecurity
Behaviour Change
Training
end-users

ABSTRACT

Cybersecurity behaviour of end-users continues to be a growing topic of conversation, both in organisations and in academia, as end-users are often said to be the last line of defence against cyberattacks. Unfortunately, end-users are often not aware that they engage in risky cyber behaviours and can, in turn, make themselves and the organisations that they work for vulnerable. Attempting to change end-user behaviour through training programs has become common practice in many organisations, a trend that is reflected in the academic literature as well. While a variety of literature reviews on the topic are available, an assessment of the effectiveness of these training programs through a meta-analysis has so far not been conducted. We carried out a meta-analysis based on a systematic literature review on the topic and an updated literature search in order to assess the overall effectiveness of cybersecurity training programs. We identified 69 studies that were eligible for inclusion.

Our analysis shows that training overall has a positive effect on end-users ($d = 0.75$, 95%CI [0.58, 0.92]), particularly when assessing predictors of behaviour such as attitudes or knowledge ($d = 1.02$, 95%CI [0.58, 1.46]). Interestingly, studies assessing changes in behaviour are not able to match these results ($d = 0.36$, 95%CI [-0.09, 0.80]), showcasing a clear inability of current training approaches to change behaviour. The effect sizes obtained in this meta-analysis can act as smallest effect sizes of interest (SESOIs) for future research on end-user cybersecurity training. Further findings with regards to the effectiveness of individual training methods and other moderators are discussed.

1. Introduction

The issue of cybersecurity has gained significant traction in recent years and has become a topic of conversation in many organisations. From large-scale attacks to minor phishing scams, companies are often faced with the responsibility of preparing their employees for such attacks (National Cybersecurity Alliance and CYBSAFE, 2023). One way to improve end-user behaviour is through comprehensive training programs (Chowdhury and Gkioulos, 2021; Coenraad et al., 2020; Prümmer et al., 2024).

Academic literature has concerned itself with cybersecurity training for a number of years now, leading to an ever-growing body of work. As outlined through a literature review conducted by the authors, certain trends among existing cybersecurity training programs can be identified (Prümmer et al., 2024). Through this literature review, we have identified that while a large variety of cybersecurity behaviours are addressed in training programs, some appear much more frequently in the literature than others. Social engineering, and phishing in particular, is often chosen as the training focus (e.g. Gordon et al., 2019), while insider threat (e.g. Carlson, 2020) and malware issues (e.g. Ikhalia et al., 2019) are addressed much less frequently. Many articles also focus on cybersecurity in general, rather than targeting a specific topic (e.g. Hepp et al., 2018; Kletenik et al., 2021). Methods of training delivery vary as well, from 'traditional' methods of training, where training material is communicated to end-users through presentations, videos or text-based methods, to more 'modern' approaches such as gamification or simulations. In particular, the recent popularity of serious gaming in educational circles is evident in the cybersecurity literature (Chen et al., 2020; van Steen and Deeleman, 2021). In the literature review we identified game-based training methods as the most commonly used training method, appearing in 38 out of 142 articles. A similar finding was made in the literature review conducted by Chowdhury and Gkioulos (2021), who identified simulation and virtualisation exercises as popular training tools in cybersecurity. Examples of these more modern methods include the game 'Anti-Phishing Phil', developed by Sheng et al. (2007) and used in Abawajy (2014) and Mayhorn and Nyeste (2011), as well as a simulation run in Loffler et al. (2021). The game 'Anti-Phishing Phil' involves players distinguishing legitimate

* Corresponding author at: Leiden University, The Netherlands, Turfmarkt 99 Room 4.03, 2511 DP The Hague, The Netherlands.
E-mail address: j.prummer@fgga.leidenuniv.nl (J. Prümmer).

https://doi.org/10.1016/j.cose.2024.104206
Received 20 September 2024; Received in revised form 30 October 2024; Accepted 7 November 2024
Available online 30 November 2024

URLs from fake ones by having the protagonist, a fish named Phil, consume legitimate URLs in order to grow. In the simulation employed by Loffler et al. (2021), players were asked to investigate misconduct by a fellow employee through clues left behind in an escape room-style intervention. As mentioned previously, more 'traditional' methods of training such as presentation-based approaches can also be found in the literature (Abraham and Chengalur-Smith, 2019; Puhakainen and Siponen, 2010).

The evaluation of the effectiveness of training plays an integral role in determining which methods can be used to change end-user behaviour. Many articles that evaluate objective behaviour change do so with regard to phishing (Baillon et al., 2019; B. Kim et al., 2020). These objective measures of behavioural change are easily implemented in phishing research, as researchers can simply count the number of end-users who clicked on a link contained in a phishing e-mail, or even entered personal information when requested. Other cybersecurity behaviours are more challenging to measure objectively and have received less attention in that respect. For these behaviours, articles describe alternative measures of success, such as improvements in attitude, intention, or knowledge (Wu et al., 2021).

While individual studies on cybersecurity training report overwhelmingly positive results, a quantitative synthesis of the effectiveness of cybersecurity training for end-users by means of a meta-analysis has so far not been conducted. Filling this gap in the literature is the goal of the present study. Consequently, this research is largely a continuation and extension of the literature review previously conducted by the authors. Narrative systematic literature reviews are informative, but quantifying success of training programs through a meta-analysis adds additional layers of information that we cannot get from individual studies and literature reviews alone. For instance, conducting a meta-analysis can shine light on whether all training methods are effective, and whether differences in effectiveness between training methods exist. Furthermore, we can investigate whether the strength of the found successes changes if precursors to behaviour are measured instead of behaviour. By conducting this meta-analysis, we are also attempting to set a baseline for what a good cybersecurity training should achieve in order to be deemed successful. By establishing an overall effect size for cybersecurity training, as well as effect sizes for a variety of subcategories, we have the ability to set up a so-called 'smallest effect size of interest' (SESOI) (Anvari and Lakens, 2021). The SESOI is the smallest effect size at which results obtained through, for example, an intervention are considered meaningful. Consequently, even if an intervention study leads to statistically significant results, these results may not be deemed to be practically relevant if they are not equal to or larger than the SESOI. Overall, it acts as an anchor for what all training programs in the future should achieve. Therefore, we conducted a meta-analysis on the effects of training of end-users on cybersecurity outcomes both generally, but also with regard to the effect of training on behaviour specifically, as well as the effect of nuances in the training design and implementation.

2. Methods

2.1. Search strategy

A systematic search of the literature was conducted in November 2021 using the following search terms:

1. cyber* OR "cyber security" OR "information security" OR "digital security" OR "computer security" OR "social engineering" OR "IT security"
2. AND intervention* OR training OR awareness OR game* OR gamification

Databases used for this search were Web of Science, ACM Digital Library, ProQuest, PubMed and PsycINFO. These databases were chosen due to the spectrum of journals included within them (e.g. journals of information security, computer science, human-computer interaction, security studies, education and others). Searches were re-run in August 2023 to include more recent publications, since a significant amount of time had passed since the initial search was conducted. Grey literature was covered by including the ProQuest Dissertations and Theses Sub-Database in our search. The initial search in November 2021 yielded 16,771 results, while the updated search led to an additional 4390 results, leading to a total of 21,161 articles. The results obtained from the initial literature search were utilised prior to this meta-analysis to conduct a systematic literature review (Prümmer et al., 2024). For this reason, the authors chose to not screen the results from the initial literature search a second time, and instead chose to use the 142 articles included in the literature review as the basis from which further screenings for this study were conducted.

2.2. Inclusion and exclusion criteria

Studies were eligible if they reported quantitative empirical data. Due to the limited number of randomised controlled trials in the selected sources, both randomised and non-randomised, as well as between- and within-subjects study designs were deemed eligible. Articles were excluded if the pre-test and post-test conducted in within-subjects designs did not assess the same outcome measure, or assessed the same outcome measures differently so that no comparison could be made. In addition, papers were included if training of end-users in relation to digital technologies was implemented within them. More specifically, training programs that focused on cyber-dependent crimes aimed at end-users (e.g. exploitations through phishing, lack of screen-locking or suboptimal password behaviour) were included. We thereby excluded training programs aimed at behaviours where digital technologies are a means of interaction. This included training programs aimed at topics such as sexting or cyberbullying. Lastly, studies were excluded if no translation to English could be found.

2.3. Study selection and data extraction

By removing duplicate records, the total number of articles obtained during the August 2023 search was reduced from 4390 to 3300. After an initial title and abstract screening carried out by two authors (JP & TvS), the number of articles deemed eligible was reduced to 177. At the title and abstract screening stage in the selection procedure, any article that was deemed relevant by one author was included in further screenings, meaning that disagreements were not discussed. These articles, as well as the 142 articles identified in the initial screening from November 2021 were combined and screened by the same two authors using the inclusion and exclusion criteria outlined previously. Here, disagreements were discussed until a consensus was reached. After the screenings were concluded, 55 articles remained. Forty-three of these articles were generated from the November 2021 search, while 12 were added through the updated search conducted in August 2023.

Data was extracted by one author (JP). Relevant data included sample sizes, as well as data used to calculate effect sizes such as means and standard deviations, binary proportions or correlations. If this data was not available, the articles were searched for alternative calculation mechanisms such as p-values, standard errors or F-tests. In cases where mean scores or standard deviations were reported for each item on a scale, the scores were transformed into a single mean and standard deviation by averaging the means and pooling the standard deviations of each item. If the sought data could not be identified in the article, relevant authors were contacted via e-mail and asked to provide the missing data. If multiple outcome measures were assessed in an article, all relevant data was extracted separately for each outcome measure. In addition, further information needed for potential moderator analyses, including training methods, platform of the training (online vs. offline) and social setting were extracted. Due to a variety of study designs, no

risk of bias assessment tool was deemed appropriate and therefore, no risk of bias assessment was conducted.

2.4. Meta-analysis

In order to prevent the inclusion of data from the same participants multiple times for a single study, we decided to first combine all independent subgroups within a study that share the same control group into a variable for each outcome measure that was assessed in the study. Then, the same outcome measures (if assessed through multiple scores) were combined into one variable for each outcome measure. Afterwards, all separate outcome measures were collated into one variable, representing the study overall. Effect size and variance calculations for two or more outcome measures were made using formulas proposed by Borenstein et al. (2021) (see below). Correlations required in these formulas were kept constant across different calculations at 0.5, as recommended by Higgins and Green (2008) in the Cochrane Handbook for Systematic Reviews of Interventions if sample correlations are unknown.

$$Y = \frac{1}{2}(Y_1 + Y_2);\ V_Y = \frac{1}{4}\left(V_{Y_1} + V_{Y_2} + 2r\sqrt{V_{Y_1}}\sqrt{V_{Y_2}}\right)$$

$$Y = \frac{1}{m}\left(\sum_{j}^{m} Y_j\right);\ V_Y = \left(\frac{1}{m}\right)^2\left(\sum_{j=1}^{m} V_i + \sum_{j\neq k}\left(r_{jk}\sqrt{V_j}\sqrt{V_k}\right)\right)$$

If participants in different training sessions within one article underwent the same training but were divided into separate subgroups, as for example seen in Alzahrani and Johnson, 2019, where 30 participants were divided into five groups of six, they were grouped together by combining the effect sizes and variances of each subgroup using the formulas above. If a shared control group was present, separate training methods in the experimental conditions were grouped together to avoid bias due to calculations based on the same control group data being included multiple times (see Baillon et al., 2019). Effect sizes from studies that reported a reduction in negative behaviour were adjusted so that all positive effect sizes reflected an increase in positive behaviour. For subgroup analyses, outcome measures were grouped together depending on how closely related the assessment was to measurements of behaviour. Objective assessments of behaviour, as well as behavioural assessments based on self-reports, were assessed together using the term 'behaviour'. In turn, measures of attitudes, intentions, knowledge etc. were assessed together under the term 'precursors'. If the same study assessed different outcome measures or employed different study designs, the relevant study data was included in each subgroup analysis of relevance and therefore more than once.

Moderator analyses were conducted for training methods (game vs discussion vs presentation vs simulation vs text vs video vs info), as well as the training platform (online vs offline) and the social setting (individual vs group) in which the training was performed. Articles grouped together for the overall meta-analysis (e.g., due to a shared control group) were kept separate for the moderator analyses if a moderator differed between groups.

2.5. Analysis strategy

The metafor package for R (Viechtbauer, 2010) was used to conduct the meta-analyses. The standardised mean difference Cohen's d was used as the effect size measure. Effect sizes were calculated using post-intervention scores of control and intervention groups if available. Alternatively, pre- and post-test assessments from the same groups were used instead. If the intervention was assessed at multiple time-points, the first assessment after the last training session was completed was used for the calculations.

Overall, seven random-effects models were fitted to the data. The first model included all studies selected for this meta-analysis to calculate an overall effect of cybersecurity training. Additionally, two models were fitted to compare the effects of assessments of behaviour and predictors, in order to assess whether predictors of behaviour were easier to change than behaviour. Lastly, due to the variety in study designs and outcome measures present in the sample, four models were fitted as subgroup analysis of these factors. More specifically, the data was divided into assessments of predictors of behaviour using pre- and post-test assessments, assessments of predictors using independent groups, assessments of behaviour using pre- and post-test assessments, and assessments of behaviour using independent groups. This was done to investigate whether the assessment of predictors vs. behaviour, as well as a more methodologically sound method, i.e. an experimental design using independent groups, would have an impact on the strength of the effect of training compared to pre- and post-test assessments. The trim-and-fill method was used to investigate publication bias in the included source material (Duval and Tweedie, 2004). Sensitivity analyses were performed using the leave-one-out method.

2.6. Moderators

To assess the impact of various training characteristics on training effects, several moderators, chosen based on areas for future research highlighted in Prümmer et al. (2024), were included in the analysis. These moderators are: the overall method with which the training was delivered (discussion vs game vs info vs presentation vs simulation vs text vs video); the use of more than one training method (multiple vs single); the platform (online vs offline vs varied) and social setting (group vs individual vs varied) in which the training was administered, as well as the modernness of the training method (modern vs. traditional). Modernness in this case aims to compare standard and long-used training methods such as texts, videos and presentations to more recently established training methods such as serious games or simulations. These moderators were tested in the overall meta-analysis.

3. Results

The search resulted in 55 articles that were eligible for inclusion. Within these articles, effect sizes could be computed for 69 studies. Overall, 100 comparisons were included in the various meta-analyses. This number is significantly higher than the number of included studies or articles due to the fact that some articles made use of multiple training methods, study designs and/or outcome measures. An overview of all included studies and comparisons, as well as their characteristics and effect sizes, can be found in Table 1, and an overview of the selection process can be found in Fig. 1.

Within the sample, 25 studies made use of independent groups to assess training effectiveness. In comparison, 45 studies employed a repeated measures design. With regard to outcome measures, 33 studies assessed behaviour either through objective measures such as click rates for phishing e-mails (e.g. Baillon et al., 2019) or participant self-reports (e.g. Alkhazi et al., 2022). Precursors of behaviour, such as intentions, knowledge or attitudes were assessed in 48 studies (e.g. van Steen and Deeleman, 2021; Veneruso et al., 2020). Sample sizes ranged from 15 (Cook et al., 2017) to 20,259 (Chatchalermpun and Daengsi, 2021) with a total sample of 81,642 across all studies.

3.1. Overall effect

The overall meta-analysis was executed by fitting a random-effects model to the data. A random-effects model was chosen due to the differences in both study samples and methodologies between studies. This decision was supported by a Q-test, which showed considerable heterogeneity in effect sizes ($Q(68) = 3305.13$, $p < 0.0001$). The analysis showed an overall medium-to-large positive effect, indicating that training has a positive effect on end-user cybersecurity ($d = 0.75$, 95%CI [0.58, 0.92]). A forest plot of all included studies can be found in Fig. 2.

Table 1
Data extracted from studies included in the meta-analysis.

Author	Subgroup	Sample Size	Outcome measure	Study Design	Training method	Platform	Social Setting	Effect Size	Variance
Abraham, 2012	Overall	98	Overall	Pre-Post	text & video	online	individual	0.97	0.0097
	Outcome Measure	98	Behaviour	Pre-Post	text & video	online	individual	1.16	0.0160
	Outcome Measure	98	Precursor	Pre-Post	text & video	online	individual	0.78	0.0101
Abraham, 2012	Overall	85	Overall	Pre-Post	text	online	individual	0.23	0.0079
	Outcome Measure	85	Behaviour	Pre-Post	text	online	individual	0.02	0.0112
	Outcome Measure	85	Precursor	Pre-Post	text	online	individual	0.44	0.0098
Abraham and Chengalur-Smith, 2019	Overall	197	Precursor	Independent Groups	text	online	individual	0.28	0.0016
Al Zaidy, 2020	Overall	25	Overall	Pre-Post	not specified	not specified	not specified	0.31	0.0321
	Outcome Measure	25	Behaviour	Pre-Post	not specified	not specified	not specified	0.55	0.0455
	Outcome Measure	25	Precursor	Pre-Post	not specified	not specified	not specified	0.06	0.0401
Al Zaidy, 2020	Overall	50	Overall	Pre-Post	not specified	not specified	not specified	0.34	0.0139
	Outcome Measure	50	Behaviour	Pre-Post	not specified	not specified	not specified	0.23	0.0154
	Outcome Measure	50	Precursor	Pre-Post	not specified	not specified	not specified	0.44	0.0220
Al-Hamar, 2010	Overall	129	Behaviour	Pre-Post	text, game & presentation	varied	varied	1.14	0.0383
Alahmari et al., 2022	Overall	79	Precursor	Independent Groups	Game	online	individual	2.26	0.0831
Albrechtsen and Hovden, 2010	Overall	168	Behaviour	Mixed	discussion	offline	group	0.09	0.0034
	Study Design	168	Behaviour	Independent Groups	discussion	offline	group	-0.04	0.0060
	Study Design	79	Behaviour	Pre-Post	discussion	offline	group	0.22	0.0033
Alkhazi et al., 2022	Overall	35	Precursor	Pre-Post	video & presentation	varied	varied	0.95	0.0107
Alkhazi et al., 2022	Overall	35	Precursor	Pre-Post	text & presentation	varied	varied	0.87	0.0107
Alkhazi et al., 2022	Overall	35	Precursor	Pre-Post	game & presentation	varied	varied	0.90	0.0107
Alkhazi et al., 2022	Overall	35	Precursor	Pre-Post	presentation	offline	group	0.55	0.0107
Alotaibi, 2019	Overall	50	Precursor	Pre-Post	game	online	individual	1.08	0.0065
Alzahrani and Johnson, 2019	Overall	30	Precursor	Pre-Post	game	offline	group	0.47	0.0980
Anzaldua Jr, 2016	Overall	32	Precursor	Independent Groups	video	online	individual	0.97	0.0370
Arain et al., 2019	Overall	560	Precursor	Independent Groups	not specified	online	individual	0.37	0.0094
Arain et al., 2019	Overall	465	Precursor	Independent Groups	not specified	online	individual	0.19	0.0198
Baillon et al., 2021	Overall	7994	Behaviour	Independent Groups	info/simulation/info & simulation	online	individual	0.34	0.0009
	Training Method	4006	Behaviour	Independent Groups	info	online	individual	0.26	0.0014
	Training Method	3789	Behaviour	Independent Groups	simulation	online	individual	0.43	0.0016
	Training Method	3805	Behaviour	Independent Groups	info & simulation	online	individual	0.34	0.0015
Baxter et al., 2016	Overall	116	Precursor	Independent Groups	game/info	online	individual	0.17	0.0444
	Training Method	78	Precursor	Independent Groups	game	online	individual	0.13	0.0641
	Training Method	83	Precursor	Independent Groups	info	online	individual	0.20	0.0595
Bodnar, 2021	Overall	44	Overall	Independent Groups	not specified	not specified	not specified	3.23	0.1584
	Outcome Measure	44	Behaviour	Independent Groups	not specified	not specified	not specified	2.91	0.1882
	Outcome Measure	44	Precursor	Independent Groups	not specified	not specified	not specified	3.55	0.2352
Chatchalermpun and Daengsi, 2021	Overall	20,259	Behaviour	Pre-Post	simulation, info & presentation	varied	varied	0.83	0.0004
Chen et al., 2020	Overall	178	Precursor	Independent Groups	game/info	online	individual	2.47	0.0141
	Training Method	117	Precursor	Independent Groups	game	online	individual	4.42	0.0485

(continued on next page)

Table 1 (*continued*)

Author	Subgroup	Sample Size	Outcome measure	Study Design	Training method	Platform	Social Setting	Effect Size	Variance
	Training Method	116	Precursor	Independent Groups	info	online	individual	2.82	0.0236
Chin et al., 2016	Overall	346	Behaviour	Independent Groups	video	online	individual	0.35	0.0012
CJ et al., 2018	Overall	8071	Precursor	Pre-Post	game	online	individual	0.35	0.0001
Clark, 2013	Overall	202	Precursor	Pre-Post	discussion, presentation & text	varied	varied	0.09	0.0005
Cook et al., 2017	Overall	15	Precursor	Pre-Post	game	online	group	0.67	0.0815
Curry et al., 2019	Overall	229	Behaviour	Pre-Post	info	online	individual	0.66	0.0043
Daengsi et al., 2022	Overall	20,134	Behaviour	Pre-Post	text & presentation	varied	varied	0.94	0.0002
DeCarlo, 2021	Overall	218	Overall	Independent Groups	game/presentation	online	individual	0.89	0.0329
	Outcome Measure	166	Behaviour	Independent Groups	game/presentation	online	individual	0.40	0.0657
	Outcome Measure	218	Precursor	Independent Groups	game/presentation	online	individual	1.37	0.0253
	Training Method	148	Precursor	Independent Groups	game	online	individual	1.42	0.0540
	Training Method	140	Precursor	Independent Groups	presentation	online	individual	0.23	0.0426
DeCusatis et al., 2022	Overall	132	Precursor	Pre-Post	game	online	individual	2.26	0.1132
Eftimie et al., 2022	Overall	235	Behaviour	Pre-Post	not specified	online	individual	0.21	0.0064
Goode, 2018	Overall	250	Precursor	Independent Groups	text, presentation & video	varied	varied	1.09	0.0205
Gordon et al., 2019	Overall	5416	Behaviour	Independent Groups	info	online	individual	-0.53	0.0079
Gundu and Flowerday, 2013	Overall	28	Precursor	Pre-Post	info	online	individual	2.53	0.3797
Hammond, 2019	Overall	340	Precursor	Independent Groups	not specified	not specified	not specified	-0.02	0.0037
Harrison, 2018	Overall	559	Behaviour	Independent Groups	text & simulation	online	individual	-0.10	0.0095
Ikhalia et al., 2019	Overall	40	Behaviour	Pre-Post	video & info	online	individual	2.86	0.1269
Jansson and von Solms, 2013	Overall	8231	Behaviour	Pre-Post	simulation	online	individual	0.34	0.0008
Kamar et al., 2022	Overall	153	Behaviour	Independent Groups	text	online	individual	-0.12	0.0386
Kävrestad et al., 2022	Overall	37	Behaviour	Independent Groups	text/game	online	individual	0.64	0.1426
	Training Method	23	Behaviour	Independent Groups	text	online	individual	1.58	0.2309
	Training Method	24	Behaviour	Independent Groups	game	online	individual	0.05	0.1715
Khan et al., 2023	Overall	154	Precursor	Pre-Post	presentation	offline	group	0.53	0.0022
Kim, 2010	Overall	85	Precursor	Pre-Post	presentation	offline	group	2.31	0.0431
Kim, 2010	Overall	127	Precursor	Pre-Post	info	online	individual	1.54	0.0172
Kim et al., 2020	Overall	1147	Behaviour	Independent Groups	info	not specified	not specified	0.31	0.0123
Kletenik et al., 2021	Overall	23	Precursor	Pre-Post	game	online	individual	1.04	0.0987
Kletenik et al., 2021	Overall	30	Precursor	Pre-Post	game	online	individual	0.90	0.0733
Kletenik et al., 2021	Overall	22	Precursor	Pre-Post	game	online	individual	1.07	0.1038
Lamour, 2008	Overall	60	Precursor	Independent Groups	presentation	varied	varied	1.94	0.0979
Martin, 2019	Overall	115	Behaviour	Pre-Post	text	online	individual	0.68	0.0107
McCrohan et al., 2010	Overall	396	Behaviour	Pre-Post	presentation	online	individual	0.17	0.0026
McKinney, 2021	Overall	2381	Behaviour	Pre-Post	not specified	not specified	not specified	0.34	0.0035
Ordonez, 2022	Overall	28	Precursor	Pre-Post	text	online	individual	0.54	0.0410
Ordonez, 2022	Overall	22	Precursor	Pre-Post	game	online	individual	0.23	0.0466
Ordonez, 2022	Overall	30	Precursor	Pre-Post	text	online	individual	1.04	0.0513
Robbins, 2020	Overall	201	Overall	Pre-Post	not specified	not specified	not specified	0.38	0.0035
	Outcome Measure	201	Precursor	Pre-Post	not specified	not specified	not specified	0.39	0.0040
	Outcome Measure	201	Behaviour	Pre-Post	not specified	not specified	not specified	0.37	0.0053
Shaw et al., 2011	Overall	39	Precursor	Pre-Post	text	online	individual	0.30	0.0203
Siponen et al., 2020	Overall	87	Precursor	Independent Groups	simulation	offline	group	0.69	0.0655
Stefaniuk, 2020	Overall	260	Precursor	Independent Groups	not specified	not specified	not specified	0.79	0.0075
Thornton and Turley, 2020	Overall	28	Overall	Pre-Post	game	online	individual	0.47	0.0126
	Outcome Measure	28	Behaviour	Pre-Post	game	online	individual	0.51	0.0306

(*continued on next page*)

Table 1 (*continued*)

Author	Subgroup	Sample Size	Outcome measure	Study Design	Training method	Platform	Social Setting	Effect Size	Variance
	Outcome Measure	28	Precursor	Pre-Post	game	online	individual	0.46	0.0296
Tschakert and Ngamsuriyaroj, 2019	Overall	67	Behaviour	Pre-Post	video, game & text	online	individual	1.27	0.3475
Tschakert and Ngamsuriyaroj, 2019	Overall	63	Behaviour	Pre-Post	video, game, text & presentation	varied	varied	0.37	0.4657
van Steen and Deeleman, 2021	Overall	178	Overall	Independent Groups	game	online	individual	0.57	0.0080
	Outcome Measure	178	Behaviour	Independent Groups	game	online	individual	0.45	0.0230
	Outcome Measure	178	Precursor	Independent Groups	game	online	individual	0.64	0.0177
Veneruso et al., 2020	Overall	20	Precursor	Pre-Post	text	offline	individual	0.85	0.0084
Veneruso et al., 2020	Overall	20	Precursor	Pre-Post	game	online	individual	1.10	0.0089
Waly, 2013	Overall	15	Precursor	Pre-Post	not specified	not specified	not specified	2.40	0.1868
White, 2022	Overall	290	Precursor	Pre-Post	text	online	individual	-0.19	0.0069
White, 2022	Overall	289	Precursor	Pre-Post	text	online	individual	-0.10	0.0069
Wu et al., 2021	Overall	110	Precursor	Independent Groups	game	varied	varied	0.12	0.0269
Younes, 2014	Overall	20	Precursor	Independent Groups	info	not specified	not specified	1.13	0.2320

3.2. Subgroup analyses

In total, six additional random-effects models were fitted to the data to assess the effect of training for a variety of conditional factors that were observed in the data. As outlined previously, two models were fitted to assess training effectiveness with regard to different outcome measures used in the selected studies. This analysis showed that, overall, assessments of predictors of behaviour, which included 48 comparisons, had a large effect on end-user cybersecurity ($d = 0.84$, 95%CI [0.64, 1.04]), while assessments of behaviour, which included 33 comparisons, had a medium effect on end-user cybersecurity ($d = 0.55$, 95%CI [0.36, 0.76]). Both effects were significant at $p < 0.0001$. The four remaining models were used to assess differences in effects due to both study design and outcome measure assessments. A forest plot of included studies, as well as results of the analyses, can be seen in Figs. 3-6. The first model, which was made up of 32 comparisons, included studies where precursors of behaviour were assessed through a repeated measures design. The analysis shows a medium-to-large positive effect ($d = 0.79$, 95%CI [0.57, 1.01]) which was significant at $p < 0.0001$ (see Fig. 3). The second model, which was made up of 17 comparisons, included studies where precursors of behaviour were assessed through an independent group design. This analysis again shows a large positive effect of training ($d = 1.02$, 95%CI [0.58, 1.46]), which was also significant at $p < 0.0001$ (see Fig. 4). The third model, which was made up of 23 comparisons, included studies where behaviour was assessed using repeated measures designs. The analysis shows a medium positive effect ($d = 0.63$, 95%CI [0.44, 0.83]), which was again significant at $p < 0.0001$ (see Fig. 5). Lastly, a model including studies where behaviour was assessed using independent groups, which was made up of 11 comparisons, showed a small positive effect of training ($d = 0.36$, 95%CI [-0.09, 0.80]) (see Fig. 6). However, this effect was not significant ($p = 0.12$).

3.3. Publication bias and sensitivity analysis

The trim-and-fill method was used to assess publication bias (Duval and Tweedie, 2004). With this method, studies in the sample that potentially lead to funnel plot asymmetry, which can be an indicator of publication bias due to missing studies, are removed from the sample (trimming). Then, a new centre of the funnel plot is estimated, after which the removed study, as well as its missing 'counterpart' (filling) is readded to the plot. For the overall meta-analysis, no publication bias was detected. See Fig. 7 for the funnel plot of all included studies. Publication bias was also assessed for all other random-effects models that were fitted to the data. No publication bias was detected in any of the fitted models.

In addition to publication bias, a sensitivity analysis was conducted to see whether any individual study within the sample had a substantial influence on the overall effect. This was achieved by using the leave-one-out method. With this method, a series of meta-analyses are carried out, with each meta-analysis excluding a different single study from the sample. The generated effect sizes for the various meta-analyses can then be compared. If a single study in the sample has a strong influence on the effect size, a difference appears between the meta-analysis that excludes that study compared to the other meta-analyses. We deemed changes in effect size significant enough to report when an effect size difference of 0.1 or more was observed. A change in effect size was detected in the models that included studies assessing precursors of behaviour and behaviour through independent groups. The removal of results obtained from a study conducted by Bodnar (2021) lowered the effect size from $d = 1.02$, 95%CI [0.58, 1.46] to $d = 0.88$, 95%CI [0.50, 1.26] for the random-effects model assessing precursors of behaviour through independent groups and from $d = 0.36$, 95%CI [-0.09, 0.80] to $d = 0.14$, 95%CI [-0.08, 0.35] for the random-effects model assessing behaviour through independent groups. No changes in effect sizes or levels of significance were observed for the remaining random-effects models.

3.4. Moderators

Data on moderators was not always available for all studies. Due to the missing data, the number of studies was reduced from 69 to a minimum of 50 for some moderator analyses (method & modernness) and 64 (multiple methods) or 65 for others (platform & social setting). The results of the moderator analyses are discussed below.

3.4.1. Training method

A moderator analysis of training methods indicated that game-based ($d = 1.03$, 95%CI [0.61, 1.45], $p < 0.0001$), information-based ($d = 0.93$, 95%CI [0.33, 1.53], $p = 0.0024$) and presentation-based ($d = 0.93$, 95%CI [0.21, 1.65], $p = 0.0113$) methods were found to have the strongest effects. All three effects were large and significant. The lowest level of effectiveness was found for discussion-based training methods ($d = 0.09$, 95%CI [-1.64, 1.82], $p = 0.92$). The remaining training methods, namely simulation-based ($d = 0.48$, 95%CI [-0.53, 1.49], $p = 0.35$), text-

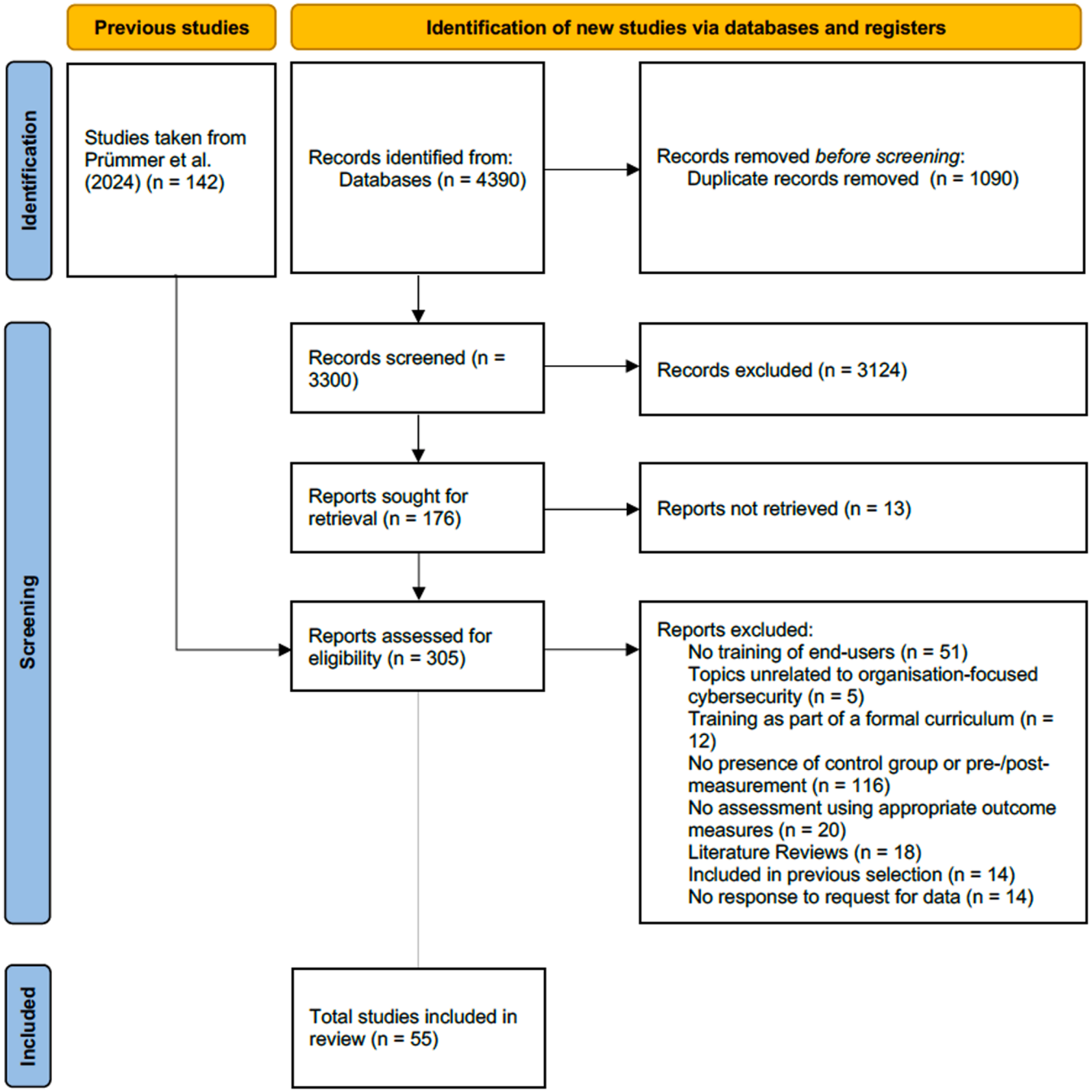

Fig. 1. Flow chart for identification of eligible studies.

based ($d = 0.44$, 95%CI [-0.09, 0.97], $p = 0.11$) and video-based ($d = 0.66$, 95%CI [-0.58, 1.89], $p = 0.30$) showed moderate and non-significant effects. Overall, no significant effect of method on training effectiveness could be observed, meaning that there was no significant difference in terms of effectiveness between any of the training methods assessed here ($QM(6) = 4.28$, $p = 0.64$).

3.4.2. Multiple training methods

While both standalone ($d = 0.80$, 95%CI [0.56, 1.03], $p < 0.0001$) and combinations of multiple training methods ($d = 0.88$, 95%CI [0.43, 1.32], $p = 0.001$) had large significant effects, no significant difference between them could be identified ($QM(1) = 0.0978$, $p = 0.75$).

3.4.3. Training platform

An overall effect of training platform could not be identified ($QM(2) = 0.05$, $p = 0.97$), indicating that no significant difference between platform types exists in the sample. Individually, all platform types, namely online ($d = 0.78$, 95%CI [0.54, 1.02], $p < 0.0001$), offline ($d = 0.78$, 95%CI [0.15, 1.40], $p = 0.01$) and a combination or variation of the two ($d = 0.84$, 95%CI [0.34, 1.35], $p = 0.001$) displayed large, significant effects on end-user cybersecurity.

3.4.4. Social setting

All social settings, namely individual ($d = 0.78$, 95%CI [0.54, 1.03], $p < 0.0001$), group ($d = 0.75$, 95%CI [0.12, 1.38], $p = 0.02$) and a combination or variation of the two ($d = 0.84$, 95%CI [0.34, 1.35], $p = 0.001$) showed large, significant effects. A significant difference between

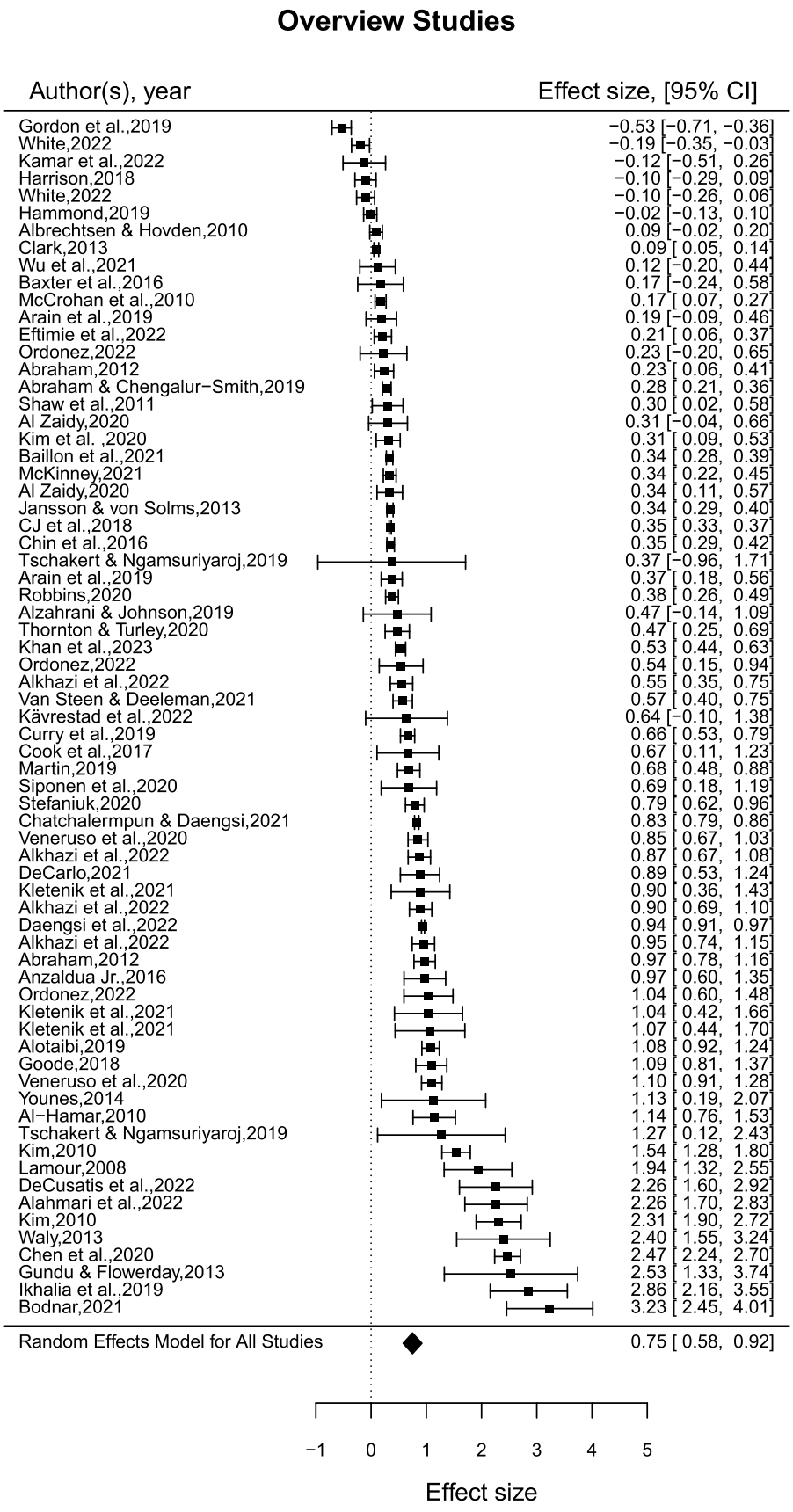

Fig. 2. Overview of included studies with corresponding effect sizes.

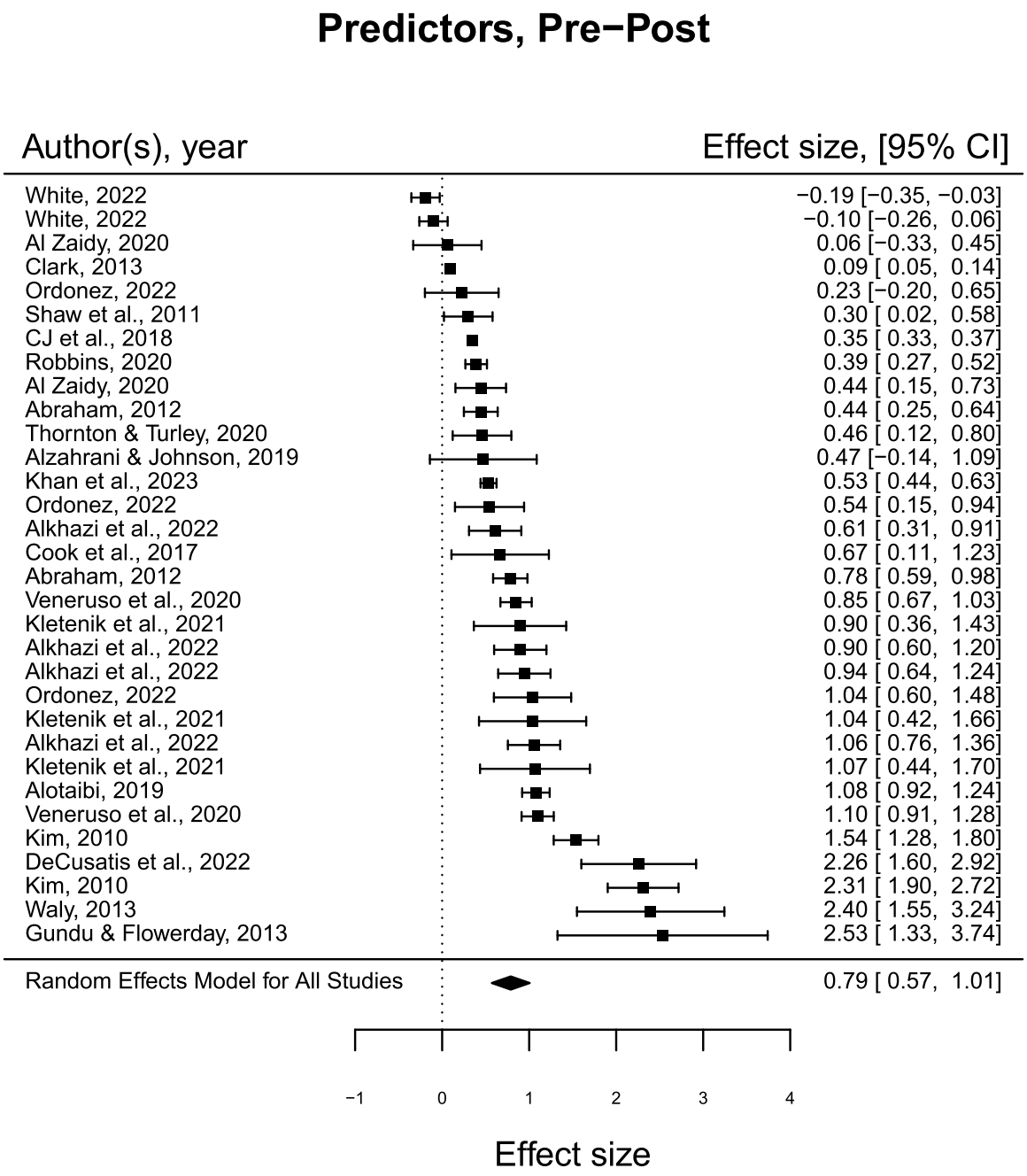

Fig. 3. Overview of studies assessing predictors through pre-post assessments with corresponding effect sizes.

them could not be identified ($QM(2) = 0.0612$, $p = 0.97$).

3.4.5. Modernness

Both modern ($d = 0.91$, 95%CI [0.53, 1.28], $p < 0.0001$) and traditional ($d = 0.71$, 95%CI [0.38, 1.04], $p < 0.0001$) training methods displayed large, significant effects with regard to their effect on end-user cybersecurity. A significant difference between them could not be identified ($QM(1) = 0.569$, $p = 0.45$).

5. Discussion

Due to the growing importance of end-user cybersecurity behaviour, training of these end-users becomes a necessity. While systematic literature reviews on this topic have been conducted previously, examining the effectiveness of these training programs aids in our further understanding of end-user behaviour and how to mitigate damages caused by it. This analysis of 69 studies on training in cybersecurity shows that, overall, training of end-users has a strong impact. While the success of training in cybersecurity was already evidenced by the findings reported in systematic reviews (e.g. Aldawood and Skinner, 2019; Hendrix et al., 2016; Prümmer et al., 2024), this confirmation of effectiveness through a meta-analysis is, in general, encouraging. Through this research, we were also able to establish a variety of smallest effect sizes of interest (SESOI) based on outcome measures and study designs, which can serve as anchors for any future research on cybersecurity training. In the following section, we will dive deeper into the results gained from the various subgroup and moderator analyses conducted in this article and assess what these results mean for research on cybersecurity training in the future.

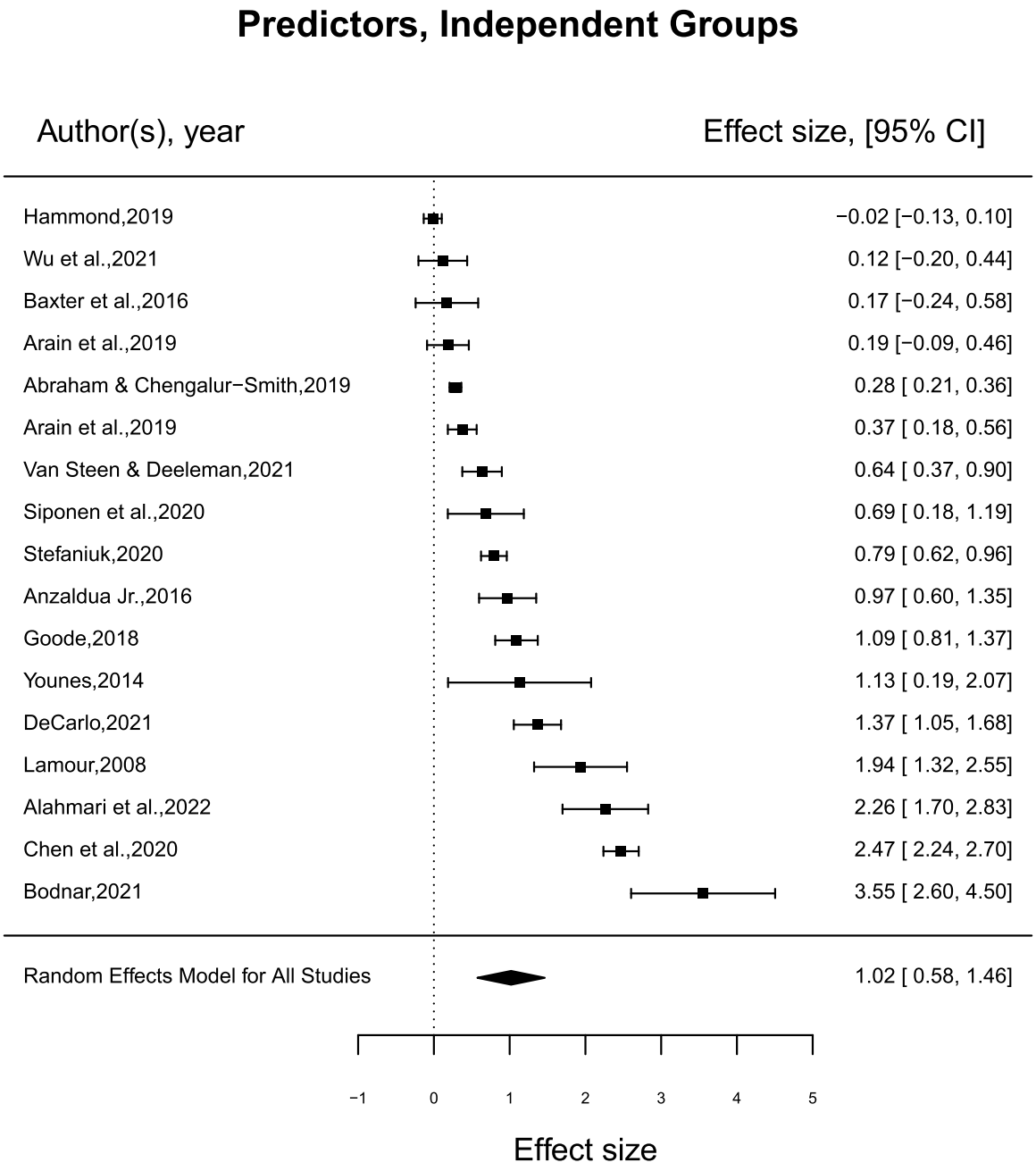

Fig. 4. Overview of studies assessing predictors through independent groups with corresponding effect sizes.

5.1. *Effect of training on behaviour*

Overall, the analysis of studies evaluating behaviour vs precursors highlighted a slight difference in effect sizes between the two. While evaluations of effectiveness using precursors produced medium-to-large effects, effects for assessments of behaviour could only be classified as medium. This difference is understandable and has been explored in a variety of behavioural theories such as the theory of planned behaviour (Ajzen, 1991), which posits that behaviour is influenced by attitudes and intentions among others, but that the strength of the effect lessens when approaching behaviour. A confirmation of this mechanism for cybersecurity behaviours provides some indication that the use of theory in general to influence cybersecurity training holds value.

While the analyses conducted on both behavioural and predictor variables proved to be insightful by themselves, we wanted to further analyse whether the study design used to evaluate these variables contributes to the size of the effect. Therefore, we decided to divide studies within these two categories further according to their study design, leading to four random-effects models being fitted to the data. Based on well-established theories of behaviour we can assume that measurements of predictors of behaviour will, on average, produce larger results than measurements of behaviour (Ajzen, 1991). This assumption was confirmed. A second assumption was made concerning the study design. Though there are potential flaws in between-subjects designs and they are not suitable for all types of research, the overestimation of effects through within-subjects designs is well-established (Charness et al., 2012). This is the case because of a variety of dependencies that occur in within-subjects designs, such as anticipation of research objectives due to repeated exposure to assessments and practice effects (Keren, 2014).

The results of this meta-analysis reveal that an overestimation of effects seems to occur when evaluating cybersecurity using within-

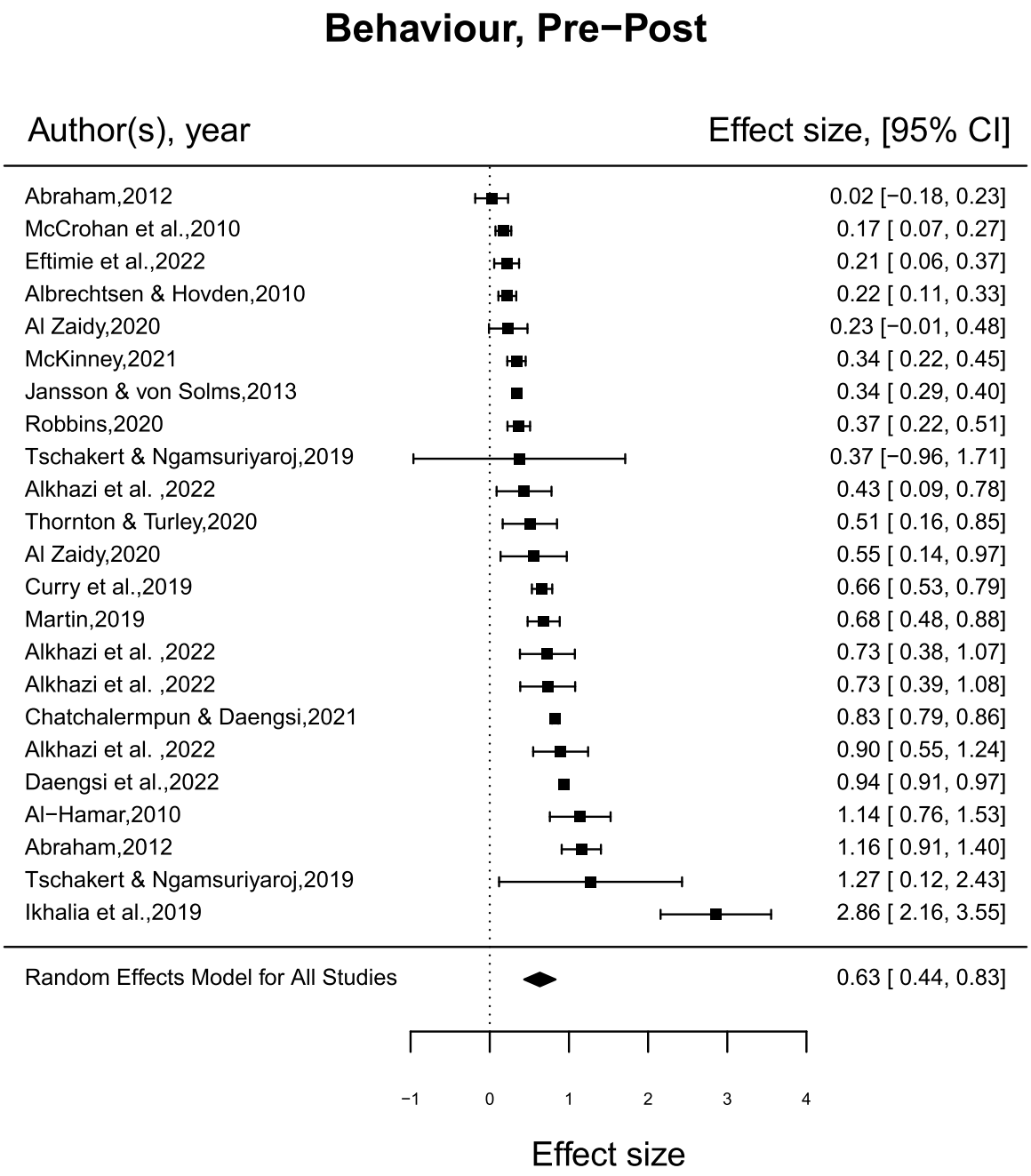

Fig. 5. Overview of studies assessing behaviour through pre-post assessments with corresponding effect sizes.

subjects designs, at least regarding assessments of behaviour. When assessing behaviour via within-subjects designs, effectiveness remains relatively high with a medium effect size of $d = 0.63$. When looking at these results in isolation, training in cybersecurity seems straightforwardly successful. Unfortunately, this is not entirely the case. Studies that assess behaviour change, which is the overall end goal of training in the first place, by using scientifically superior methods report only small effects that are not significant. Furthermore, this effect size seems to be largely driven by one study in particular (Bodnar, 2021), as the results of the leave-one-out analysis show that the effect size decreases further from $d = 0.36$ to $d = 0.14$ when this study is removed. This finding is highly relevant considering the necessity of end-user training to ensure the continued safety of organisations, as well as the high costs often associated with implementing training.

Though we cannot say conclusively why there is such a discrepancy between the effects of training on precursors and behaviour, there seems to be more at play than just a continuation of effects established by behavioural models and theories. One potential explanation is that the training methods that are currently in use are not entirely suitable to address behaviour, and instead are more tailored towards achieving changes in other factors. Many of the articles included in this sub-analysis state that their goal is to achieve changes in awareness, rather than changes in behaviour (e.g. Al-Hamar, 2010; CJ et al., 2018; Daengsi et al., 2022). Another explanation could be that users are aware of the behaviours they need to perform, but conditions in their environment are preventing them from doing so. More specifically, the amount of effort often required from end-users to engage in the variety of security practices that are taught in these training programs is too extensive to implement, specifically alongside other responsibilities such as job duties (Beautement et al., 2008; Guo et al., 2011). Established theories of behaviour such as protection motivation theory recognise these discrepancies. Within protection motivation theory, the concept of

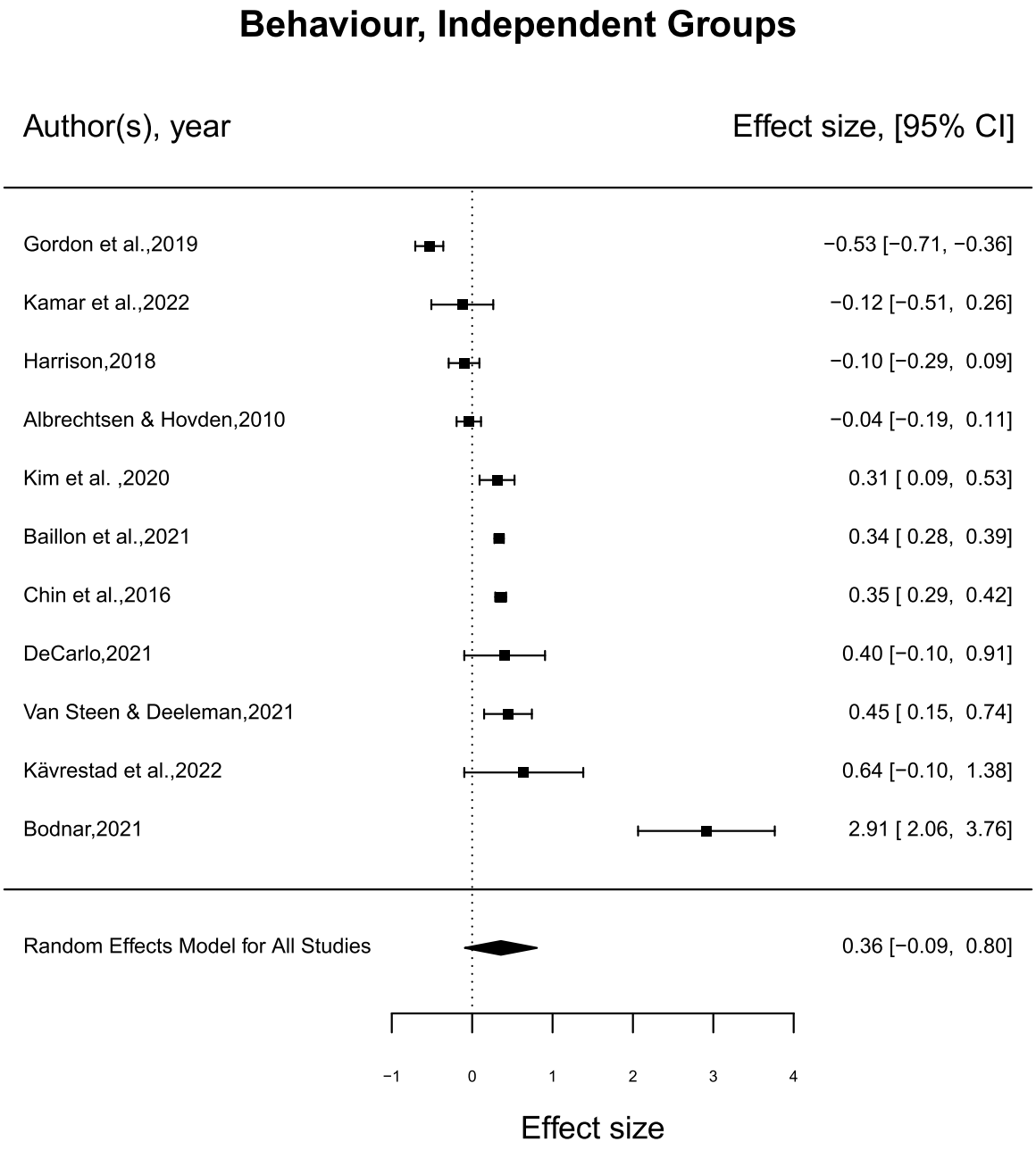

Fig. 6. Overview of studies assessing behaviour through independent groups with corresponding effect sizes.

self-efficacy implies that a high belief in one's ability to perform a behaviour is necessary for the behaviour to occur. Similarly, if the costs associated with performing the behaviour are too high, a behaviour is less likely to occur (Rogers, 1983). Using insights from these theories to adapt training programs to the realities that end-users face could improve the effect of training programs on behaviour.

Overall, any future training programs aimed at changing cybersecurity behaviour can use the small effect size ($d = 0.36$) found in this meta-analysis as a SESOI/benchmark for effectiveness testing, as this effect size currently constitutes the most accurate representation of the effects current training programs have on end-user behaviour. Using new insights and developments in how to train end-users should lead to increased effect sizes, and thus effectivity of the training programs. The SESOI is therefore a starting point for future training development and not simply a goal that needs to be met for a training initiative to be successful.

5.2. Effects of moderators

The evaluation of a range of moderators provided valuable insights into a variety of training characteristics and their influence on effectiveness. For example, while some training methods such as games, presentations and information-based approaches produced large and significant effects, other methods did not. Most surprisingly, discussion-based approaches were least successful and did not contribute to changes in end-user cybersecurity. An explanation for this could be that only one study adopted this method to train end-users (Albrechtsen and Hovden, 2010). In general, the findings with regard to individual training method effectiveness should be interpreted with caution, particularly due to the limited number of studies included for some methods. Another potentially insightful factor we wanted to investigate, was whether more traditional techniques of training, such as text-based, video-based or presentation-based approaches, contributed more, less or equally to effectiveness than modern approaches such as games or

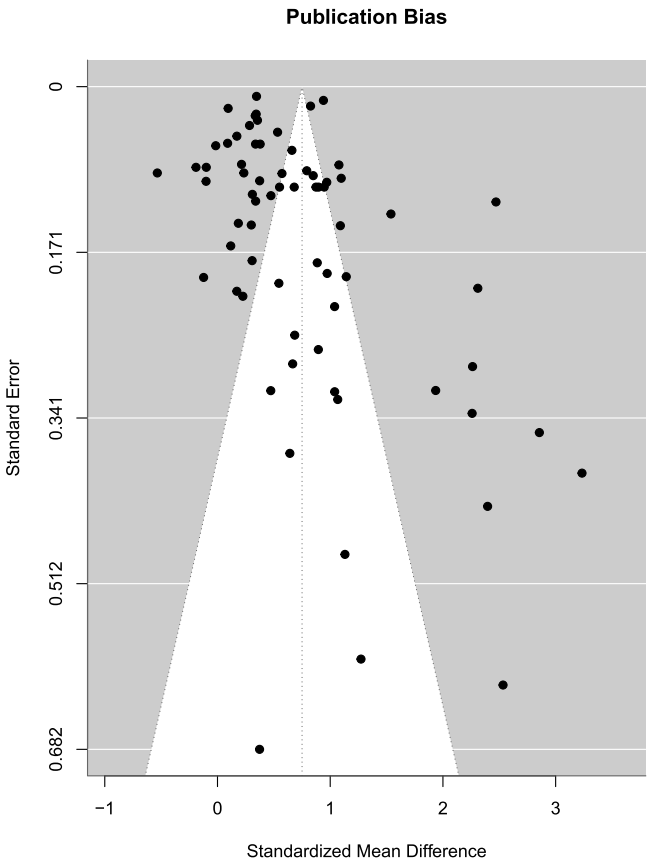

Fig. 7. Assessment of publication bias.

simulations. Our analysis shows that both approaches have a large and significant effect. Based on this, the modernness of approaches does not seem to impact effectiveness differently than traditional techniques of teaching.

Some authors have speculated about the benefits and drawbacks of including a combination of training methods in cybersecurity training programs (Abawajy, 2014; P. Kim, 2010). By adding a moderator that assesses the inclusion of one vs multiple training methods, we are able to provide some insight and guidance on how valuable this approach might be. Our findings show that while a combination of methods produced significant, large effects, so did training programs that included standalone methods. The potential risks of overstimulating participants and the increased work and resources for organisations that go into planning and organising multiple training mechanisms do not appear to be worth the effort, at least based on the data included in this meta-analysis. As none of the studies included in this paper administered longitudinal training programs spanning several months, these findings only apply to the administration of multiple training methods at a single moment in time. Similar findings were made for training platforms and social settings. Online training, in-person training and a variation of both all showed large effects. The same was determined for individual training, group training and a variation of both. Since no significant difference between platforms and social settings could be determined, it seems sensible to argue that these training characteristics should be determined based on convenience and suitability to the overall training method, rather than implications for effectiveness.

5.3. Limitations

Although we were able to include a large number of studies in this meta-analysis, some concessions concerning the inclusion criteria had to be made. While the highlighted difference between findings from within-subjects and between-subjects research is valuable, the inclusion of both methodologies within the same analysis might lead to some limitations, such as an overestimation of the overall effectiveness of cybersecurity training. In addition, even though we would have preferred to assess the effects of training on behaviour entirely through objective measurements, this was not possible due to the limited number of studies that made use of this type of outcome measure. Therefore, objective and self-reported measurements of behaviour were combined, which again might lead to an overestimation of effectiveness. Overall, coding was often difficult as some studies did not report how exactly the training was implemented, e.g., which method was used or whether the training was online or in-person. This could also be a contributing factor as to why some moderator analyses that were conducted contained only limited numbers of studies. For example, discussion-based methods were only covered in one study, video-based methods in two and simulations in three. By contrast, game- and text-based methods were used in 18 and 11 studies respectively. This discrepancy in how (in-)frequently different training methods were encountered in the literature makes interpretation of results with regard to training method effectiveness difficult. Furthermore, we are not able to pinpoint why specific studies are more successful than others in the sample, particularly when comparing studies that utilised, for example, similar training methods, as theoretical underpinnings and other background information necessary to make those determinations was often lacking in the source material. Lastly, it was not possible to conduct an analysis on long-term effectiveness of training, due to the limited number of studies reporting results beyond an initial evaluation after the training program was administered. More specifically, only four studies in the sample (Abraham, 2012; Albrechtsen and Hovden, 2010; Al-Hamar, 2010; Siponen et al., 2020) conducted follow-up assessments and timelines in which follow-ups were conducted ranged from 2 weeks to 6 months after an initial effectiveness evaluation.

5.4. Future research

Through this meta-analysis, several gaps within the literature have been outlined, particularly pertaining to theoretical underpinnings of training programs, a lack of differentiation between cybersecurity behaviours and minimal changes in end-user behaviour compared to predictors of behaviour. These gaps offer a variety of opportunities for future research to address them. In general, future research on cybersecurity training of end-users should focus on building a theoretical basis that aids in determining what is needed within the training programs to change behaviour. As outlined in Prümmer et al. (2024), it appears as if a majority of the training programs included in this meta-analysis are developed based on common-sense determinations, rather than a clear exploration of goals and needs. A lack of differentiation between target behaviours (such as phishing or screen locking) and training methods could also be observed. Not all cybersecurity behaviours require the same thought processes and actions. Not engaging with a phishing e-mail is linked to continued vigilance and conscious checking of various factors such as senders and links, whereas screen locking is much more related to habit formation. Training these behaviours the same way might not be the best solution. In addition, more insights from participants and particularly employees themselves are needed to determine if other factors hinder the implementation of secure behaviours. While knowledge about cybersecurity or the intention to act securely might be high, organisational influences such as a lack of security culture within the organisation could play a role. Other external influences could be a depleted compliance budget (Beautement et al., 2008) due to other policies and guidelines the employees already need to comply with. Furthermore, difficulty in adhering to security guidelines could be caused by a lack of automatic solutions, or a high amount of friction employees need to overcome to be secure. Reducing this friction often requires copious amounts of effort and dedication that is very likely not available to these end-users. As outlined in Collins and

Hinds (2021), employees were much more likely to act securely when solutions were enabled automatically, encouraged through reminders, or easier to implement due to shortcuts. Overall, more research is needed to determine why the changes in behaviour are so minimal.

6. Conclusion

This meta-analysis found that while training significantly increases predictors of end-user behaviour, such as attitudes or knowledge, changes in behaviour can only be observed minimally, particularly when evaluated using a between-subjects study design. The lack of theoretical models to inform training creation observed in this sample of studies could be a contributing factor. Furthermore, there is a potential unsuitability of current training methods to address actual behaviour, as there is no differentiation in how certain cybersecurity behaviours are trained. Here, it is not necessarily the case that one method of training is superior over another in general, as shown by the various moderator analyses that were conducted as part of this analysis, but that we need to find methods of training that are suitable for each cybersecurity behaviour we are attempting to address individually. Future research should be concerned with identifying unique characteristics of the different cybersecurity behaviours that end-users engage in and outlining training methods that address these characteristics.

CRediT authorship contribution statement

Julia Prümmer: Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Tommy van Steen:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Bibi van den Berg:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Papers marked with an asterisk (*) were included in the meta-analysis

Abawajy, J., 2014. User preference of cyber security awareness delivery methods. Behaviour Inf. Technol. 33 (3). https://doi.org/10.1080/0144929X.2012.708787.

* Abraham, S., 2012. Exploring the effectiveness of information security training and persuasive messages. ProQuest Dissertations and Theses. State University of New York at Albany.

* Abraham, S., Chengalur-Smith, I., 2019. Evaluating the effectiveness of learner controlled information security training. Comput. Secur. 87. https://doi.org/10.1016/j.cose.2019.101586.

Ajzen, I., 1991. The theory of planned behavior. Organ. Behav. Hum. Decis. Process. 50 (2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T.

* Al Zaidy, A., 2020. Impact of training on employee actions and information security awareness in academic institutions. ProQuest Dissertations and Theses. Northcentral University.

* Alahmari, S., Renaud, K., Omoronyia, I., 2022. Moving beyond cyber security awareness and training to engendering security knowledge sharing. In: Information Systems and E-Business Management. https://doi.org/10.1007/s10257-022-00575-2.

* Albrechtsen, E., Hovden, J., 2010. Improving information security awareness and behaviour through dialogue, participation and collective reflection. An intervention study. Comput. Secur. 29 (4). https://doi.org/10.1016/j.cose.2009.12.005.

Aldawood, H., Skinner, G., 2019. An academic review of current industrial and commercial cyber security social engineering solutions. In: Proceedings of the 3rd International Conference on Cryptography, Security and Privacy, pp. 110–115. https://doi.org/10.1145/3309074.3309083.

* Al-Hamar, M.K., 2010. Reducing the risk of e-mail phishing in the state of qatar through an effective awareness framework. PQDT - UK & Ireland. Loughborough University (United Kingdom).

* Alkhazi, B., Alshaikh, M., Alkhezi, S., Labbaci, H., 2022. Assessment of the impact of information security awareness training methods on knowledge, attitude, and behavior. IEEe Access. 10, 132132–132143. https://doi.org/10.1109/ACCESS.2022.3230286.

* Alotaibi, F.F.G, 2019. Evaluation and enhancement of public cyber security awareness. PQDT - UK & Ireland. University of Plymouth (United Kingdom).

* Alzahrani, A., Johnson, C., 2019. Autonomy motivators, serious games, and intention toward ISP compliance. Int. J. Serious Games 6 (4). https://doi.org/10.17083/ijsg.v6i4.315.

Anvari, F., Lakens, D., 2021. Using anchor-based methods to determine the smallest effect size of interest. J. Exp. Soc. Psychol. 96, 104159. https://doi.org/10.1016/j.jesp.2021.104159.

* Anzaldua Jr, R., 2016. Does information security training change hispanic students' attitudes toward the perception of risk in the management of data security. ProQuest Dissertations and Theses. Northcentral University.

* Arain, M.A., Tarraf, R., Ahmad, A., 2019. Assessing staff awareness and effectiveness of educational training on IT security and privacy in a large healthcare organization. J. Multidiscip. Healthc. 12, 73–81. https://doi.org/10.2147/JMDH.S183275.

* Baillon, A., Bruin, J.de, Emirmahmutoglu, A., Veer, E.van de, Dijk, B.van, 2019. Informing, simulating experience, or both: a field experiment on phishing risks. PLoS. One 14 (12). https://doi.org/10.1371/journal.pone.0224216.

* Baxter, R.J.E.W., Holderness Jr., D.K., Wood, D.A, 2016. Applying basic Gamification techniques to IT compliance training: evidence from the lab and field. J. Inf. Syst. 30 (3). https://doi.org/10.2308/isys-51341.

Beautement, A., Sasse, M.A., Wonham, M., 2008. The compliance budget: managing security behaviour in organisations. In: Proceedings of the 2008 New Security Paradigms Workshop. Association for Computing Machinery, pp. 47–58. https://doi.org/10.1145/1595676.1595684.

* Bodnar, R.C., 2021. Improving the understanding of information security adherence in higher education: a quantitative ex-post facto design study [Ph.D., Northcentral University]. *ProQuest Dissertations and Theses* (2794477533). ProQuest Dissertations & Theses Global.

Borenstein, M., Hedges, L.V., Higgins, J.P., Rothstein, H.R., 2021. Introduction to Meta-Analysis. John Wiley & Sons.

Carlson, A., 2020. Combating insider threat with proper training. ProQuest Dissertations and Theses. Utica College.

Charness, G., Gneezy, U., Kuhn, M.A., 2012. Experimental methods: between-subject and within-subject design. J. Econ. Behav. Organ. 81 (1), 1–8. https://doi.org/10.1016/j.jebo.2011.08.009.

* Chatchalermpun, S., Daengsi, T., 2021. Improving cybersecurity awareness using phishing attack simulation. IOP. Conf. Ser. Mater. Sci. Eng. 1088 (1). https://doi.org/10.1088/1757-899X/1088/1/012015.

* Chen, T., Stewart, M., Bai, Z., Chen, E., Dabbish, L., Hammer, J., 2020. Hacked time: design and evaluation of a self-efficacy based cybersecurity game. In: Proceedings of the 2020 ACM Designing Interactive Systems Conference, pp. 1737–1749. https://doi.org/10.1145/3357236.3395522.

* Chin, A.G., Etudo, U., Harris, M.A., 2016. On mobile device security practices and training efficacy: an empirical study. Inf. Educ. 15 (2). https://doi.org/10.15388/infedu.2016.12.

Chowdhury, N., Gkioulos, V., 2021. Cyber security training for critical infrastructure protection: a literature review. Comput. Sci. Rev. 40. https://doi.org/10.1016/j.cosrev.2021.100361.

* CJ, G., Pandit, S., Vaddepalli, S., Tupsamudre, H., Banahatti, V., Lodha, S., 2018. PHISHY - a serious game to train enterprise users on phishing awareness. In: Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, pp. 169–181. https://doi.org/10.1145/3270316.3273042.

* Clark, C.Y., 2013. A study on corporate security awareness and compliance behavior intent. ProQuest Dissertations and Theses. Pace University.

Coenraad, M., Pellicone, A., Ketelhut, D.J., Cukier, M., Plane, J., Weintrop, D., 2020. Experiencing cybersecurity one game at a time: A systematic review of cybersecurity digital games. Simul. Gam. 51 (5). https://doi.org/10.1177/1046878120933312.

Collins, E.I.M., Hinds, J., 2021. Exploring Workers' subjective experiences of habit formation in cybersecurity: a qualitative survey. Cyberpsychol. Behav. Soc. Network. 24 (9), 599–604. https://doi.org/10.1089/cyber.2020.0631.

* Cook, A., Smith, R.G., Maglaras, L., Janicke, H., 2017. SCIPS: using experiential learning to raise cyber situational awareness in industrial control system. Int. J. Cyber Warfare Terrorism 7 (2). https://doi.org/10.4018/IJCWT.2017040101.

* Curry, M., Marshall, B., Correia, J., Crossler, R.E., 2019. InfoSec process action model (IPAM): targeting insiders' weak password behavior. J. Inf. Syst. 33 (3). https://doi.org/10.2308/isys-52381.

* Daengsi, T., Pornpongtechavanich, P., Wuttidittachotti, P., 2022. Cybersecurity awareness enhancement: a study of the effects of age and gender of thai employees associated with phishing attacks. Educ. Inf. Technol. (Dordr) 27 (4), 4729–4752. https://doi.org/10.1007/s10639-021-10806-7.

* DeCarlo, S.M., 2021. Measuring the application of knowledge gained from the gamification of cybersecurity training in healthcare. In: Dissertation Abstracts International: Section B: The Sciences and Engineering, 82. ProQuest Information & Learning. Issues.

* DeCusatis, C., Alvarico, E., Dirahoui, O., 2022. Gamification of Cybersecurity training. In: Proceedings of the 1st International Workshop on Gamification of Software Development, Verification, and Validation, pp. 10–13. https://doi.org/10.1145/3548771.3561409.

Duval, S., Tweedie, R., 2004. Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. Biometrics 56 (2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x.

* Eftimie, S., Moinescu, R., Racuciu, C., 2022. Spear-phishing susceptibility stemming from personality traits. IEEe Access. 10, 73548–73561. https://doi.org/10.1109/ACCESS.2022.3190009.

* Goode, J., 2018. Comparing training methodologies on Employee's cybersecurity countermeasures awareness and skills in traditional vs. socio-technical programs. ProQuest Dissertations and Theses. Nova Southeastern University.

* Gordon, W.J., Wright, A., Glynn, R.J., Kadakia, J., Mazzone, C., Leinbach, E., Landman, A., 2019. Evaluation of a mandatory phishing training program for high-risk employees at a US healthcare system. J. Am. Med. Inf. Assoc. 26 (6). https://doi.org/10.1093/jamia/ocz005.

* Gundu, T., Flowerday, S.V., 2013. Ignorance to awareness: Towards an information security awareness process. SAIEE Africa Res. J. 104 (2). https://doi.org/10.23919/SAIEE.2013.8531867.

Guo, K.H., Yuan, Y., Archer, N.P., Connelly, 2011. Understanding Nonmalicious security violations in the workplace: a composite behavior model. J. Manag. Inf. Syst. 28 (2), 203–236. https://doi.org/10.2753/MIS0742-1222280208.

* Hammond, S.T., 2019. Threat and coping appraisals on information security awareness training effectiveness: a quasi-experimental study. ProQuest Dissertations and Theses. Capella University.

* Harrison, B., 2018. Does Anti-phishing training protect against organizational cyber attacks?: An empirical assessment of training methods and employee readiness. ProQuest Dissertations and Theses. State University of New York at Buffalo.

Hendrix, M., Al-Sherbaz, A., Bloom, V., 2016. Game based cyber security training: are serious games suitable for cyber security training? Int. J. Ser. Games 3 (1). https://doi.org/10.17083/ijsg.v3i1.107.

Hepp, S.L., Tarraf, R.C., Birney, A., Arain, M.A., 2018. Evaluation of the awareness and effectiveness of IT security programs in a large publicly funded health care system. Health Inf. Manag. J. 47 (3). https://doi.org/10.1177/1833358317722038.

Higgins, J.P., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions*.

* Ikhalia, E., Serrano, A., Bell, D., Louvieris, P., 2019. Online social network security awareness: mass interpersonal persuasion using a Facebook app. Inf. Technol. People 32 (5). https://doi.org/10.1108/ITP-06-2018-0278.

* Jansson, K., von Solms, R., 2013. Phishing for phishing awareness. Behaviour Inf. Technol. 32 (6). https://doi.org/10.1080/0144929X.2011.632650.

* Kamar, E., Howell, C.J., Maimon, D., Berenblum, T., 2022. The moderating role of thoughtfully reflective decision-making on the relationship between information security messages and SMiShing victimization: an experiment. Justice Q.. https://doi.org/10.1080/07418825.2022.2127845.

* Kävrestad, J., Hagberg, A., Nohlberg, M., Rambusch, J., Roos, R., Furnell, S., 2022. Evaluation of contextual and game-based training for phishing detection. Future Internet. 14 (4). https://doi.org/10.3390/fi14040104.

Keren, G., 2014. Between-or within-subjects design: A methodological dilemma. A Handbook Data Anal. Behav. Sci. 1, 257–272.

* Khan, N.F., Ikram, N., Murtaza, H., Riphah, M., 2023. Evaluating protection motivation based cybersecurity awareness training on Kirkpatrick's Model. Comput. Secur. 125. https://doi.org/10.1016/j.cose.2022.103049.

* Kim, B., Lee, D.-Y., Kim, B., 2020. Deterrent effects of punishment and training on insider security threats: a field experiment on phishing attacks. Behav. Inf. Technol. 39 (11). https://doi.org/10.1080/0144929X.2019.1653992.

* Kim, P., 2010. Measuring the effectiveness of information security training: A comparative analysis of computer -based training and instructor -based training. ProQuest Dissertations and Theses. Robert Morris University.

* Kletenik, D., Butbul, A., Chan, D., Kwok, D., LaSpina, M., 2021. Game on: teaching Cybersecurity to novices through the use of a serious game. J. Comput. Sci. Coll. 36 (8).

* Lamour, J., 2008. Impact of user awareness and training of infosec practitioners on data security. In: Dissertation Abstracts International Section A: Humanities and Social Sciences, 68. ProQuest Information & Learning. Issues.

Loffler, E., Schneider, B., Asprion, P.M., Zanwar, T., 2021. CySecEscape 2.0-A virtual escape room to raise Cybersecurity awareness. Int. J. Serious Games 8 (1). https://doi.org/10.17083/ijsg.v8i1.413.

* Martin, J., 2019. Phishing in dark waters: a quasi-experimental approach with evaluating cyber-security training for End-users. ProQuest Dissertations and Theses. University of South Florida.

Mayhorn, C.B., Nyeste, P.G., 2011. Training users to counteract phishing. ProQuest Dissertations and Theses. North Carolina State University.

* McCrohan, K., Engel, K., Harvey, J., 2010. Influence of awareness and training on cyber security. J. Internet Commerce 9 (1). https://doi.org/10.1080/15332861.2010.487415.

* McKinney, G.B.J., 2021. Assessing corporate impacts of cyber training on email phishing attacks [Ed.D., Southern Nazarene University]. *ProQuest Dissertations and Theses* (2618814610). ProQuest Dissertations & Theses Global.

National Cybersecurity Alliance, & CYBSAFE. (2023). *The Annual Cybersecurity attitudes and behaviors report*. https://www.cybsafe.com/whitepapers/cybersecurity-attitudes-and-behaviors-report/.

* Ordonez, G., 2022. Comparing social engineering training in the context of healthcare [M.*Sc*., Purdue University]. *ProQuest Dissertations and Theses* (2838333188). ProQuest Dissertations & Theses Global.

Prümmer, J., Steen, T.van, Berg, B.van den, 2024. A systematic review of current cybersecurity training methods. Comput. Secur. 136, 103585. https://doi.org/10.1016/j.cose.2023.103585.

Puhakainen, P.P., Siponen, M., 2010. Improving employees' compliance through information systems security training: An action research study. MIS Q. 34 (4).

* Robbins, M.S., 2020. Exploring the impact of information security awareness training on knowledge, attitude, and behavior: A K-12 study. ProQuest Dissertations and Theses. Northcentral University.

Rogers, R.W., 1983. Cognitive and physiological processes in fear appeals and attitude change: a revised theory of protection motivation. Soc. Psychol.: Source Book 153–176.

* Shaw, R.S., Keh, H.C., Huang, N.C., Huang, T.C., 2011. Information security awareness on-line materials design with knowledge maps. Int. J. Distance Educ. Technol. 9 (4). https://doi.org/10.4018/jdet.2011100104.

Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L.F., Hong, J., Nunge, E., 2007. Anti-phishing Phil: the design and evaluation of a game that teaches people not to fall for phish. In: Proceedings of the 3rd symposium on Usable privacy and security. Association for Computing Machinery, pp. 88–99. https://doi.org/10.1145/1280680.1280692.

* Siponen, M., Puhakainen, P., Vance, A., 2020. Can individuals' neutralization techniques be overcome? A field experiment on password policy. Comput. Secur. 88. https://doi.org/10.1016/j.cose.2019.101617.

* Stefaniuk, T., 2020. Training in shaping employee information security awareness. Entrepreneur. Sustain. Iss. 7 (3). https://doi.org/10.9770/jesi.2020.7.3(26).

* Thornton, D., Turley, F., 2020. Analysis of player behavior and EEG readings in a Cybersecurity game. In: Proceedings of the 2020 ACM Southeast Conference, pp. 149–153. https://doi.org/10.1145/3374135.3385276.

* Tschakert, K.F., Ngamsuriyaroj, S., 2019. Effectiveness of and user preferences for security awareness training methodologies. Heliyon. 5 (6). https://doi.org/10.1016/j.heliyon.2019.e02010.

* van Steen, T., Deeleman, J.R.A., 2021. Successful gamification of cybersecurity training. Cyberpsychol. Behav. Soc. Networ.. https://doi.org/10.1089/cyber.2020.0526.

* Veneruso, S.V., Ferro, L.S., Marrella, A., Mecella, M., Catarci, T., 2020. CyberVR: an interactive learning experience in virtual reality for Cybersecurity related issues. In: Proceedings of the International Conference on Advanced Visual Interfaces. https://doi.org/10.1145/3399715.3399860.

Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Softw. 36 (3), 1–48.

* Waly, N.S., 2013. Organisational information security management: The impact of training and awareness: evaluating the socio-technical impact on organisational information security policy management. PQDT - UK & Ireland. University of Bradford (United Kingdom).

* White, G.L., 2022. Infrastructure cyber-attack awareness training: effective or not?. Int. J. Inf. Secur. Privacy 16 (1). https://doi.org/10.4018/IJISP.291702.

* Wu, T., Tien, K.-Y., Hsu, W.-C., Fu-Hsiang, W., 2021. Assessing the effects of Gamification on enhancing information security awareness knowledge. Appl. Sci. 11 (19). https://doi.org/10.3390/app11199266.

* Younes, W., 2014. Cybersecurity education (training and awareness) for K-12 faculty and staff in allegheny county. In: Dissertation Abstracts International Section A: Humanities and Social Sciences, 75. ProQuest Information & Learning. Issues.