# Beyond Technicalities:<br>Assessing Cyber Risk by Incorporating Human Factors

Wenjing Huang, Sasha Romanosky, Joe Uchill
RAND Corporation

DRAFT

March 2024

## Abstract

The process of assessing a firm's cyber risk most often relies on the technical characteristics of the firm's computing infrastructure, such as network configurations and software patching practices. While important, these approaches often ignore the human factors that affect a firm's cyber risk, such as an individual's knowledge and awareness, or their workplace setting. Research that has examined the correlation between human factors and cyber risk provides insights only into the *strength of the correlation* between these variables – what is often referred to in measurement theory as *validity*. Research has yet to address issues related to *reliability* – the other critical aspect of measurement -- such as how easily and consistently the variables can be collected in the workplace. Indeed, since most cyber incidents are the result of human failure, there becomes increasing urgency to understand, estimate, and *influence* the effect of individuals on a firm's overall security posture. This absence of consideration for human risk factors represents both a glaring omission, and an opportunity for better ways to measure and manage a firm's cyber risk. Therefore, this research seeks to fill this gap by creating a holistic approach to assessing cyber risks using modern psychometric techniques. We first identify the set of factors that have demonstrated empirical validity. We then survey cybersecurity experts and researchers to solicit their opinion about both validity and reliability of these variables based on their professional expertise, and we identify factors that would be most suitable either as predictors of human cyber risk, or opportunities by the firm to reduce cyber risk. Organizations can use these results to identify the human risk factors that may be most correlated with cybersecurity incidents, to assess the qualification of job candidates, and to assess the effectiveness of training programs for their employees.

Keywords: cybersecurity, psychometrics, validity, reliability, risk assessment

Acknowledgements: We would like to thank each of the respondents of our survey in addition to Masha Sedova, Jason Nurse, Julie Hanley, Matthew C., and RAND's Institute of Civil Justice and RAND's Feinberg Center for Catastrophic Risk for their generous support.

# Introduction

As the adage goes, humans are the weakest link in security.[1] Schultz (2005) argues that "information security is primarily a people problem, not a technical problem," and based on an analysis of 23,896 cybersecurity incidents, the 2022 Verizon Data Breach Investigations Report (VBIR) found an alarming proportion of breaches were the direct results of either deliberate or accidental human failure. Specifically, it found that employees falling victim of phishing attacks accounted for between 60% and 80% of breaches (VBIR, 2022). The World Economic Forum found that 95% of cybersecurity issues were the result of human error, and that 43% of all breaches were the result of either intentional or accidental employee behavior (WEF, 2022).

In addition, employees falling victim to business email compromise/email account compromise attacks led to a loss of nearly $2.4 billion in 2020 (FBI, 2021). These attacks use fraudulent emails to request wire transfers using deceptions of overdue invoices, or stories of managers stuck in foreign countries, and needing emergency funds. The security firm, SpyCloud (2021) analyzed 1.7 billion sets of stolen credentials from 755 datasets leaked in 2021 and found that 70% of victims had reused passwords. The concern being that these poor password behaviors make it even easier for hackers to compromise user accounts.

However, the process of assessing a firm's cyber risk is most often based on the technical characteristics of the firm's computing infrastructure, such as network configurations, security technologies, and software patching practices. Indeed, enterprise cyber risk scores provided by firms such as Bitsight, Security Scorecard, and UpGuard provide scores that are largely based on the technical security posture and configuration of a firm's public facing IP systems.[2] In addition, cyber insurance questionnaires focus almost entirely on technical security or data collection practices of the applicant (Romanosky et al, 2017). Ubiquitous security frameworks and standards such as those developed by NIST, DHS's Cybersecurity Infrastructure Agency (CISA), New York's Department of Finance (NYDFS), all speak to collections of hundreds of security controls that organizations may, or should implement.[3] While important, these approaches often ignore the human factors that drive a firm's cyber risk, such as individual security competency, and other individual or cultural factors.

In response to this disconnect, there has been a growing body of research that investigates the factors of human cyber risk, and their contribution to overall enterprise cyber risk. These efforts, which date back at least as far as 1996 (Zurko & Simon), provide insights only into the *strength* of the correlation (i.e. whether a given factor is statistically correlated with a harmful outcome or not), what is often referred to in measurement theory as *validity*. Unfortunately, this research has yet to address data *reliability*, such as how practically and consistently the data could be collected. That is, they lack application of formal

[1] The earliest published citation to this specific phrase was from Dhamija and Perrig (2000). However, a somewhat earlier citation illustrating the inevitable failure of humans can be found in Jeremiah 10:14, "Every man is stupid and without knowledge." And of course, the relative frailty of strong computer systems in the face of humans is exemplified in this cartoon, https://xkcd.com/538/, last accessed December 1, 2023.

[2] See https://www.gartner.com/reviews/market/it-vendor-risk-management-solutions/vendor/securityscorecard/product/security-scorecard-platform/alternatives. Last accessed December 1, 2023.

[3] See https://www.nist.gov/cyberframework, https://www.cisa.gov/sites/default/files/2023-03/cisa_cpg_checklist_v1.0.1_final.pdf, https://www.dfs.ny.gov/industry_guidance/cybersecurity, and https://content.naic.org/cipr-topics/cybersecurity, last accessed December 5, 2023.

psychometric methods necessary in order to construct a holistic approach to understanding human cyber risk, and ultimately measuring enterprise risk.

Despite these research efforts, and the clear recognition of the individual's role in cybersecurity, our understanding of the specific factors driving an individual's behavior and techniques to assess or improve these behaviors remain critically underdeveloped. Therefore, this research seeks to answer the following questions:

- Which factors does the existing literature find are most strongly correlated with individual cybersecurity risk?
- Which of these factors are most justified based on psychometric theory?
- Given this, what is the most appropriate framework for assessing human cyber risk?

Given the lack of formal analysis in this area, we believe these questions of fundamental research are necessary in order to better understand the problem and develop practical solutions. In order to improve employee security awareness and overall organization's security posture, different approaches may be needed depending on the firm's structure and industry, and may not require all the human related . Moreover, as we find in this research, not all factors identified in this report are appropriate for assessing an organization's security posture. One may only need a subset of these factors to use for different purposes, e.g. for evaluating effectiveness of specific training programs, or for assessing the overall workforce security posture.

## Related Literature

This work is informed by multiple bodies of literature. First, our work relates to research and cybersecurity frameworks that seek to measure and assess cybersecurity risk. Frameworks like the NIST Cybersecurity Framework, ISO-27001, DHS's Cyber Protection Goals, and the Payment Card Industry Data Security Standard (PCI-DSS) are examples of guidance that companies can use to compare their cyber security posture against best practices and industry requirements.[4]

Second, our research is informed by commercial software tools that measure and evaluate employee cyber risk. For example, Elevate security offers a diagnostic tool that collects data about employee behavior for the purpose of assessing and managing employee cyber risk.[5] CybSafe provides a software tool for firms to use to test both existing employees, and job candidates on their cybersecurity hygiene practices.[6] In addition to their commercial projects, CybSafe has also developed a Security Behavior Database which is a taxonomy of 148 security-related behaviors that could increase a firm's cybersecurity risk (e.g. whether the individual uses strong passwords, reuses passwords, or ignores or reports phishing emails, etc.).[7] Each of these entries is also accompanied by a risk score (from 1-4) with higher scores representing greater risk to the firm. SANS (an information security training company) has also developed a maturity model [8] to help organizations benchmark and improve the overall effectiveness of their information security awareness and training programs.[9]

[4] See https://www.nist.gov/cyberframework, https://www.cisa.gov/sites/default/files/2023-03/cisa_cpg_checklist_v1.0.1_final.pdf, https://www.pcisecuritystandards.org/, last accessed December 5, 2023

[5] See https://elevatesecurity.com/, last accessed June 1, 2023.

[6] See https://www.cybsafe.com/, last accessed June 1, 2023.

[7] See https://www.cybsafe.com/research/security-behaviour-database/ last accessed October 11, 2023.

[8] See https://www.sans.org/security-awareness-training/resources/maturity-model/, last accessed December 1, 2023.

[9] See https://www.sans.org/blog/sans-2022-security-awareness-report/, last accessed June 1, 2023.

We also leverage academic research related to understanding how laboratory interventions can best influence an individual's propensity to comply with corporate cyber security policies. For example, Dalal et al (2022) provide a meta-review of the human factors in cybersecurity, to include discussion of antecedents of human risk. They distinguish between individual (employee)-level factors and specific work environment factors, and illustrate how users individually, and collectively, influence cybersecurity outcomes, such as cybersecurity behaviors, and incidents.

Finally, we leverage the body of psychometric literature which provides a theoretical basis for evaluating existing cybersecurity assessment approaches, and for guiding the development of the framework in this research. Psychometrics is an important field in psychology, with the goal to develop psychology as a quantitative rational science. Initially defined as "the art of imposing measurement and number upon operations of the mind" (Galton, 1879, p149; Jones and Thissen, 2007), psychometrics as a field of study came into increasingly common use as psychology developed, reaching prominence as the name of the subdiscipline with the foundation of the Psychometric Society[10] in 1935 and the publication of Guilford's (1936) Psychometric Methods.

Psychometrics, now a common approach used widely in all behavioral and social sciences, provides a way to measure and quantify complex constructs whose outcomes are considered indicators of attributes of interest that cannot be directly observed (i.e. latent constructs such as personality and aptitudes). It involves the development, validation, and application of tests and other measurement instruments that connect observable phenomena (e.g., responses to items in an IQ-test) to theoretical attributes (e.g., intelligence) (Borsboom, 2005).

Psychometrically-sound tests are designed with the goals of being *reliable* and *valid*, meaning that they consistently measure what they intend to measure, and that they accurately reflect the traits being assessed. These tests can be used for a variety of assessment or prediction purposes, such as job candidates' capabilities, severity of psychological disorders, and student achievements. The outcomes of these assessments can be used for recruitment, clinical diagnosis and evaluating educational programs (Rao and Sinharay, 2007).

## Research Approach

Our analytic approach for developing a defensible framework for assessing human cyber risk consisted of the following steps. First, we conduct an extensive literature review in order to identify the factors[11] that have been shown to be statistically correlated with a cyber incident, or are used as controls in a broader regression analysis. This enables us to begin with a known, empirically tested set of factors.

Next, we incorporate insights from psychometric theory in order to adopt a common assessment for these factors based on *validity* and *reliability*, two widely used psychometric criteria. Most prior research used statistical methods to establish correlation between the factors and outcomes (which varies from one another), but few adopted proper psychometric techniques to establish measurement validity and

[10] See What is Psychometrics? - Psychometric Society, https://www.psychometricsociety.org/what-psychometrics, last accessed Dec 1, 2023.

[11] We use *factors* to refer to indicators, variables or elements that are potentially related to human cyber risk. We observed that these terms were used in the literature interchangeably without consensus. In social science literature, indicators are observed variables that are associated with underlying unobserved latent constructs (i.e. factors); elements could be observed or unobserved. Variables are usually observed. Here we use factors loosely to represents many terms as they include simple observed variable such as age and gender, as well as groups of variables (e.g. a set of conspicuous behaviors), unidimensional latent constructs (e.g. impulsiveness and workload), and more complex latent constructs (e.g. KSA's as in Knowledge, Skills and Abilities).

reliability of the proposed factors. We determine that the factors identified in literature provide some evidence of *validity* based on significant correlation, but lack a second important criteria, *reliability*. Therefore, in order to assess reliability, we surveyed 49 cybersecurity practitioners and cybersecurity researchers to evaluate each factor along four aspects of reliability: accessibility, interpretability, standardized measurement, and consistency.[12]

Based on an evaluation of these results, we present and discuss the factors that have both been shown to be valid (e.g. that have been statistically correlated with security outcome in our literature review or rated high on validity assessments in our survey) and meet a minimum threshold of reliability.

# Factors of Human Cyber Risk

In order to identify factors of human risk that are statistically correlated with cyber security incidents, we conducted an extensive literature review. We performed the review in a systematic manner, searching multiple databases for research papers relating to "cybersecurity human factors," "cybersecurity employee risk," and "cybersecurity human risk." Specifically, we searched an internal, multiplatform database aggregator searching EBSCO Academic Search Complete, ACM Digital Library, Google Scholar, IEEE Explore, JSTOR, SAGE Premier, Science Direct, Scopus, Taylor & Francis, and Web of Science. We then iteratively searched through the references in the papers we uncovered during our search.

We initially identified approximately 350 research papers meeting our criteria. We filtered out papers that were unrelated to our topic, resulting in 161 papers. We then further refined our criteria to papers based on predictive, empirical studies that employed quantitative methods with statistically significant results (rather than presenting hypothetical risks), resulting in a final set of 63 papers.

From those studies, we extracted variables used to predict insecure behavior, including poor security hygiene, susceptibility to social engineering, likelihood of participating in insider threat, et cetera.

Once the variables were identified as potential representation (i.e. factors) of human cyber risk, we then classified them based on the *agency of control* of the variable, producing three categories, *Individual*, *Workplace*, and *External threat*. *Individual* factors such as impulsiveness or financial debt are broadly controlled or influenced most directly by the individual. *Workplace* factors such as policies and practices and workload are both apparent and modifiable by the organization. *External threat*, including the how professional hackers value human targets, are not under the control of a workplace or the employee, but of a malicious agent.

Next, we describe in detail each of these factors grouped under the categories of *Individual*, *Workplace*, and *External threat*.

## *Individual*

### Demographic Traits

A review of literature about susceptibility to phishing by Baki & Verma (2023) aggregating 35 studies on the impact of gender and 29 studies on age, published between 2004 and 2020, suggested that women and younger people were statistically more vulnerable to online scams. Modic et al. (2011) also found that women were more vulnerable to scams, relative to men. That younger people were more vulnerable than

[12] While the issues of validity and reliability involve much discussion in the psychometrics literature, terms that could be formally tested and estimated, we use validity and reliability rather loosely as how they are interpreted in the cybersecurity literature and understood by security practitioners.

older people is a potentially counterintuitive result, given that older people are traditionally depicted in popular culture as more vulnerable.

A battery of surveys with different metrics for cybersecurity from Anwar et al. (2017), found that men reported higher computer and cybersecurity competence and experience than women, which may hint at mechanisms by which gender causes different security outcomes.

## Personality Traits

The so-called *Big 5* personality traits (openness, conscientiousness, extraversion, agreeability and neuroticism), are often used to predict workplace behavior, and in particular, cybersecurity behavior. However, the direction of effect found within past research bas been inconsistent. For example, these traits have shown stability in the short term (Cobb-Clark, 2012) and “moderate to high” (Rantanen, 2007) stability in the long term.

In eight papers comprising ten experiments testing Big 5’s effects on secure behavior, the results were mixed. Of the seven experiments dealing with Big 5’s impact on phishing, three (Modic et al., 2011, Enos et al., 2006, and Fraustein & Flowerday, 2020) showed that openness had significant positive impact on secure behavior. Three papers (Modic et al, 2011; Lawson et al., 2020, and Albladi et al,. 2017) showed that extraversion had a negative effect, two showed that conscientiousness had a negative effect (Halevi et al. 2015, and Fraustein & Flowerday, 2020), while one experiment showed that conscientiousness had a positive effect (Albladi et al., 2017). One experiment showed neuroticism had a negative effect (Halevi et al, 2013), while three showed the opposite (Fraustein & Flowerday, 2020; Albladi et al, 2017, and Enos et al, 2006.) and two showed a positive impact for agreeableness.

The inconsistency is likely at least in part due to the fact that adversaries can tailor a phishing email to appeal to different personality types – appealing to fear may lure neuroticism, but appealing to disappointment may lure conscientiousness. That specific set of results can be seen with the work of Halevi et al. (2013) who found that the only Big 5 personality trait impacting phishing compliance was neuroticism (and only in women), but in 2015 used a specifically crafted lure to phish users with conscientiousness.

In three experiments not dealing with phishing, people scoring high on openness were worse at securing their Facebook privacy security settings (Halevi et al., 2013), while conscientious people performed positively on a written test of security behaviors, and open and conscientious people were more motivated by compliance (Uffen et al., 2013).

Impulsiveness is a surprisingly multi-dimensional collection of traits that can be measured through different metrics, and has shown to be more consistently correlated with security behavior. The Barratt Impulsiveness Scales (BIS; Patton, Stanford & Barratt, 1995) is a mechanism to describe the attributes of impulsiveness, which includes motor impulsiveness (acting on the spur of the moment, without thinking), non-planning impulsiveness (considering the present but not the future) and attentional impulsiveness (not paying attention). Two studies that tested security behavior against the BIS (Hadlington, 2017; Aivazpour & Rao, 2022) found that motor and attentional impulsiveness both had a significant impact. A third study (Aivazpour & Rao, 2019) also found that motor impulsiveness led to poor privacy behavior. Another study tested only “cognitive impulsiveness,” which maps to non-planning, and found what it called “marginally significant results” (a p-value of 0.101) (Pattinson et. al, 2012).

Modic et al. (2011) monitored secure behavior based on four factors, as well as a general “self-control” metric. Self-control and premeditation (thinking through potential outcomes of a scam) correlated with secure behavior, and urgency (the drive to complete a task no matter the consequence) was inversely correlated. Surprisingly, they also found sensation seeking – risk seeking behavior – to be inversely correlated with secure behavior, which the researchers concluded that this group may be less intellectually curious than others. Premeditation was found to be similar to non-planning, but was the only study to find significant results.

An fMRI study conducted by Neupane et al. (2016) found that people presented with a malware warning (as would be issued by security software) activated the medial prefrontal cortex – a brain region less-active in those graded as impulsive using the BIS.

### Knowledge, Skills, and Abilities (KSAs)

Knowledge, skills and abilities (KSA) describe different aspects of expertise: knowledge refers to grasp of information, skills refers to the ability to apply that information to a mechanical task and abilities refers to innate or more general talent. For example, a designer might have knowledge of men's fashion, the skill of physically assembling a suit or customer service skills. Training is an attempt to provide minimal KSA, though workplace cybersecurity training is too infrequent to impart comprehensive knowledge, skill or ability.

The relationship between KSA and actual behavior is hazy, possibly for the same reasons that the relationship between training and behavior is hazy. Baraković and Baraković Husić (2023) found that knowledge, awareness and behavior only occasionally held significant correlations. "Computer and smartphone sharing is associated with cyber hygiene knowledge [in] some segments, i.e., knowledge regarding https, ransomware, and GPS. Password creation is associated with the correct answer regarding passwords, while password management is not," wrote the authors, who assessed a partial existence of a relation. Zwilling et al. (2022) found similarly mixed results, with knowledge related to installation of antimalware software but not to sharing of personal information. Studies from Egelman et al. (2015,2016) found that knowledge corelated with updating systems, using passwords to protect cell phones (in a time before that was a norm), and creation of strong passwords in real world settings and the ability to identify a phishing website in a lab setting. Cain et al. (2018) found no link between knowledge and self-reported history of breaches, and that self-reported experts were less likely to behave securely.

Tempestini et al. (2023) found no correlation comparing a knowledge test to a real-world behavior assessment grouping subjects into "no-," "low-" and "medium-risk" tiers. Cain et al. (2018) also found no link between knowledge and self-reported history of breaches, and interestingly, self-reported experts were less likely to behave securely.

The mixed results could potentially relate to tasks where intended behavior could be mitigated by lapses in focus and discipline. Password creation is a one-time task that requires active attention, while password management is a continuous process where users may take shortcuts. Sharing of personal information may be subject to lack of focus or discipline, while antimalware software might require only a single moment of discipline to download and install. Tests for KSA are not standardized and differ between studies; it is also possible that mixed results stem from poorly developed tests that are not designed with psychometrically-sound properties.

## *Workplace*

### Cognitive Load

Research has shown how cognitive load such as workload, workplace distractions including deadlines (Yeng et. al, 2022, and Chowdhury, 2019), total email quantity, and phishing email density (Sarno et al., 2021) are statistically correlated with people making security mistakes, both in lab and real-world settings. Workload creates two impediments to discipline (Chowdhury, 2019): People who are preoccupied with deadlines tend to behave more carelessly, while at the same time, complex tasks often coax people to take shortcuts to expedite processes (Khan, 2022). An early human factors study by Beautement et al. (2008) identified five potential costs in the cost/benefit analysis employees do to determine security compliance – four of five of which stem from workload or cognitive load.

Indeed, Chris Hadnagy, founder and former head of the largest social engineering competition in the world, observes that these distractions often determine human receptiveness to attacks, "[t]his is what social engineers prey on – that we're all busy. So when a guy who's sitting in his office, stressed – who maybe got in an argument with his wife last night, who on the way to work got a flat tire, and now his boss is chewing him out because he's got three reports due and a meeting in fifteen minutes – gets an email that says, "Boss wants you to read this Excel file on end-of-year budgets"" He'll just click it" (Uchill, 2015).[13]

## Training

Baki & Verma (2023) analyzed 28 papers regarding the effectiveness of training, and found overwhelmingly consistent positive correlation between training and security outcomes.[14] That said, only four of the studies estimated the effectiveness of training over time, and showed inconsistent results regarding the lasting effect of training over time. Further, in a review of 10 papers from the Baki & Verma (2023) review, seven (Martin, 2019; Wen et al., 2019; Gokul et al., 2018; Moreno-Fernandez et al., 2017; Lastdrager et al. 2017; Kunz et al., 2016; and Stockhardt et al., 2016) did not conduct real world testing for effectiveness, either relying on pre- and post-test questionnaires or other systems where users knew they were being tested. Users were therefore cued to focus and removed from the context of workplace distractions and stressors like cognitive load – situations where knowledge is not indicative of behavior.

In addition, Elevate Security (a cyber security training vendor) and the Cyentia Institute (a data security analysis company) found that in telemetry-based study of 114,000 employees across 2,000 organizational departments that employees acted inconsistently to training (Edwards, 2021). Employees who had a single course of training had a nearly identical click through rate on phishing emails (11%) compared with those with no training (12%.). Click through continued to drop after two (6%) and three (5%) trainings, before beginning to skyrocket after additional trainings (8% after three and 14% after four). These data become easier to conceptualize with a basic understanding of phishing training. Enterprises often assign additional training every time a user clicks on a real or simulated phishing email, meaning users are only trained until they stop clicking and some users continue to click no matter what the trainings. It is unclear from these tests if employees viewed the trainings as punishment and reformed their behavior to prevent punishment, or if employees legitimately learned new material based on additional training, though the fear appeals section shows only a temporary value to fear.

We also consider *fear appeals*, efforts by the firms meant to modify behavior by reminding an employee (even through reprimand) of the consequences of a cybersecurity event – including consequences to the company or society. Current research tends to focus on snapshot studies of how inciting fear affects security in the immediate short-term (Dupuis et al, 2022) and explaining how the fear mechanism works in those studies. Studies include theory of planned behavior (Guo & Yuan, 2012), protection motivation

[13] Indeed, Hadnagy, among the world's most knowledgeable experts on social engineering, was himself once phished on the way to his competition (Uchill, 2015).

[14] Note also that training was not standardized across research testing, encompassing everything from comics to games to mindfulness training (Baki & Verma, 2015). Training was also not standardized in format, application or content across different commercial training providers and different enterprises enforcing training.

theory (Boss et al., 2015; Posey et al., 2015) and deterrence theory (Warkentin, 2011). It is ultimately unclear what the mechanism is or what makes one appeal more successfully than another.

Fear appeals is often a component of training programs, but can also be introduced as part of warnings when people engage in risky behavior. While the immediate effects of inducing fear may help drive better cybersecurity behaviors, it is unclear how long the effect lasts. Dupuis et al. (2022) found a positive short-term result, stating, "Fear, as a short-lived emotion, can indeed be effective in the short term. Snapshot-like studies, like the one reported here, might lead us to conclude that fear is indeed indicated and efficacious. Yet, it may backfire in the long term due to the negative long term affects it can trigger." Here, "negative long term affects," may refer to multiple kinds of effects. It could refer to security fatigue (Stanton et al. 2016, Furnell & Thompson 2009), where the individual is so overwhelmed by security warnings and mitigations that they simply accept failure as inevitable, or it could refer to an employee's ill-will that has been a documented result of fear appeals (Reeves et al., 2023, Dupuis et al, 2022).

There are several known shortcomings of fear appeals research (Renaud & Dupuis, 2019), such as how studies were rarely based in the field, and how studies largely do not measure the amount of fear incited by an intervention. While surveys and lab cybersecurity tests have shown positive results, the broader literature around fear appeals has generated skepticism, including several studies and meta-analyses raising some doubt about effectiveness. One meta-analysis questioned whether fear appeals worked in health decisions (French et al., 2017), other researchers have shown that research into appeals homogenizes low and high efficacy groups who appear to respond differently (Kok et al., 2018), and a meta review by Floyd (2000) demonstrated that, in general, fear appeals were less successful than bolstering self-efficacy.

## Insider Threat

One major divide in risk attributes is the difference between the potential to accidentally enable a threat and becoming an intentional insider threat, when an employee steals from, or deliberately sabotages their employer. Intentional insider threat, the subject of this section – often just referred to as insider threat – could be predicted based on factors that, by our ontology, would be considered individual factors or workplace factors.

Individual factors related to insider threat may include the following. Several government agencies have produced guidelines for identifying potential indicators of insider threat behavior. For example, NATO's Cooperative Cyber Defense Centre of Excellence (CCDCOE, Kont et al, 2018) and CISA (2020) identify indicators at individual level, including what we will define as *financial situation* (e.g. debt and spending that exceeds income), *conspicuous psychological status* (e.g. depression) and *behaviors* (e.g. involvement with individuals or groups opposing the core beliefs of the organization, aggression and unwillingness to comply to rules), *major life changes* (e.g. marriage and divorce), and other *personal risk factors* (e.g. criminal record, addiction, etc.). Both CCDCOE and CISA make clear that many of these individual-level indicators could be innocuous on their own, with the majority of mental illness not resulting in criminality. Nurse et al also present a very comprehensive framework for understanding, and hopefully managing, insider threat (Nurse et al, 2014).

CCDCOE (Kont et al, 2018) insider threat guidance includes what we will call *conspicuous workplace behaviors* that may be indicative of ongoing malfeasance including frequent unexplained use of data copying equipment, volunteering for projects that would increase access to data, excessive overtime use of personal computer equipment in high security areas, or concerning statements. These are not meant to be causal factors – rather, they may be clues of active attempts to subvert a workplace. There are, however, substantial workplace causal factors as well. CCDCOE (Kont et al, 2018) mentions impending termination of contract (which CISA extends to include being passed over for promotion and other workplace slights).

*Workplace hostility* is a core reason for insider threat. 84 percent of insiders were motivated in whole or in part by a desire for revenge, with 57 percent identified by coworkers as being disgruntled (Keeney, 2005). Reasons for discontent include, firing, being passed over for promotion, demotion, workplace conflicts and policy disagreements, and perceived lack of respect.

While our ontology differentiates between attitudes, traits or behaviors associated with risk caused by the workplace and the longstanding attitudes, traits and behaviors that exist before an employee enters a company, these are not clean distinctions when discussing insider threat. CCDCOE (Kont et al, 2018) lists aggression, bragging, poor social interaction and unwillingness to comply to rules as indicators of insider threat – each of which may be caused by either a workplace issue (aggression due to interpersonal conflict) or an individual issue (aggression due to longstanding disposition).

### Adherence to Corporate Policies

Organizations and corporates can create a *culture* of security through formal or informal, enforced or unenforced rules (i.e. *policies* and *procedures*). These rules could mitigate intended or actual behavior. Alternatively, an enterprise could use technology to block employees from completing risky actions. For example, blocking social media from office computers dramatically reduces the odds a user can be phished over social media on the office network. Data on workload suggests that high-workload users often ignore rules that are not backed by technological blocks to increase efficiency. Both rules and blocks could be seen as mediating factors for potential risky behavior.

### Security Behaviors

If the underlying assumptions of this paper are true – that is, if certain individuals are higher risk than other individuals – then previous security events are likely to be predictive of future events. That is to say, the human risk factors that were a component of the first event are largely the same after the event. This would likely hold true whether an attack was successful or an event was detected and mitigated by security products or personnel.

With that in mind, data on employee behaviors aggregated by modern security products could be used as an estimator of human risk, particularly in comparing different employees behavior. Common security technologies include detection of employees going to unsafe websites, attempts to install disallowed software or hardware, evidence of ongoing breaches, unusual account behavior, delays in applying security updates or clicks on phishing messages (either legitimate or simulated emails sent to the real account without warning). Security intelligence offer services to detect password breaches on the dark web, and while leaked passwords from a breached third-party may or may not be evidence of personal culpability, evidence an employee was using corporate email accounts for personal accounts or reusing workplace passwords in other context is bad security behavior.

Indeed, CISA (2020) identifies a number of such behaviors as indicators of insider threat, for example, “direct correspondence with competitors, email messages with abnormally large attachments or amounts of data, Domain Name System (DNS) queries associated with Dark Web activities, use of activity masking tools (e.g., virtual private networks [VPN] or the Onion Router [Tor]),” and other related behaviors.

## *External threat*

A number of factors regarding the value of the real or externally perceived target employee may increase or decrease the likelihood an employee is targeted. Those include attackers targeting specific job titles, roles or industries (Janofsky, 2022, Mullen, 2018), people with specific external activities (including job seekers) (Breitenbacher & Osis, 2020[15]), or even specific nations during times of geopolitical strife (CISA, 2022; both threat intelligence firms and government agencies, including CISA, provide this threat intelligence information).

A user's level of access can also increase risk – all else being equal, an employee with more *access to critical data* or systems will have higher risk than one who does not as targets of *spear phishing or other similar attacks* directed at the employee.

## Summary of variables

Based on the literature review and discussion above, we summarize the identified variables below.

First, we consider that human cyber risk is a function of many characteristics of the *Individual*, such as their demographic characteristics (age, gender and race[16]), their experience, knowledge, and abilities regarding cybersecurity practices, as well as the individual's personality traits and beliefs. We consider that these variables are within the control of the individual. That is, it is primarily the individual, herself, who is able to most directly affect her personality, KSAs, life situation and risk factors.

Second, we recognize that human risk is a function of the individual's *Workplace*, such as the security culture formed either explicitly through annual training, or implicitly through corporate culture, norms, policy and other practices. In addition, cyber risk may be affected by the individual's job and related duties and stressors experienced at work. These data reflect conditions of the workplace, or behaviors implemented at the workplace, and so are within the primary control of the organization (rather than within the control of the end-user).

Third, we consider that human cyber risk is affected by *External threats* to the person and workplace. For example, criminals or nation-state actors may evaluate an employee based on their role to persuade the individual take action that would result in a cybersecurity incident. An employee may also be persuaded by an external threat actor to access a system without authorization, or exceed their authorization in order to access or corrupt information for personal or financial gain.

We summarize these factors in Table 1.

Table 1. Classification of factors affecting human cyber risk

### Individual

| | |
|---|---|
| **Demographic traits** | • Age<br>• Gender<br>• Race |
| **Personality traits** | • Big five Personality traits: extroversion, agreeableness, openness, conscientiousness, neuroticism<br>• Impulsiveness |

[15] See https://web-assets.esetstatic.com/wls/2020/06/ESET_Operation_Interception.pdf, last accessed December 7, 2023.

[16] Note that we did not find relevant quantitative study on race as a factor for assessing human cyber risk. We included it because it is routinely included in studies as control variables along with age and gender (i.e. the two variables we did find significant statistical results as reviewed in the Demographic Traits section).

| | |
|---|---|
| **Life circumstances and behaviors related to insider threat** | • Financial situation (debt; whether spending exceeds income; etc.)<br>• Standard conspicuous psychological status (e.g. depression) or behaviors (e.g. volunteering for suspicious activities or groups)<br>• Major life change (marriage, divorce, etc.)<br>• Personal risk factors (criminal record, addiction) |
| **KSA** | • Knowledge, Skills and Abilities (KSA) |
| **Workplace** | |
| **Cognitive load** | • Workload<br>• Workplace distractions (e.g. email volume, noise, activity) |
| **Training** | • Security awareness training (e.g. simulated phishing emails)<br>• Fear appeals |
| **Insider threat** | • Workplace hostility (job satisfaction / desire for revenge / office conflicts / passed over for promotion / fired / demoted/ etc.) |
| **Adherence to corporate policy** | • Organizational and corporate culture, policies and procedures |
| **Security behavior** | • Data on employee behaviors collected from security software, e.g. visited websites, malware infections, etc. |
| **External Threat** | |
| **Value of target employee** | • Global threat intelligence about the attacker groups and capabilities<br>• Access to privileged or classified information<br>• Spear phishing, or other similar targeted attacks directed at the employee<br>• Job Title / Role |

Source: RAND analysis

## Evaluating Factors of Human Cyber Risk

Now that we have identified factors from literature review that are correlated with human cyber risk, we next seek to evaluate the factors based on criteria informed by the field of psychometrics and security practitioners. This body of psychometrics research is primarily concerned with the measurement of latent variables – variables such as psychological attributes or traits that are hidden and cannot be directly observed, but can be inferred indirectly through a statistical model from other observable variables that can be directly observed or measured (i.e. manifest variables such as behaviors). Typical psychometric procedures focus on statistical modeling that relates latent variables to manifested variables, or latent variable modeling, including for example, classical test theory, structural equation modeling, and item response theory. A primary result of psychometric procedures is for developing standardized scales or assessment tools. While this process is purely based on quantitative analysis (e.g. see PROMIS as a national standard used in the clinical research community for a rigorously tested patient reported outcome measurement tool as demonstrated in Huang et al. (2016)), the steps that lead to a pool of candidate manifest variables are almost always a mix of qualitative and quantitative methods. For example, see examples of how to measure complex scenarios such as Barriers to Mental Health Care in the Military in

Acosta et al. (2018) and national worker's wellbeing (Chari et al, 2022) in CDC NIOSH's latest WellBQ.[17]

This body of research has used different definitions of the factors, different methods (a mix of quantitative and qualitative), and security outcomes (some simulated attacks, some real-world consequences and some lab-based vignettes). This makes it difficult to compare these results along a standardized scale regarding the validity of the factors studied. Therefore, we surveyed cybersecurity practitioners and academic cybersecurity researchers to seek expert opinion by giving them the list of factors which have been identified to be potential indicators of human related cybersecurity risk in the current literature that could lead to negative security outcomes.

Ideally, we would want to conduct a single research experiment and measure all variables at the same time in order to compare their relative strengths. However, we did not have access to data sources that covered all the relevant factors. The lack of empirical data made it impossible to conduct psychometric analysis to test these candidate factors. Hence instead, we relied on the expert survey results to create an initial candidate pool for future testing (i.e. more formal scale development) in order to build psychometrically-sound measurement tools. We discuss this limitation in the Discussion section of this research.

## Survey Instrument

Participants for the survey were selected according to their experience either as cybersecurity expert practitioners, or researchers studying in the field of cyber risk, cybersecurity, or human risk. We employed a non-probabilistic sampling approach (Guest et al, 2006) using a convenience sample of experts known to the authors as well as inviting self-identified cyber practitioners through social media (Mastodon and BlueSky). Overall, we received 49 completed survey responses.

The survey information was presented as a spreadsheet where participants were asked to score their *beliefs* about both the validity and reliability for each factor listed in Table 1, according to a 5-point Likert scale. See Appendix A for the instructions and factors listed in the spreadsheet.

Specifically, participants were instructed to evaluate each factor according to one measure of validity, and four measures of reliability on a 5-point Likert scale using the following criteria: "1 = worst, 2 = not so good, 3 = neutral/okay, 4 = good, and 5 = the best":

- Validity: does the variable measure what is meant to be measured (i.e. is it a good indicator for measuring human cyber risk)?
- Reliability: How precise and consistent can the variable be measured in regard to:
    - *Accessibility*: how easy can the data be observed or collected?
    - *Interpretability*: how easy is the variable to understand or explain?
    - *Standardized Measurement*: can the measurement be standardized or benchmarked against industry norm?
    - *Consistency*: can the data be consistently collected over time, under repeated circumstances?

In addition, participants were invited to include an overall assessment (again, based on their beliefs) about whether they would use the factor to measure human cyber risk, and to provide any final comments. Next, we provide the survey results. Average scores for validity and each of the four reliability variables are shown in Table 2.

[17] https://www.cdc.gov/niosh/twh/wellbq/default.html

## Quantitative Survey Results

| | Factors | Validity | Accessibility | Interpretability | Measurement | Consistency |
|---|---|---|---|---|---|---|
| **Individual** | **Age** | 2.4 | 4.8 | 4.2 | 4.6 | 4.7 |
| | **Gender** | 1.6 | 4.5 | 3.7 | 4.1 | 4.3 |
| | **Race** | 1.6 | 3.9 | 3.2 | 3.4 | 4.1 |
| | **Financial situation** | 3.0 | 2.3 | 3.1 | 3.1 | 2.9 |
| | **Standard conspicuous behaviors** | 3.0 | 2.0 | 2.2 | 1.9 | 1.8 |
| | **Major life change** | 2.8 | 2.4 | 2.9 | 2.6 | 2.4 |
| | **Personal risk factors** | 3.2 | 2.7 | 3.1 | 2.9 | 2.7 |
| | **Big five** | 2.4 | 2.6 | 2.5 | 2.5 | 2.5 |
| | **Impulsiveness** | 3.5 | 2.2 | 2.4 | 1.9 | 2.0 |
| | **KSA** | 3.4 | 3.6 | 3.4 | 3.2 | 3.6 |
| **Workplace** | **Workload** | 3.7 | 3.4 | 3.4 | 2.8 | 3.2 |
| | **Workplace distractions** | 3.7 | 3.4 | 3.3 | 2.7 | 3.1 |
| | **Workplace hostility** | 4.1 | 2.9 | 3.1 | 2.5 | 2.6 |
| | **Org. policies procedures** | 3.9 | 4.5 | 4.1 | 3.8 | 4.2 |
| | **Fear appeals** | 3.1 | 2.1 | 2.3 | 1.8 | 1.9 |
| | **Simulated phishing emails** | 3.0 | 4.6 | 4.1 | 4.1 | 4.4 |
| | **Data on employee behaviors** | 3.8 | 4.1 | 3.7 | 3.6 | 4.1 |
| | **Firm awareness training** | 3.2 | 4.6 | 4.3 | 4.1 | 4.6 |
| **External Threat** | **Global threat intel** | 3.8 | 3.1 | 3.5 | 2.9 | 3.1 |
| | **Access to privileged info** | 4.0 | 4.6 | 4.5 | 4.2 | 4.5 |
| | **Spear phishing attacks** | 4.1 | 3.7 | 4.2 | 3.5 | 3.6 |
| | **Job title-role** | 3.5 | 4.9 | 4.3 | 4.1 | 4.8 |

Table 2. Average validity and reliability scores

Note: darker green colors illustrate higher scores, while darker red scores reflect lower scores. Scores are based on a 5-point Likert scale where 1 = worst, 2 = not so good, 3 = neutral/okay, 4 = good, and 5 = the best.

The survey results generally agree with prior research. As can be seen, most of the individual factors scored quite low (less typically than 3 out of 5), relative to all other factors, both in regard to validity (i.e. strength of correlation with security incidents) and the reliability metrics. Notable exceptions include the demographic factors (age, gender, race) which were (understandably) considered to be quite reliability measurable, but not good indicators for predicting human cyber risk.

While three of the workplace factors (i.e. workload, workplace distractions and hostility) were considered to be quite valid, i.e. strongly correlated with security incidents, security experts did not feel that they could be reliability measured by the organization. Fear appeals stood out as scoring very low on overall reliability. How one follows organization policy and procedure is found to be a good indicator for human cyber risk and generally can be reliably measured. The two exceptions were results from simulated phishing emails and firm awareness training where both were considered to have good reliability but rather average validity rating. Contrary to popular beliefs, our experts did not find them to be very indicative of human cyber risk. We discuss this more in the next section.

Finally, with the exception of *global threat* intelligence, both the external threat (e.g. *job title and role*) and security behaviors (*data on employee behaviors*) factors were considered to be highly valid (i.e.

strong predictors of a security incident), and the data can be reliably collected by the organization using commercial cyber security monitoring soft wares and services.

These results are also displayed graphically in Figure 1.

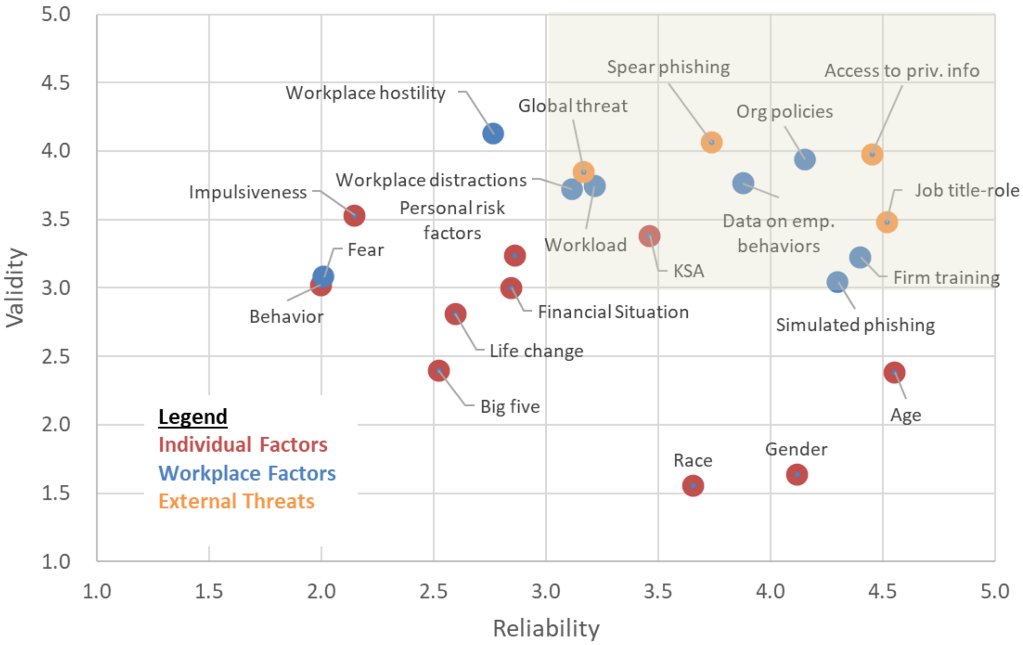


Figure 1. Validity and reliability of human risk factors
Note: The reliability value is the average of accessibility, interpretability, standard measurement, and consistency. The shaded area represents the set of factors that received an average score of 3 (neutral/okay) or greater for both validity and reliability.

In Figure 1 we consider factors that would meet a threshold of 3 or greater on both validity and reliability, shown within the shaded area, to be the ones that are sufficient indictors for assessing human cyber risk. Note that we consider (not require) a threshold of 3 as one approach for differentiating among high and low scoring factors.

Given the scores provided by our cyber security experts, only one individual factor (KSA) would be included, while all of the workplace factors would also be included (*Workload, Workplace distractions, Organization policies and procedures, Simulated phishing emails, Firm awareness training, and Data on employee behaviors*) except for *Workplace hostility and Fear appears*. Further, all external threat factors (*Global threat intelligence, Access to privileged information*, *Spear phishing attacks,* and *Job title-role*) were considered both valid and reliable. Of those factors outside the shaded area, both the ratings and comments from the experts (discussed more below) show that variables such as *Workplace hostility* and *Impulsiveness,* while considered to be valid, were considered too unreliable to be an appropriate measure.

For variables on the lower end of reliability (i.e. scoring less than 3), we would recommend exercising caution on collecting reliable data when using them as indicators of human cyber risk. For the variables that are low on validity (i.e. age, gender and race), we do not recommend using them as potential indicators for human cyber risk, but they could still be reliable data sources if there is interest in exploring how subgroups (e.g. young vs. old) differ in these other indicators that are being assessed.

## Expert Comments

In addition to the numerical scoring of each factor, respondents were invited to provide comments and reasoning behind their scores, which we summarize below.

### Individual Factors

In regard to the demographics and other personal matters, most respondents felt these factors were too intrusive, unethical, and “probably illegal.” On the other hand, some respondents felt that these factors, while difficult to obtain in some cases, could be worthwhile to consider as predictors. Other respondents commented that personal issues like addiction and criminal behavior could manifest by creating either *more* or *less* secure behavior by individuals. Another respondent noticed that factors regarding life changes don’t occur very often and would therefore become very difficult to use as predictors of insecure behavior.

In regard to the Big 5 factors, one respondent noted that such personality assessments are complicated enough that they require qualified people to administer them, making these factors more difficult to implement. One commenter noted that use of these factors is really just “pseudoscience,” while other respondents suggested that the traits were too susceptible to mood or emotional variation to be useful as a predictor of cyber security risk. Overall, most participants felt that the Big 5 personality measures were simply too difficult to measure effectively or reliably via self-reporting. These comments are in line with the expert ratings and supported by the mixed results from our literature review.

Regarding KSAs, one respondent noted that these factors are most useful for technical jobs (e.g. cybersecurity, information technology, etc.), rather than non-technical roles. In practice the value of KSA may be mediated by job title, access to data or other external risk factors – something that could hold true for all factors.

### Workplace Factors

In regard to workload factors, people questioned how the data could be collected and measured, but also noted that one person’s level of workload (or rather, the point at which someone feels overloaded) varies from person to person, and so calibrating this measure across all users in an organization becomes problematic.

In addition, one respondent mentioned the importance that leadership has in identifying and reducing workplace distractions and conflicts. Information could be collected from performance reviews and team-manager meetings, but that sharing the information beyond human resources would be problematic. One respondent suggested that any monitoring should be done, “for actions, not intent – we’re not Minority Report!” Another respondent questioned the value of collecting these data, hypothesizing that “sh[*]tty places to work are harder to secure.”

Another respondent noted that there is greater difficulty collecting accurate data from employees who pose the greatest risk because they are the ones most likely to conceal their beliefs or views, stating, “people don't like to admit they don't like their employer!” Though, on the other hand, the more hostile or disgruntled is an employee, the clearer would be signals of their discontent.

In regard to corporate policies, one respondent mentioned, “a real, un-scoped audit of controls would be invaluable, but I doubt anyone would allow that.” Another respondent mentioned that, “this is far & away the most important variable.”

In regard to corporate security awareness training, many respondents didn’t think that this was very effective at all (at least not for them), and can lead to a “false sense of security.” For example, in regard to simulated phishing testing, respondents were agnostic, with some feeling, “[t]his is potentially a great way to test employee resilience to induced risk. I'm not sure I've seen a capability of this sort that's done a great job testing this, however,“ while others felt it was “insane and a worst practice. Doing it destroys trust and increases risk.” Given there are various reasons one may fall for a simulated phishing email, three of our experts do not agree that it is a good indicator for human cyber risk at all, while others found it quite good, hence the average rating of 3 on validity. Despite the mixed feelings, one of the experts commented that “perhaps failing once or twice is understandable, it is those that are consistently failing the simulated phishing attacks (e.g. 4 or 5 times) that we should be more concerned about.”

### External Threat

In regard to threat, one respondent noted, “[y]ou can’t steal what you can’t access,” and “[a]ccess and title are proxies for the notion of consequence, which is a critical element of a risk calculation,” while another respondent mentioned the value of infrastructure security that pushes back the onus on the user to behave properly, stating, “[l]east privilege, compartmentalization, and reliable backup/recovery systems are the real deal here. By the time you're installing spyware and [anti-virus], you've already lost.” These insights further highlight the importance of firms implementing a diverse set of security controls in order to obviate opportunities for data and information compromises.

## Implementing the Framework for Assessing Human Cyber Risk

So far, we have presented a comprehensive framework for assessing human cyber risk. We discussed factors that are more useful than others according our survey and literature review. To our third research question *what is the most appropriate framework for assessing human cyber risk*, we believe the most appropriate assessment should be the one that is adapted and implemented for each enterprise with specific goal in mind (e.g. for finding insider threat or for improving employee’s overall security behaviors. Next, we discuss methods that combine modern psychometrics and machine learning (or AI more broadly) to help implement such framework in actual practice. There are important considerations regarding data collection and analysis, which we discuss next.

### Data Collection

A significant component of data reliability is the practicality of actual data collection. For example, we observe that many of the factors examined in past research efforts were not studied in real-world settings, but rather in a controlled, yet consented, lab setting.

**Individual-level data.** Demographic data can be reliably and repeatedly collected from standard human resource forms and corporate databases, however this brings obvious concerns. The inclusion of these factors as part of an assessment (evaluation) of human cyber risk is fraught with legal, ethical, and privacy issues, which should not be ignored (Bauer et al., 2020). Information about the individual’s KSAs can be collected from annual training and assessments, and may provide a proxy for the person’s true skills. Information about the individual’s personality traits and beliefs, however, are more difficult to reliably capture through existing inventories that rely on self-reports as data source. And so much more reliable measures would need to be developed before these individual factors could be appropriate for inclusion in an assessment of human cyber risk.

**Work Environment.** These data reflect conditions of the workplace, or behaviors implemented at the workplace, and are more easily observable by the organization. On the other hand, personal traits such as impulsiveness, as well as psychological status such as hostility and fear are harder to observe. Measurement of these latent variables often require self-reports, which introduces opportunities for biased reports, and is also subject to data availability and privacy challenges.

**External threat.** Most of the variables in this group (such as the person's title or role in the organization, or whether or not the individual has privileged access to important information) can be objectively and reliably collected. Variables within this group are controlled (affected) by the individual (she decides to change her job), the firm (granting access to privileged information), and external parties, such as attackers targeting the individual with spear phishing attacks.

In addition, we recognize that some variables, such as those related to insider threat, exhibit a particular feature in that they become strongly reliable when they are present, but strongly unreliable when absent. For example, with workplace hostility, it may be true that many people unusually upset with their company may not behave in a way that outwardly indicates they have become a risk. Hostility would not be a reliable indicator for those people. However an employee whose grievances have caused them to get into aggressive fights with management may be providing usable predictive, data, even if hostility is equally measurable across all employees.

## Data Analysis

Modern technology gives rise to big data from non-traditional sources and formats (e.g. social media interactions). How to process and analyze the vast and unstructured format of these data requires collaboration and creativity across disciplines (see discussions by Rauthmann, 2020). The developments in machine learning and large language models make it possible to analyze massive unstructured and text-based behavioral data that add to the traditional psychological research methods (Woo et al., 2020). However, not carefully analyzing the relationships between the massive amount of data may lead to poor interpretability of the results and harm predictive quality and validity. One potential cause is that big data often include groups of variables that are inherently related to each other due to shared underlying common *latent constructs*. These groups of data are also *noisy* in the sense that they come with measurement errors. It would be nice to condense them to a few dimensions as latent constructs before blending them with the other variables that clearly stand on themselves as direct representatives of the human factors (e.g. age can be a clearly-coded variable, whereas impulsivity has to be measured indirectly through asking a number of questions as in the Barratt Impulsiveness Scales). Here, we introduce the idea of pre-processing the measurement of these latent constructs using psychometrics before moving onto other methods such as traditional regression, ML or AI more broadly for causal analysis or prediction purpose.

For dealing with *latent constructs*, modern psychometrics such as Item Response Theory (IRT) can be applied to create standardized psychometric scales or to validate or scoring existing scales. IRT is based on the idea that the probability of a correct response to an item (e.g. a question on a test to assess one's math ability) is a mathematical function of the person (i.e. the respondent's underlying math ability) and item parameters (e.g. how easy or difficult the question is). There are a family of IRT models depending how to categorize them, e.g. the popular three parameter logistic model (3PL, see more details about IRT models in Thissen & Steinberg, 2009). There are many software applications that implement IRT. For example, a popular one for multidimensional IRT scale development and scoring is IRTPRO (Cai, Thissen & du Toit, 2011). When data to be analyzed are gathered using existing validated psychometric scales, one could simply follow the scoring instructions by the test developers (usually by taking the sum score), or one could use IRT to compute IRT scale scores for more refined measures (see more information on test scoring in Thissen & Wainer, 2001). For the factors without such existing instruments, we suggest creating new psychometrically-sound scales, (a process often referred to as "scale development or calibration") and appropriately validate the use of them among the population or the scenarios applied.[18] This calibration process involves estimating the item parameters in the IRT model, evaluating the item characteristics, and picking the best functioning items for defining the corresponding latent construct(s). See more details in Bjorner et al. 2007 for scale development. A product of this process is the factor scores generated for the latent constructs. The implementation these scales can also take the form of computer adaptive tests (Wainer et al., 1990) to expediate and automate the scoring process.

Latent variable modeling (or IRT more specifically) takes care of measurement errors of the latent construct by incorporating such errors directly into model estimation, leading to factor scores that are more reliable than the group of variables it was derived from. This process reduces the amount of (redundant) data down to a few more well-defined latent constructs, hence improving the overall quality of data to be analyzed in the following steps. With these pre-calculated latent factor scores along with the other variables that are not redundant, one then proceeds with traditional regression, or ML for causal inference or prediction purposes. With less amount of data used as predictors, the results would be more interpretable. A recent study by Wang et al. (2023) showed that Psychometrics can be integrated into the process as a first step for improving AI's predictive power, explanatory power, and quality assurance. This process could also be iterative in order to find the best assessment for human cyber risk. For example, one could use AI and Machine Learning to aid the psychometric development of standardized scales (e.g. Alexander et al, 2020 and Gonzalez, 2021).

# Discussion

We believe that this, and other research related to understanding the factors driving human cyber risk, can be applied in a number of important ways. First, in line with the bulk of research in this area, data regarding these factors can be used to *measure*, *assess*, and *predict employee risk*. For example, organizations like Elevate Security use related data in order to identify which employees are responsible for most security incidents within an organization. And tools like CybSafe use related data to gauge employee security awareness and maturity over time. The data can also be used (subject to important privacy, ethical and legal limitations) to estimate or evaluate a potential job candidate regarding their cyber security hygiene practices.

Proper cybersecurity risk assessment is also a critical component of cyber insurance underwriting. The better an insurance carrier is at measuring and differentiating cyber risk, the better it will be at pricing insurance policies according to the applicant's probability of a claimable cyber event. Indeed, based on available research, carriers rarely include human cyber risk as part of their underwriting process (Romanosky et al., 2017), though more recent research has shown progress in including human risk factors (Nurse, 2020). However, they are improving, and we see a legitimate opportunity for carriers to incorporate these factors when assessing (and tracking) a firm's cybersecurity risk.

Effective risk assessment is also the goal of cybersecurity frameworks and standards, such as NIST's Cybersecurity Framework, DHS's Cyber Performance Goals, the State of New York's Department of Financial Services Cybersecurity Program, and the National Association of Insurance Commissioners Model Data Security Law.[19] Currently, these frameworks provide either passing reference to employee cyber risk, or no mention at all. Again, we see a significant opportunity to incorporate human cyber risk

[18] Note some scales (e.g. corporate policy) may need to be developed and validated for each enterprise and could be scenario-specific.

[19] See https://www.nist.gov/cyberframework, https://www.cisa.gov/sites/default/files/2023-03/cisa_cpg_checklist_v1.0.1_final.pdf, https://www.dfs.ny.gov/industry_guidance/cybersecurity, and https://content.naic.org/cipr-topics/cybersecurity, last accessed December 5, 2023.

factors into these assessments, especially given the volume of cyber incidents resulting from human failures and mistakes.

In addition, the insights from this work can be used to identify and craft opportunities for *intervention* in order to reduce a firm's cybersecurity risk. That is, network managers can examine each of the factors in turn to determine which are potentially within their control to modify, and which are not. For example, as shown in Figure 2, the *Workplace* and *External Threat* factors are largely within the firm's control to affect, e.g. through better user education and awareness training, promoting workplace culture, improving adherence to policies and practices, as well as tailoring user access to critical information assets. We also recognize the interdependency between individual factors, workplace, and external threat given that they may each influence one another.

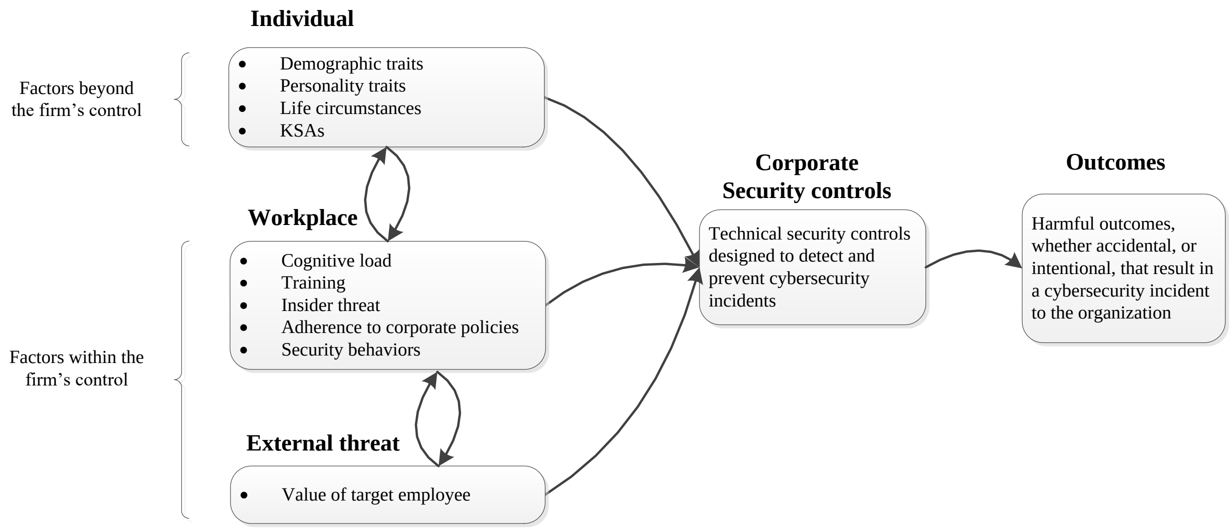

Individual
Demographic traits
Personality traits
Life circumstances
KSAs
Factors beyond the firm's control
Workplace
Cognitive load
Training
Insider threat
Adherence to corporate policies
Security behaviors
Factors within the firm's control
External threat
Value of target employee
Corporate Security controls
Technical security controls designed to detect and prevent cybersecurity incidents
Outcomes
Harmful outcomes, whether accidental, or intentional, that result in a cybersecurity incident to the organization

Figure 2. Opportunities for intervention

However, it is also just as important to recognize (and account for) the factors that are *not* within the firm's control. Clearly, many of the *Individual* factors are either driven by the individual themselves and/or therefore either completely exogenous to the firm (e.g. age, race), or may at best be only indirectly affected by the firm (e.g. financial circumstances or day to day emotional disposition).

Further, variables related to insider threat may not be relevant to firms which do not experience such concerns.[20] And so there would be no need for organizations looking to take advantage of this research to invest in resources to track and report on these factors.

## Limitations

This work has a number of limitations. First, the human factors that we collected are based on a review of existing literature, and corporate employee and personnel tools. While we believe we were rigorous in this search, it is possible that some important variables were omitted our analysis.

Further, the insights regarding reliability of these human factors are based on the beliefs and experiences of a sample of experts. While we made every attempt to reach out to a diverse group, our results may be biased, and so further research should be conducted to validate and replicate our results.

[20] While we are not aware of any such firms, we remain optimistic that such an exceptional organization exists.

We did not examine the correlation between these factors and the type of cyber incident they cause, nor did we distinguish between the average end-users vs. the intentional insider threat.

Actual individual-level data across multiple organizations were not available to us. However, if such data were available, we would focus on specific settings (e.g. looking at end-users at a certain type of organization such as government national security agencies) and collect all relevant data from human resource records, security control and monitoring software, and other self-reports on behaviors, psychological status and other latent variables such as awareness, intentions and corporate culture in order to develop psychometrically-sound instruments that can be applied to the targeted population.

We did not examine negative personality traits such as narcissism, Machiavellianism, and psychopathy (the so called *dark triad*) as predictors of human cyber risk, though we believe these could warrant further examination.[21] Dark triad research that appeared during our literature review was theoretical and not based on data. We also did not discuss the situation of employees with mental impairment that could be protected under the Americans with Disabilities Act (ADA). [22]

Finally, an important artifact of the human factor in cyber security is that "an individual's intention and behavior lead to outcomes that are unexpected and of a much larger magnitude than imagined" (Dalal et al, 2022). Indeed, as described above with users inadvertently being deceived by phishing attacks, the action may simply involve a single click on a link, or diligently complying with an urgent request. However, the outcome can cost the company millions of dollars (Romanosky, 2016).

Possible explanations for this disconnect between people's actions and outcomes has similarly been explored in behavioral privacy research, which seeks to explain why people disclose personal information much more eagerly online, compared to offline. Romanosky and Acquisti (2009) provide a number of explanations. First, the benefits of these transactions are often immediate (e.g. a click of a button enables the user to view the website, use the software, watch a video, or access a file), while the harms are: potentially not realized until much later (e.g. a hacker may not attack for months or years), intangible, or indirect. (e.g. costs are borne by the firm, not necessarily by the user committing the acts), or manifested as a probability rather than certainty (e.g. increased risk of future identity theft, or security breach).

To understand why people behave (seemingly irrationally) the way they do, we have to understand the intention and motivation behind such behaviors, but assessing these is challenging (e.g. asking people's intentions and motivations is awkward and likely won't generate true answers), hence this topic is beyond the scope of the current study.

# Conclusion

This research sought to answer three questions regarding evaluating the human factors related to cyber risk:

- Which factors does the existing literature find are most strongly correlated with individual cybersecurity risk?
- Which of these factors are most justified based on psychometric practice and theory?
- What is the most appropriate framework for assessing human cyber risk?

To answer these questions, we first performed an extensive literature review of research papers that seek to correlate human factors with cyber security outcomes, such as self-reported or observed unsecure

[21] See https://www.psychologytoday.com/us/basics/dark-triad, last accessed December 7, 2023.
[22] See Mental Health Conditions in the Workplace and the ADA at https://adata.org/factsheet/health

behaviors, sharing information over social networks, behaviors believed to lead to cyber incidents, propensity to be tricked by phishing emails, etc.. In addition, we fielded a survey in which we asked cybersecurity experts to provide their beliefs and comments regarding the factors already identified. We received 49 responses. Specifically, we asked them to evaluate 22 factors according to one measure of validity, and four measures of reliability: accessibility, interpretability, standardized measurement, and consistency.

Based on a threshold of 3 out of a 5-point likert scale, as shown in Figure 3, we find that only one of the individual factors (*KSA*) would be suitable for inclusion in a human cyber risk framework, while most workplace factors (*Workload, Workplace distractions*, adherence to *Organizational policies and procedures*, *Data on employee behaviors*, *Firm training*, and *Simulated phishing emails*) would be suitable. Finally, all of the external threat factors (*Global threat intelligence*, *Access to privileged information*, *Spear phishing attacks,* and *Job title-role*) would be suitable for a human cyber risk framework.

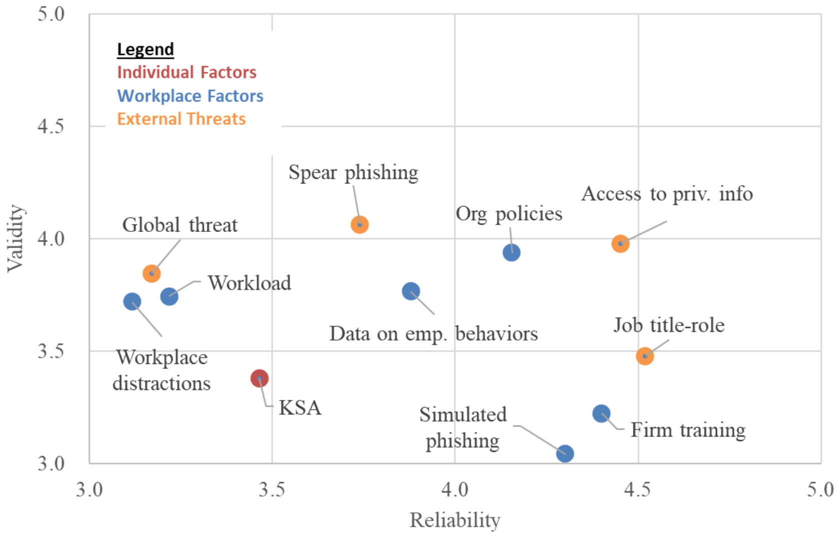


Figure 3. Most suitable factors related to human cyber risk

Overall, we presented a comprehensive framework for assessing human cyber risk, highlighting which factors are more useful than others, and discussed methods that combine modern Psychometrics with other methods such as regression, ML and AI for assessing human cyber risk. We believe that these insights can be used (or tailored) by companies and government agencies to better train and evaluate their workforce in order to avoid cybersecurity incidents, to design interventions to reduce risk, and to evaluate the overall effectiveness of such training and intervention programs.

We also believe that the most appropriate framework for assessing human cyber risk should be tailored for each enterprise and designed for each specific scenario applied. While technology made the availability of more data possible, the impact of its use on society is profound, leading to both benefits (e.g. exciting possibilities to understand human behaviors beyond what traditional behavioral sciences could ever reach) and risks (i.e. the extent to which such data can reveal about people's personal life makes violation of privacy a serious concern). Hence we believe that the methods used to handle these data in order to predict future human cyber risk require caution and the extent to which one could interpret the results should be handled carefully to prevent misuse and over-interpretation.

# References

Acosta, J.D., Huang, W., Edelen, M., Cerully, J., Soliman, S., & Chandra, A. (2018). *Measuring Barriers to Mental Health Care in the Military*. Santa Monica, CA: RAND Corporation (RR-1762-OSD).

Albladi, S.M. & Weir, G. R. S. (2017). *Personality traits and cyber-attack victimisation: Multiple mediation analysis*. 2017 Internet of Things Business Models, Users, and Networks, Copenhagen, Denmark, 1-6, doi: 10.1109/CTTE.2017.8260932.

Alexander, L., Mulfinger, E., & Oswald, F. L. (2020). Using Big Data and Machine Learning in Personality Measurement: Opportunities and Challenges. *European Journal of Personality*, *34*(5), 632-648. https://doi.org/10.1002/per.2305

Anwar, M., He, W., Ash, I., Yuan, X., Li, L., & Xu, L. (2017). Gender difference and employees' cybersecurity behaviors. *Computers in Human Behavior, 69*, 437–443. https://doi.org/10.1016/j.chb.2016.12.040

Aivazpour, Z., & Rao, V. S. (2019). Impulsivity and information disclosure: Implications for privacy paradox. *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Aivazpour, Z., & Rao, V. S. (2022). A replication study of the impact of impulsivity on risky cybersecurity behaviors. *AIS Transactions on Replication Research, 8*, Article 3.https://doi.org/10.17705/1atrr.00074

Baki, S., & Verma, R. M. (2023). Sixteen years of phishing user studies: What have we learned? *IEEE Transactions on Dependable and Secure Computing, 20*(2), 1200–1212. https://doi.org/10.1109/tdsc.2022.3151103

Bauer, T. N., Truxillo, D. M., Jones, M. P., & Brady, G. (2020). Privacy and cybersecurity challenges, opportunities, and recommendations: Personnel selection in an era of online application systems and big data. In S. E. Woo, L. Tay, & R. W. Proctor (Eds.), *Big data in psychological research* (pp. 393–409). American Psychological Association. https://doi.org/10.1037/0000193-018

Beautement, A, Sasse, A. M. & Wonham. M. (2008).The compliance budget: managing security behaviour in organisations. In Proceedings of the 2008 new security paradigms workshop, pp. 47-58.

Borsboom, D. (2005). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge University Press.

Boss, S. R., Galletta, D. F., Lowry, P. B., Moody, G. D., & Polak, P. (2015). What Do Systems Users Have to Fear? Using Fear Appeals to Engender Threats and Fear that Motivate Protective Security Behaviors. *MIS Quarterly (MISQ), 39*(4), 837–864. https://ssrn.com/abstract=2607190

Baraković, S., & Baraković Husić, J. (2023). Cyber hygiene knowledge, awareness, and behavioral practices of university students. *Information Security Journal A Global Perspective, 32*(5), 347–370. https://doi.org/10.1080/19393555.2022.2088428

Bjorner, J.B., Chang, C.-H., Thissen, D. & Reeve, B.B. (2007). Developing tailored instruments: item banking and computerized adaptive assessment. *Quality of Life Research, 16*, 95-108.

Breitenbacher, D., & Osis, K. (2020). *OPERATION IN(TER)CEPTION: Targeted Attacks Against European Aerospace and Military Companies*. Retrieved December 7, 2023 from https://web-assets.esetstatic.com/wls/2020/06/ESET_Operation_Interception.pdf,.

Cai, L., Thissen, D., & du Toit, S. H. C. (2011). IRTPRO for Windows. [Computer software]. Lincolnwood, IL: Scientific Software International.

Cain, A. A., Edwards, M. E., & Still, J. D. (2018). An exploratory study of cyber hygiene behaviors and knowledge. *Journal of Information Security and Applications, 42*, 36–45. https://doi.org/10.1016/j.jisa.2018.08.002

Chowdhury, N. H., Adam, M. T. P., & Skinner, G. (2019). The impact of time pressure on cybersecurity behaviour: a systematic literature review. Behaviour & *Information Technology, 38*(12), 1290–1308. https://doi.org/10.1080/0144929X.2019.1583769

Cobb-Clark, D. A., & Schurer, S. (2012). The stability of big-five personality traits. *Economics Letters, 115*(1), 11–15. https://doi.org/10.1016/j.econlet.2011.11.015

Chari, R., Sauter, S.L., Petrun Sayers, E.L., Huang, W., Fisher, G.G. & Chang, C.C. (2022). Development of the National Institute for Occupational Safety and Health Worker Well-Being Questionnaire. *Journal of Occupational and Environmental Medicine*, *64*(8), 707-717. doi: 10.1097/JOM.0000000000002585

CISA (2020). *Insider Threat Mitigation Guide (2020).* Cybersecurity Infrastructure Security Agency.

CISA (2022). *Russian state-sponsored and criminal cyber threats to critical infrastructure. (n.d.). Cybersecurity and Infrastructure Security Agency CISA*. Retrieved December 8, 2023, from https://www.cisa.gov/news-events/cybersecurity-advisories/aa22-110a.

Dalal, R.S., Howard, D.J., Bennett, R.J., Posey, C., Zaccaro, S.J., & Brummel, B.J. (2022). Organizational science and cybersecurity: abundant opportunities for research at the interface. Journal of *Business and Psychology*, *37*(1), 1-29. doi: 10.1007/s10869-021-09732-9.

Dhamija, R., & Perrig, A. (2000). *Déjà Vu: A User Study, Using Images for Authentication*. Proceedings of the 9th USENIX Security Symposium, Denver, Colorado, USA, August 14 –17, 2000.

Dupuis, M., Renaud, K., & Jennings, A. (2022). Fear might motivate secure password choices in the short term, but at what cost? *Proceedings of the Annual Hawaii International Conference on System Sciences*.

Edwards, B. (2021) *Elevating Human Attack Surface Management.* Cyentia Institute.

Egelman, S., & Peer, E. (2015, April). Scaling the security wall: Developing a security behavior intentions scale (sebis). In Proceedings of the 33rd annual ACM conference on human factors in computing systems 2873-2882.Egelman, Serge, Harbach, M  & Peer, E. (2016).Behavior ever follows intention? A validation of the Security Behavior Intentions Scale (SeBIS). In Proceedings of the 2016 CHI conference on human factors in computing systems,5257-5261.Enos, F., Benus, S., Cautin, R. L., Graciarena, M., Hirshberg, J.B., & Shriberg, E. (2006). *Personality Factors in Human Deception Detection: Comparing Human to Machine Performance.* Columbia University.

Floyd, D. L., Prentice-Dunn, S., & Rogers, R. W. (2000). A meta-analysis of research on protection motivation theory. *Journal of Applied Social Psychology, 30*(2):407– 429.

Federal Bureau of Investigation (2021). *Federal Bureau of Investigation Internet Crime Report 2021*.

Frauenstein, E. D., & Flowerday, S. (2020). Susceptibility to phishing on social network sites: A personality information processing model. *Computers & Security, 94*(101862). https://doi.org/10.1016/j.cose.2020.101862

French, D. P., Cameron, E., Benton, J. S., Deaton, C., & Harvie, M. (2017). Can Communicating Personalised Disease Risk Promote Healthy Behaviour Change? A Systematic Review of Systematic Reviews. *Annals of behavioral medicine: a publication of the Society of Behavioral Medicine, 51*(5), 718–729. https://doi.org/10.1007/s12160-017-9895-z

Furnell, S., & Thomson, K.L. (2009). Recognising and Addressing ‘Security Fatigue. *Computer Fraud and Security, 11*,  7–11.

Gokul, C.J., Pandit, S., Vaddepalli, S., Tupsamudre, H., Banahatti, V., & Lodha, S. (2018). Phishy-a serious game to train enterprise users on phishing awareness. *Proceedings of the 2018 annual symposium on computer-human interaction in play companion extended abstracts,* 169-181.

Gonzalez O. (2021). Psychometric and Machine Learning Approaches to Reduce the Length of Scales. *Multivariate behavioral research, 56*(6), 903–919. https://doi.org/10.1080/00273171.2020.1781585

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18,*59–82.

Guilford, J.P. (1936). *Psychometric Methods*. New York, NY: McGraw-Hill Book Company.

Guo, K. H., & Yuan, Y. (2012). The effects of multilevel sanctions on information security violations: A mediating model. *Information & Management, 49*(6), 320–326. https://doi.org/10.1016/j.im.2012.08.001

Hadlington, L. (2017). Human factors in cybersecurity; examining the link between Internet addiction, impulsivity, attitudes towards cybersecurity, and risky cybersecurity behaviours. *Heliyon*, 3, e00346. doi: 10.1016/j.heliyon.2017. e00346

Halevi, T., Lewis, J., & Memon, N. (2013). A Pilot Study of Cyber Security and Privacy Related Behavior and Personality Traits. *Proceedings of the 22nd International World Wide Web Conference*, 737–744.

Halevi, T., Memon, N., & Nov, O. (2015). Spear-phishing in the wild: A real-world study of personality, phishing self-efficacy and vulnerability to spear-phishing attacks. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2544742

Huang, W., Stucky, B.D., Edelen, M.O., Tucker, J. S., Shadel, W.G., Hansen, M., & Cai, L. (2016). Calibration of the Spanish PROMIS® smoking item banks. *Nicotine & Tobacco Research, 18*(7), 1635-1641. DOI 10.1093/ntr/ntw005.Insider Threat Mitigation Guide (2020). CISA.

Janofsky, A. (2022). *Iran-linked cyberspies expand targeting to medical researchers, travel agencies. Therecord.Media*. Retrieved December 8, 2023, from https://therecord.media/iran-linked-cyberspies-expand-targeting-to-medical-researchers-travel-agencies

Jones, L. V., & Thissen, D. (2007). *A history and overview of psychometrics.* In C.R. Rao and S. Sinharay (eds). Handbook of Statistics, 26: Psychometrics (pp.1-27). Amsterdan: North Holland.

Keeney, J.D., Kowalski, E., Cappelli, D., Moore, A., Shimeall, T., & Rogers, S. (2005). *Insider Threat Study: Computer System Sabotage in Critical Infrastructure Sectors*. Carnegie Mellon Software Engineering Institute.

Khan, N., Houghton, R..J., & Sharples, S. (2022). Understanding factors that influence unintentional insider threat: a framework to counteract unintentional risks. *Cognition, Technology & Work, 24*(3), 393–421. https://doi.org/10.1007/s10111-021-00690-z

Kok, G., Peters, G.-J. Y., Kessels, L. T. E., ten Hoor, G. A., & Ruiter, R. A. C. (2018). Ignoring theory and misinterpreting evidence: The false belief in fear appeals. *Health Psychology Review, 12*(2), 111–125. https://doi.org/10.1080/17437199.2017.1415767

Kont, M., Pihelgas, M., Wojtkowiak, J., Trinberg, L, & Osula. A. (2018). *A. NATO CCDCOE Insider Threat Detection Study*. NATO CCDCOE.

Kunz, A., Volkamer, M., Stockhardt, S., Palberg, S., Lottermann, T., & Piegert, E. (2016). Nophish: evaluation of a web application that teaches people being aware of phishing attacks. *Informatik 2016,* 509-518.

Lastdrager, E., Gallardo, I. C., Hartel, P., & Junger, M. (2017). *How effective is anti-phishing training for children?* Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017), 229–239.

Lawson, P., Pearson, C. J., Crowson, A., & Mayhorn, C. B. (2020). Email phishing and signal detection: How persuasion principles and personality influence response patterns and accuracy. *Applied ergonomics, 86*, 103084. https://doi.org/10.1016/j.apergo.2020.103084

Martin, J. (2019). *Phishing in Dark Waters: A Quasi-Experimental Approach with Evaluating Cyber-Security Training for End-Users*. USF Tampa Graduate Theses and Dissertations.

Modic, D., & Lea, S.E. (2012). How Neurotic are Scam Victims, Really? The Big Five and Internet Scams. *Law & Humanities eJournal*. DOI:10.2139/ssrn.2448130

Moreno-Fernández, M. M., Blanco, F., Garaizar, P., & Matute, H. (2017). Fishing for phishers. Improving Internet users' sensitivity to visual deception cues to prevent electronic fraud. Computers in *Human Behavior, 69*, 421–436. https://doi.org/10.1016/j.chb.2016.12.044

Mullen, J. (2018). Top CFOs are being targeted by a sophisticated email scam. CNN. Retrieved on December 9, 2023 from https://www.cnn.com/2018/12/04/tech/london-blue-email-hackers/index.html

Mumford, G. (2009, September 1). Preventing cyber attacks. *Monitor on Psychology*, *40*(8). https://www.apa.org/monitor/2009/09/cyber-attacks

Neupane, A., Saxena, N., Maximo, J. O., & Kana, R. (2016). Neural markers of cybersecurity: An fMRI study of phishing and malware warnings. *IEEE Transactions on Information Forensics and Security, 11*(9), 1970–1983. https://doi.org/10.1109/tifs.2016.2566265

Nurse, J., Buckley, O., Legg, P. A, Goldsmith, M., Creese, S., Wright, G. R.T., & Whitty, M. (2014). *Understanding Insider Threat: A Framework for Characterising Attacks*. 2014 IEEE Security and Privacy Workshops, San Jose, CA, USA, 2014, pp. 214-228, doi: 10.1109/SPW.2014.38.

Nurse, Jason RC, Louise Axon, Arnau Erola, Ioannis Agrafiotis, Michael Goldsmith, and Sadie Creese. "The data that drives cyber insurance: A study into the underwriting and claims processes." In 2020 International conference on cyber situational awareness, data analytics and assessment (CyberSA), pp. 1-8. IEEE, 2020.

Pattinson, M., Jerram, C., Parsons, K., McCormac, A., & Butavicius, M. (2012). Why do some people manage phishing e-mails better than others? *Information Management & Computer Security, 20*(1), 18-28. https://doi.org/10.1108/09685221211219173

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of clinical psychology, 51*(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:6<768::aid-jclp2270510607>3.0.co;2-1

Posey, C., Roberts, T. L., & Lowry, P. B. (2015). The impact of organizational commitment on insiders' motivation to protect organizational information assets. *Journal of Management Information Systems, 32*(4), 179–214. https://doi.org/10.1080/07421222.2015.1138374

Rantanen, J., Metsäpelto, R. L., Feldt, T., Pulkkinen, L., & Kokko, K. (2007). Long-term stability in the Big Five personality traits in adulthood. *Scandinavian journal of psychology, 48*(6), 511–518. https://doi.org/10.1111/j.1467-9450.2007.00609.x

Reeves, A., Calic, D., & Delfabbro, P. (2023). Generic and unusable": Understanding employee perceptions of cybersecurity training and measuring advice fatigue. *Computers & Security, 128*,. https://doi.org/10.1016/j.cose.2023.103137

Renaud, K., & Dupuis, M. (2019). *Cyber security fear appeals: Unexpectedly complicated*. Proceedings of the New Security Paradigms Workshop, 42–56.

Rao, C. & Sinharay, Sandip. (2007). *Handbook of Statistics, Vol. 26: Psychometrics*.

Romanosky, S. (2016). Examining the costs and causes of cyber incidents. *Journal of Cybersecurity, 2*(2), 121-135. https://doi.org/10.1093/cybsec/tyw001

Romanosky, S., Ablon, L., Khuen, A., & Jones, T. (2017) Content Analysis of Cyber Insurance Policies: How Do Insurance Companies Price Cyber Risk?, *Journal of Cybersecurity*, *5*(1), 1-19.

Rauthmann, J. F. (2020). A (More) Behavioural Science of Personality in the Age of Multi–Modal Sensing, Big Data, Machine Learning, and Artificial Intelligence. *European Journal of Personality*, *34*(5), 593-598. https://doi.org/10.1002/per.2310

Sarno, D. M., & Neider, M. B. (2021). So Many Phish, So Little Time: Exploring Email Task Factors and Phishing Susceptibility. In Human Factors: *The Journal of the Human Factors and Ergonomics Society, 64*(8), 1379–1403. SAGE Publications.

Schultz, E. (2005). The human factor in security. Computers & Security, 24(6), 425-426.

SpyCloud (2022). *Annual Identity Exposure Report: 2022*.

Stanton, B., Theofanos, M. F., Prettyman, S. S., & Furman, S. (2016). Security Fatigue. *IT Professional, 18*(5), 26–32. https://doi.org/10.1109/mitp.2016.84

Stockhardt, S., Reinheimer, B., Volkamer, M., Mayer, P., Kunz, A., Rack, P., & Lehmann, D. (2016). *Teaching Phishing-Security: Which Way is Best? In ICT Systems Security and Privacy Protection* (pp. 135–149). Springer International Publishing.

Tempestini, G., Rovira, E., Pyke, A., & Di Nocera, F. (2023). The Cybersecurity Awareness INventory (CAIN): Early Phases of Development of a Tool for Assessing Cybersecurity Knowledge Based on the ISO/IEC 27032. *Journal of Cybersecurity and Privacy, 3*(1), 61–75. MDPI AG. Retrieved from http://dx.doi.org/10.3390/jcp3010005

Thissen, D. & Steinberg, L. (2009). Item response theory. In R. Millsap & A. Maydeu-Olivares, *The Sage handbook of quantitative methods in psychology* (Pp. 148-177). London: Sage Publications.

Thissen, D. & Wainer, H. (Eds) (2001). *Test Scoring*. Mahwah, NJ: Lawrence Erlbaum Associates.

Uchill, J. (2015, July 31). *Chris Hadnagy on the Def Con hackers posing as your coworkers. Christian Science Monitor*. Retrieved December 9, 2023 from https://www.csmonitor.com/World/Passcode/2015/0731/Chris-Hadnagy-on-the-Def-Con-hackers-posing-as-your-coworkers

Uffen, J., & Breitner, M. H. (2013). *Management of Technical Security Measures: An Empirical Examination of Personality Traits and Behavioral Intentions*. 2013 46th Hawaii International Conference on System Sciences (HICSS). IEEE.

Verizon (2022). *Verizon Data Breach Investigations Report: 2022*.

Wang et al. (2023). *Evaluating General-Purpose AI with Psychometrics*. Published by Microsoft. https://www.microsoft.com/en-us/research/publication/evaluating-general-purpose-ai-with-psychometrics/

Wainer, H., Dorans, N. J., Green, B. F., Steinberg, L., Flaugher, R., Mislevy, R. J., & Thissen, D. (1990). *Computerized adaptive testing: A primer*. Lawrence Erlbaum Associates, Inc.

Warkentin, M., Johnston, A. C., & Shropshire, J. (2011). The influence of the informal social learning environment on information privacy policy compliance efficacy and intention. *European Journal of Information Systems: An Official Journal of the Operational Research Society, 20*(3), 267–284. https://doi.org/10.1057/ejis.2010.72

Wen, Z. A., Lin, Z., Chen, R., & Andersen, E. (2019). *What.Hack: Engaging anti-phishing training through a role-playing phishing simulation game.* Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

World Economic Forum, (2022). *The Global Risks Report 2022, 17th Edition*. Retrieved December 9, 2023 from https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2022.pdf, last accessed December 1, 2023.

Woo, S. E., Tay, L., & Proctor, R. W. (Eds.). (2020). *Big data in psychological research*. American Psychological Association. https://doi.org/10.1037/0000193-000

Yeng, P. K., Fauzi, M. A., Yang, B., & Nimbe, P. (2022). Investigation into Phishing Risk Behaviour among Healthcare Staff. *Information, 13*(8), 392. https://doi.org/10.3390/info13080392

Zurko, M. E. & Simon, R. T. (1996). User-centered security. In Proceedings of the 1996 workshop on New security paradigms, 27-33.Zwilling, M., Klien, G., Lesjak, D., Wiechetek, Ł., Cetin, F., & Basim, H. N. (2022). Cyber security awareness, knowledge and behavior: A comparative study. *Journal of Computer Information Systems, 62*(1), 82–97. https://doi.org/10.1080/08874417.2020.1712269.

# Appendix A

## Survey Instrument

Below are the instructions and scoring for each variable along a 5-point Likert scale.

### Instructions

We are policy researchers on human behaviors and cyber security . We are working on a project to measure employee cyber risk. That is, the risk that employees pose to a firm (rather than technical cybersecurity controls). We have identified a number of variables that may be correlated with employee cyber risk, which are listed in the worksheet in Colum B. We would like you to score each variable according to the metrics described using the following 5-point Likert scale:

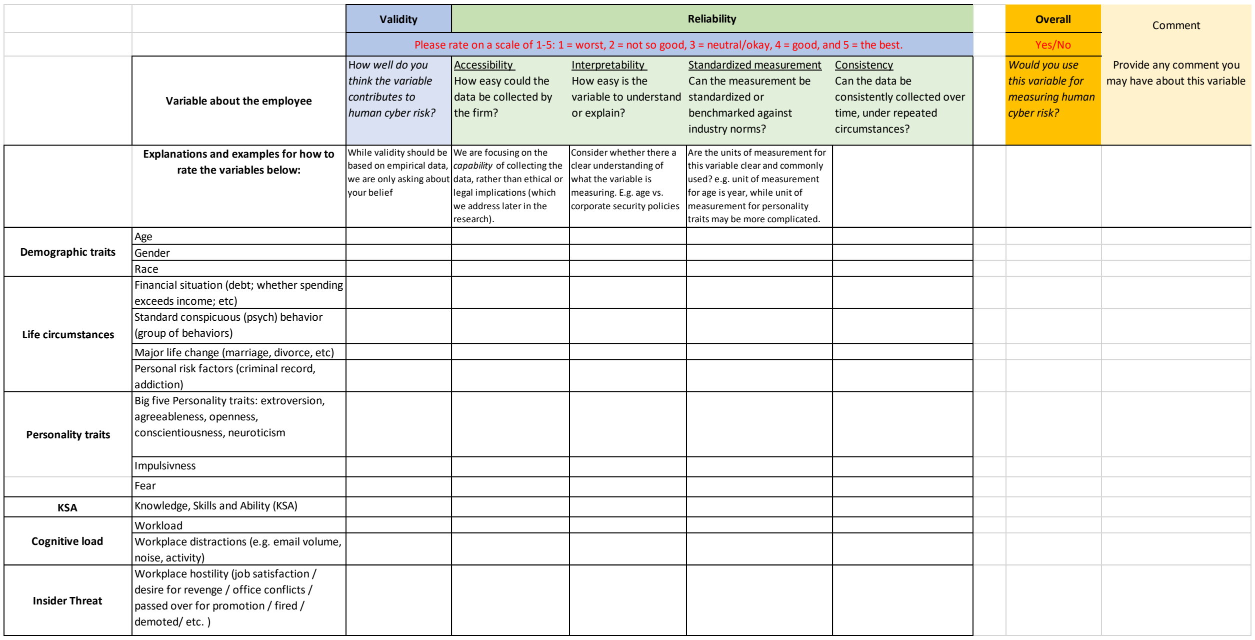
1 = worst
2 = not so good
3 = neutral/okay
4 = good
5 = the best

In addition, we included an "Overall" metric to allow you to provide a broad sense of whether you feel the metric would be useful as a potential metric for measuring employee cyber risk (Yes/No). While you may not know the correlation to human cyber risk for certain, we ask that you provide your best estimate of your beliefs. There is also a Comment column, which you can use to provide any comments, if you like.

Thank you for your participation!

### Scoring

The spreadsheet provided the following variables, and instructions:



| | | Validity | Reliability | | | | | Overall | Comment |
|---|---|---|---|---|---|---|---|---|---|
| | | Please rate on a scale of 1-5: 1 = worst, 2 = not so good, 3 = neutral/okay, 4 = good, and 5 = the best. | | | | | | Yes/No | |
| | **Variable about the employee** | *How well do you think the variable contributes to human cyber risk?* | Accessibility<br>How easy could the data be collected by the firm? | Interpretability<br>How easy is the variable to understand or explain? | Standardized measurement<br>Can the measurement be standardized or benchmarked against industry norms? | Consistency<br>Can the data be consistently collected over time, under repeated circumstances? | | *Would you use this variable for measuring human cyber risk?* | Provide any comment you may have about this variable |
| | **Explanations and examples for how to rate the variables below:** | While validity should be based on empirical data, we are only asking about your belief | We are focusing on the *capability* of collecting the data, rather than ethical or legal implications (which we address later in the research). | Consider whether there a clear understanding of what the variable is measuring. E.g. age vs. corporate security policies | Are the units of measurement for this variable clear and commonly used? e.g. unit of measurement for age is year, while unit of measurement for personality traits may be more complicated. | | | | |
| **Demographic traits** | Age | | | | | | | | |
| | Gender | | | | | | | | |
| | Race | | | | | | | | |
| **Life circumstances** | Financial situation (debt; whether spending exceeds income; etc) | | | | | | | | |
| | Standard conspicuous (psych) behavior (group of behaviors) | | | | | | | | |
| | Major life change (marriage, divorce, etc) | | | | | | | | |
| | Personal risk factors (criminal record, addiction) | | | | | | | | |
| **Personality traits** | Big five Personality traits: extroversion, agreeableness, openness, conscientiousness, neuroticism | | | | | | | | |
| | Impulsivness | | | | | | | | |
| | Fear | | | | | | | | |
| **KSA** | Knowledge, Skills and Ability (KSA) | | | | | | | | |
| **Cognitive load** | Workload | | | | | | | | |
| | Workplace distractions (e.g. email volume, noise, activity) | | | | | | | | |
| **Insider Threat** | Workplace hostility (job satisfaction / desire for revenge / office conflicts / passed over for promotion / fired / demoted/ etc. ) | | | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Corporate policy** | Organizational policies, procedures, and technical security controls | | | | | |
| | Firm-provided security awareness training | | | | | |
| **Value of target employee** | Global threat intelligence about the employee | | | | | |
| | Access to privileged or classified information | | | | | |
| | Spear phishing, or other similar targeted attacks directed at the employee | | | | | |
| | Job Title / Role | | | | | |
| **Secure behavior** | Simulated phishing emails, or other company-initiated tests | | | | | |
| | Data on employee behaviors collected from security software, eg. Visited websites, malware infections, etc. | | | | | |