The Influence Of Security Related Stress And Self-Efficacy On Actual Security Behaviors Over Time

Early Stage Paper

September 15, 2024

Seth Hastings
Tandy School of Computer Science
College of Engineering & Computer Science
The University of Tulsa
seth-hastings@utulsa.edu

Tyler Moore
School of Cyber Studies
College of Engineering & Computer Science
The University of Tulsa
tyler-moore@utulsa.edu

Bradley Brummel
Department of Psychology
University of Houston
bjbrummel@uh.edu

Sal Aurigemma
Department of Information Technology Management
University of Hawai'i at Manoa
sa8@hawaii.edu

Abstract

Researchers have long postulated that individual differences can affect performance on cybersecurity-related tasks. They have compared constructs such as Security Related Stress (SRS), New General Self-Efficacy (NGSE), and Security Related Self-Efficacy (SRSE) to intended behaviors, typically measured through surveys. Measuring intended behavior rather than observed behavior presents many drawbacks. In this paper, we explore the predictive relationships between these constructs and multi-factor authentication (MFA) metrics derived from 115 users' authentication logs and, we found significant relationships between SRS and decreased success authenticating, as well as increased time away following failed authentication attempts, and increased time spent locked out from digital resources. We found that low NGSE is associated with less successful authentication, but surprisingly we also observed that the highest NGSE users did not perform significantly better than their moderately effective peers. In line with our hypotheses, low NGSE correlates with a significant increase in instances of lock-out from digital resources compared to moderate NGSE users.

1 Introduction

A growing body of research has investigated the human factors associated with Information Security Policy (ISP) compliance and performance by measuring compliance intention, as real compliance performance data is hard to acquire. One body of such research focuses on Security Related Stress, as developed by D'Arcy et al., 2014. This and subsequent work explores security related psychological constructs and how these individual differences influence self reported ISP Compliance intention. While we discuss related work in greater detail in the next section, this study directly answers the call issued by several papers:

- "...the research would be strengthened by a longitudinal design with a lag between the collection of the dependent and independent variables or through measures of actual ISP violations obtained from independent sources" (D'Arcy et al., 2014, p. 307).
- "Even though it is plausible to assume that behavioral intention (i.e., compliance intention) can predict actual behavior, future research should consider measuring actual behaviors to clearly establish the relationship between information security-related technostress and information security compliance" (Hwang and Cha, 2018, p. 290).
- "Opportunity 4: Cybersecurity scholars should seek to study novel stress outcomes" (Singh et al., 2023, p. 115).

Despite the perceived value of measuring actual security behaviors, most research thus far has not gathered such data. Warkentin and Mutchler, 2014, January refer to this as the "holy grail" of behavioral research in their chapter on behavior information security management, a helpful reference that surveys the theories and methods applied to behavioral information security research. A recent meta-analysis analyzed studies that measure the relationship between self-efficacy and security behavior (Borgert et al., 2024). Out of 52 peer-reviewed studies investigating behavior in the meta analysis, only one measured actual performance with security tasks. Kwak et al., 2020 tasked participants with identifying phishing emails in a laboratory setting, not in the field.

In this paper, we do measure actual security behavior in the form of multi-factor authentication activities in an enterprise setting. We examine the relationship between Security Related Stress,

New General Self-Efficacy, Security Related Self-Efficacy, and observed user MFA performance across a seventeen month period. We developed a dataset of organic user authentication events to serve as ground truth, rather than relying on users self-reported experience with MFA. Gaining insight into these patterns and relationships could inform targeted interventions, improve usability, and aid in identification of compromised accounts. This work, while preliminary, offers a novel analysis of the relationships between psychology and multi-factor authentication performance.

2 Related Work

Security-related stress (SRS) has been a focal point in understanding the relationship between psychological factors and information security compliance. D'Arcy et al., 2014 established security related stress as a second order construct made up of security related Overload, Uncertainty, and Complexity. This construct establishes a relationship between stressors and information security policy (ISP) violation intention. Their study identified key stressors such as security demands, overload, complexity, and uncertainty, which contribute to SRS.

Moody and Galletta, 2015 expanded this research by exploring the impact of stress on online information retrieval performance. They proposed an "inverted-U" relationship, where moderate stress levels could potentially enhance performance, while both low and high stress levels negatively affect it. Their findings highlighted the significance of time constraints and information scent in influencing user stress and performance.

Ament and Haag, 2016 provided an empirical test of a multidimensional construct of security-related stress, revealing mixed effects on ISP compliance intentions. They introduced different stressors, including invasion of privacy and job insecurity, showing how these factors collectively contribute to overall security-related stress. In the same year, Lee et al., 2016 investigated the impact of work overload and privacy invasion as stressors in information security stress (ISS). They found that work overload significantly influences ISS, particularly in technical security-oriented organizations. Attitudes toward ISP compliance, prior security knowledge, and perceived security

threats were all identified as mitigating factors.

Belk et al., 2017 examined the difference in authentication performance across authentication devices for users categorized as either field-dependent or field-independent. They highlighted that visual perceptiveness, or the ability to interpret the surrounding environment by processing information in visible light, plays a significant role in authentication performance.

Hwang and Cha, 2018 focused on the role of technostress creators and role stress in employees' information security compliance. Their results indicated that technostress negatively impacts compliance by diminishing organizational commitment, with promotion focus moderating the relationship between technostress and role stress.

Furthering this, D'Arcy and Teh, 2019 examined the daily variability of security-related stress and its impact on ISP compliance, emphasizing the role of emotions in the coping process. They highlighted how certain stressors, categorized as hindrance stressors, can deplete employee resources leading to negative outcomes such as psychological strain and reduced compliance. Maier et al., 2019 explored how personality traits influence the perception of technostress. They identified that IT mindfulness, a dynamic trait, significantly moderates technostress perceptions, suggesting that some users are more susceptible than others to stress induced by security demands.

Nasirpouri Shadbad and Biros, 2020 provided a comprehensive overview of technostress, identifying five key stressors: techno-overload, techno-invasion, techno-complexity, techno-uncertainty, and techno-insecurity. They found that these stressors lead to adverse outcomes such as reduced productivity, ISP non-compliance, and discontinued IT use. Cram et al., 2021 introduced the concept of security fatigue, identifying its antecedents and consequences on ISP compliance. They described security fatigue through symptoms such as frustration, tiredness, and hopelessness, which significantly impact employee compliance behaviors.

Kim et al., 2022 used eye-tracking technology to study the impact of technostress on cognitive load. Their study differentiated between low-stress and high-stress individuals, showing that high-stress participants exhibited more distractions and slower task completion times.

Jeon et al., 2023 focused on the emotional responses of employees to security policy compli-

ance, particularly the role of frustration. They found that frustration negatively impacts compliance, but this effect can be mitigated by providing autonomy to employees.

Finally, there are three recent meta-analyses on this topic. Yuan et al., 2023 investigated the effects of specific technostressors on strain and job performance, including techno-complexity, techno-insecurity, and techno-uncertainty. The results revealed that techno-complexity and techno-insecurity were significant predictors of both strain and job performance. Employees facing high levels of these stressors experienced increased strain and decreased job performance. Interestingly, the study found that techno-uncertainty did not have a significant impact on job performance, suggesting that not all technostressors equally affect employees. Their paper also highlighted the moderating roles of demographic factors, such as age and job experience, in shaping the relationship between technostressors and job outcomes. They concluded that tailored interventions considering these demographic factors could help mitigate the negative effects of technostress on employees.

Concurrently, Aggarwal and Dhurkari, 2023 conducted a comprehensive meta-analysis to investigate the association between stress and information security policy (ISP) non-compliance intention. They found a weak positive correlation between stress and ISP non-compliance, indicating that higher stress levels are associated with a slight increase in non-compliance. Notably, the study emphasized the role of demographic characteristics such as age, country, and employment status in moderating this relationship.

Lastly, Singh et al., 2023 provided a systematic review of the literature on stress in the cybersecurity profession, focusing on the appraisal process of security demands and the outcomes of stress beyond mere compliance. Their review identified unique stressors faced by cybersecurity professionals, such as constant exposure to high-stakes security threats and the pressure to maintain vigilance against potential breaches. Their call to action highlighted the need to go beyond compliance metrics and consider the holistic impact of stress on cybersecurity professionals.

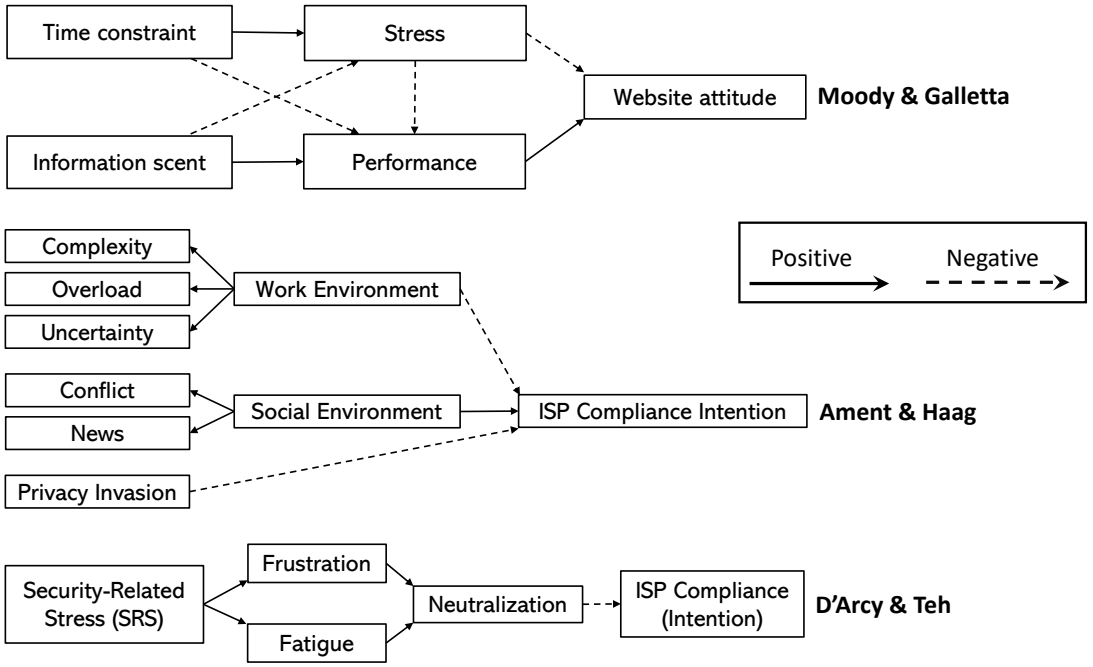

Figure 1: SRS-related Behavioral Models

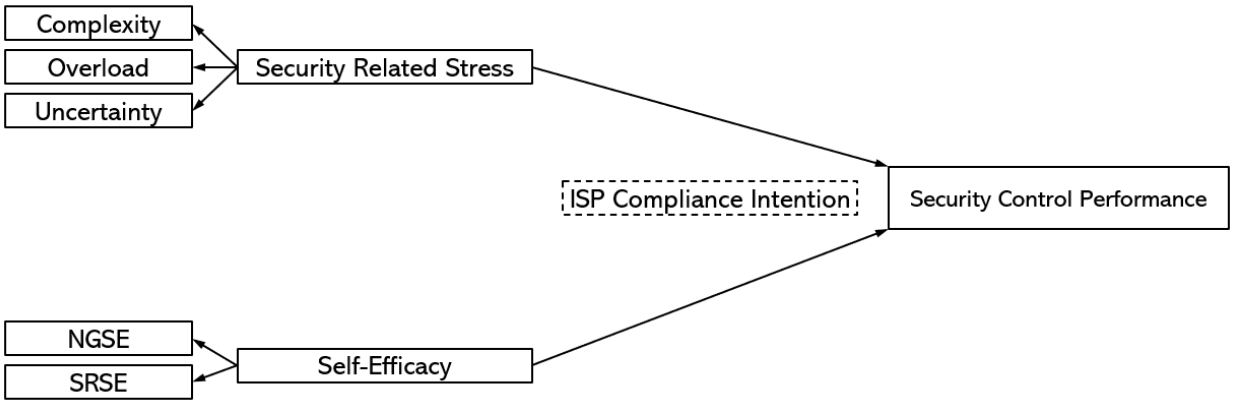

Figure 2: This Study

Figure 1 shows the models used in this research by Ament and Haag, D'arcy and Teh, and Moody and Galletta. Across these different models of the impact of stress on cyber security, the main commonality is that they predict attitudes of intentions with the link from intention to future behavior left to be assumed beyond the model. Our work builds on this growing body of research through a longitudinal study of actual security control performance derived from Azure sign-in

logs at a University, as shown in Figure 2. We compare self-reported stress and efficacy measures to observed security control performance without relying on self-reported compliance intentions.[1] This enables a more direct measure of any links between stress, self-efficacy, and security control performance.

3 Methodology

3.1 Psychological measures

Measures We take constructs measuring security-related Overload, Complexity, and Uncertainty from D'Arcy et al., 2014. New General Self-Efficacy (NGSE) and Security Related Self-Efficacy (SRSE) were adapted from Chen et al., 2001 and Compeau and Higgins, 1995 respectively. The Overload, Complexity, and Uncertainty constructs are relatively new (circa 2014), and the authors were unable to find a study where they were used in analysis of authentication performance.

The two efficacy measures are designed to capture an individual's belief in their own abilities. Computer Self-Efficacy was adapted as Security-Related Self-Efficacy, for example SRSE Item 1 reads:

> "Regarding the use of 2FA for my [EDU] accounts, we could configure and use 2FA. . . if there was no one around to tell me what steps to follow."

And the Computer Self-Efficacy item it was adapted from reads:

> "I could complete my job using the technology if...there was no one around to tell me what to do."

We similarly adapt the other items without making substantive changes to the wording. The NGSE items and Security Related Stress construct items are used verbatim. In related work we reference studies that look at both state, or situational stress and trait-like representations of stress; our study

[1] The constructs used in this study were selected and collected prior to the formation of the security control performance measure.

assumes that the SRS variables have trait-like properties attributable to the individual, which is necessary to predict over time.

Data source Between October 12, 2020, and January 18, 2021, 167 people at the author's university completed an IRB approved survey of psychological variables and attitudes towards MFA, 162 of the participants chose to participate in the study. The survey was collected via Qualtrics, and was composed of items for the five referenced psychological constructs, several additional items on security policy at [University], and user sentiment about MFA recently after its mandatory roll-out. The ability to compare user differences from survey data to observed network behavior affords a unique opportunity to draw connections between psychology and organic security control performance.

Constructs are examined as superscores, averages across all items within a user and construct. The Security Related Stress constructs are 1-7 Likert scale responses, and NGSE is Likert 1-5. SRSE questions were posed as binary response with a follow up confidence range of 1-10 for affirmative answers. "False" responses were coded as "0".

3.2 Authentication performance metrics

Data source We collect authentication events data from the author's university between November 8, 2021, and March 1, 2023. Follow the methodology developed by Hastings et al., 2024, we define events as the occurrences reflected in log data that users directly experience, beginning when an authentication to a particular application is initiated, and terminated upon the eventual success, failure, or abandonment (> 600 second lapse of activity) of the authentication attempt. By extracting these events from raw authentication logs, we can measure the interactive components of authentication while reducing the noise in the raw data, such as applications accessing resources or non-interactive authentications occurring in the background.

In total, this dataset includes 24,326 complete authentication events across 4 semesters from 115 network users who participated in the survey, with an average of 53 events per user per

semester. After filtering for users who were active across all semesters, the study is left with 19,515 events from 111 users with an average of 44 events per user each semester. Events are single row representations of complete interactions. Attributes are fairly intuitive, including the elapsed time, result, application being authenticated to, form of authentication used, types of errors encountered, and more. In the dataset used for analysis, we summarize these events over monthly time periods for each user, and describe the specific metrics next.

Measures We developed several performance metrics to capture not only the success users have with authentication, but also the amount of errors they encounter, and the associated time costs to a user or organization.

- Success Rate: The number of successful events divided by the number of total events for a particular user within a Period.
- Success Rank: Success Rate over a given Period relative to peers (least successful user ranked 1)
- Elapsed Time: The mean time per event in seconds across a given Period and user[2].
- Days Locked Out: The number of days within each month that a user could not successfully authenticate to any service. We require two or more consecutive, separate, failed authentication events resulting in over 6 hours unauthenticated to consider a user locked out.
- Time Away (TA): The time in minutes between a failed authentication event and the next attempted authentication; summed over the full month Period within a user. This is another measure of time cost to the user and organization.
- Friction: An error rate; the number of errors for a user in a given Period divided by the number of events they had in that Period.
- Period: An integer index variable tracking which monthly time period a given user observation is associated with.

Descriptive statistics for variable are shown and discussed in Section 4.1.

[2]Note that this captures the time between the first row of data associated with an event and the last row of the event.

Table 1: Hypothesized Relationships

	Success Rate	Success Rank	Elapsed Time	Timey Away	Days Locked Out	Friction
NGSE	H1a: +	H1b: +		H1c: -	H1d: -	
SRSE	H2a: +	H2b: +		H2c: -		H2d: -
Overload	H3a: -	H3b: -		H3c: +	H3d: +	H3d: +
Complexity	H4a: -	H4b: -		H4c: +	H4d: +	
Uncertainty	H5a: -	H5b: -	H5c: +	H5d: +		

3.3 Research Hypotheses

As prior work examines relationships with compliance intention, rather than actual behavior, we hypothesized with fresh eyes; hypotheses may diverge from the expectations of prior work. The anticipated relationships between psychological constructs and response variables are shown in Table 1. New General Self-Efficacy (NGSE) measures the confidence someone has in their ability to be successful in their daily lives and overcome challenges. Given this, we expect those with higher NGSE will overcome errors more often, and have a higher Success Rate and Success Rank, relative to their peers. Similarly, those with greater confidence are more likely to seek help when they can't log in, resulting in lower Time Away and fewer Days Locked Out.

Security-Related Self-Efficacy (SRSE), measures the confidence someone has to succeed with technical security controls. We expect someone with greater SRSE to use security controls more proficiently, leading to the same positive relationships as NGSE. Since SRSE is specific to security controls, and not a general efficacy measure, we don't necessarily expect someone with higher SRSE to be more likely to seek help when locked out.[3] Similarly to NGSE, we expect someone with high SRSE to have lower Time Away, as they are more confident overcoming security related challenges. Unlike NGSE, we expect those with higher SRSE to be relatively lower in Friction, a measure of the frequency of errors encountered. This reduction in Friction is expected to come from a reduction in user errors relative to a low SRSE individual, and prior work indicates that the vast majority of authentication errors are user errors.

[3] A review of the events causing lock-outs shows a vast majority of errors are configuration errors. Thus, we expect someone's proficiency to have little bearing on their chances of getting locked out.

Overload, a measure of the user's perception of excessive demands placed upon them by security controls, is expected to have negative impacts on performance. We hypothesize that higher levels of Overload is associated with lower Success Rates due to the increased cognitive burden leading to more frequent mistakes and reduced perseverance in resolving errors. Consequently, users experiencing high Overload are expected to exhibit higher Time Away. Additionally, Overload is likely to result in more Days Locked Out and higher Friction rates, as the strain from excessive security demands leads to more frequent errors and failures.

Complexity captures contexts in which security requirements require significant time or effort to learn and understand. While multi-factor authentication may be a new experience for some users, its usage is relatively static; consequently, we don't expect a great difference in raw performance for users who have higher security related complexity. As perceived complexity may drive the level to which a user engages with the security control, high complexity users may also be more prone to seeking compensatory tools, such as a password manager, to offload some of the burden. With those considerations, no hypotheses were made about the relationship with success rate or fortitude. Instead, we hypothesize that users with high complexity will also have longer Time Away, as they may expend more time or effort to address a failure. Similarly, we hypothesize a positive relationship between complexity and how long or often a user is locked out of their account.[4]

Uncertainty measures the user's perception of the unpredictability and lack of easy understanding related to security controls, policies, and procedures. We expect higher levels of Uncertainty is associated with lower success rates and higher mean elapsed time, as users may be less confident in their ability to navigate the authentication process, leading to mistakes and longer time spent authenticating.

[4]Complexity was not observed to have significant relationships throughout analysis, so we omit it from discussion for brevity

4 Analysis

4.1 Summary Statistics

As we move into analysis, we describe our independent variables in Table 2 and response variables in Table 3. Examining our independent variables, we note large correlations between SRSE and NGSE, and similarly sized inverse correlation with Overload and Complexity. NGSE shows inverse correlations with all three SRS constructs, and Overload has large positive correlations with Uncertainty and Complexity.

	Mean	SD	Correlations				
			1.	2.	3.	4.	5.
1. SRSE	7.82	2.20	**.94**				
2. NGSE	4.17	.57	.32	**.86**			
3. Overload	2.90	1.29	-.32	-.19	**.87**		
4. Uncertainty	3.93	1.15	-.06	-.08	.30	**.83**	
5. Complexity	3.54	1.06	-.30	-.25	.58	.11	**.73**

Table 2: Means, Standard Deviations, Correlations, Cronbach's Alpha for Independent Variables.

	Mean	SD	Correlations							
			6.	7.	8.	9.	10.	11.	12.	13.
6. Success Rate	.94	.18	**.76**							
7. Success Rank	49.3	13.9	.74	**.63**						
8. Elapsed Time	37.2	151.1	-.11	-.10	**.56**					
9. Days Locked Out	17.7	19.4	.27	.54	-.03	NA				
10. Time Away(hrs)	64.88	238.6	-.16	-.18	.03	-.19	**.31**			
11. Friction	.08	.19	-.71	-.54	.16	-.20	.29	**.73**		
12. Period	7.04	4.06	-.27	-.28	.08	-.21	.11	.21	NA	

Table 3: Means, Standard Deviations, Correlations for Monthly Period Response Variables, Cronbach's Alpha on the diagonal[6]

Moving to our response variables, we first notice the large correlation we anticipated between Success Rate and Success Rank. Success Rate has a similarly large inverse correlation with Fric-

[6]calculated using the first 10 months of data to limit missing-ness and avoid the influence of the observed natural experiment in SP23

tion, as failures driven by the errors experienced during authentication, and both relationships are echoed by the Success Rank variable. Next, we see a positive correlation between Elapsed Time and Friction, and an inverse relationship between both Days Locked Out and Time Away with Friction. These results are largely intuitive, but the positive relationships between Success Rate and Rank with the negative performance metric Days Locked Out are puzzling.

	1. SRSE	2. NGSE	3. Overload	4. Uncertainty	5. Complexity
6. Success Rate	-0.03	**-0.11**	**-0.11**	-0.02	**-0.06**
7. Success Rank	-0.03	-0.00	-0.03	-0.05	-0.03
8. Elapsed Time	0.03	0.01	0.00	0.01	-0.01
9. Days Locked Out	0.02	0.05	0.02	-0.00	0.05
10. Time Away(hrs)	**-0.12**	-0.08	0.02	-0.00	0.00
11. Friction	-0.03	0.01	**0.08**	-0.05	**0.08**

Table 4: Correlation Results between Independent and Response Variables, correlations significant at the 0.05 level in bold font

Finally, we examine the correlations between Independent and Response variables in Table 4[7]. We observe significant inverse correlation between Success Rate and NGSE, Overload, and Complexity. The relationships with Overload and Complexity are intuitive, as those stressors increase, authentication success would naturally decrease. The inverse relationship with NGSE is counter-intuitive, as we expect those with higher generally self-efficacy to perform in line with their elevated confidence. We explore this result more in later sections.

Time Away has a significant inverse correlation with SRSE; users with higher Security Related Self-Efficacy are correlated with less Time Away after authentication failure, which matches our intuition. Friction had significant correlations with both Overload and Complexity. Friction is a simple measure of errors per event; this suggests as a user has increasing Security Related Overload or Complexity, they experience more errors.

[7]We omit the Period variable from these correlations, as the various self-reported construct superscores were collected at a single point

4.2 Single Predictor Regressions

A series of single predictor regressions were conducted to evaluate our hypotheses against within user averages across construct items we call construct superscores. Single item regressions were performed using authentication event data aggregated within users across a monthly time period.[8] Natural log transforms were used for both the construct averages and response variables, enabling an intuitive reading of each beta values as an elasticity.[9] Using hypothesis **H3c** in Table 5 as an example: a .88 beta value means a 1% increase in Security Related Overload is associated with in a .88% increase in mean Time Away after failure per month. We set .05 as our threshhold significance for hypotheses support, and bold results that reach significance when listing hypotheses.

Simple regressions supported five of our twenty-one hypotheses. Two additional hypotheses were inversely related but significant: NGSE shows a negative relationship with Success Rate and positive relationship with Days Locked Out. Users with higher NGSE had lower success rates and had more days in which they were locked out of digital systems. Three Overload relationships were supported: Success Rate, Success Rank, and Time Away. Highly overloaded users were less successful, and spent more time away from their accounts after a failed event. A 1% increase in Uncertainty was associated with a 1.09% increase in Time Away after a failed event, and a .04% decrease in Success Rate.[10][11]

4.3 Multiple Regression Analysis

Next, we move beyond single regression first through incorporating two control variables, then moving to multi-construct regressions. The control variable Period has a monthly frequency, inclusion of this variable into our initial regressions allows us to observe if users' performance changed over time. When re-running our regressions adding Period, all previously significant relationships

[8] Regressions revealed that weekly periods capitalized on chance and found significant (but small) relationships where none existed on the semester or monthly time scales.

[9] When both the dependent Y and independent X are log-transformed, the coefficient β in the regression model can be interpreted as an elasticity, which represents the percentage change in Y for a one percent change in X

[10] These results are qualitatively unchanged when using the bi-weekly or per semester datasets

[11] Analysis was replicated on datasets including the summer months, anticipating this data would be less reliable due to reduced student activity. Results confirmed this intuition, yielding less significant relationships across the board.

Table 5: Hypotheses, Support Indicators, and Regression Statistics

Hypothesis	Construct	Metric	Supported	Beta
H1a	**NGSE**	**Success Rate (+)**	**No**	**-0.16**
H1b	NGSE	Success Rank (+)	No	−0.24
H1c	NGSE	Time Away (-)	No	0.63
H1d	**NGSE**	**Days Locked Out (-)**	**No**	**0.51**
H2a	SRSE	Success Rate (+)	No	−0.03
H2b	SRSE	Success Rank (+)	No	−0.05
H2c	SRSE	Time Away (-)	No	−0.55
H2d	SRSE	Friction (-)	No	0.03
H3a	**Overload**	**Success Rate (-)**	**Yes**	**-0.05**
H3b	**Overload**	**Success Rank (-)**	**Yes**	**-0.12**
H3c	**Overload**	**Time Away (+)**	**Yes**	**0.88**
H3d	Overload	Days Locked Out (+)	No	−0.06
H3e	Overload	Friction (+)	No	−0.15
H4a	Complexity	Success Rate (-)	No	−0.03
H4b	Complexity	Success Rank (-)	No	−0.07
H4c	Complexity	Time Away (+)	No	0.27
H4d	Complexity	Days Locked Out (+)	No	−0.11
H5a	**Uncertainty**	**Success Rate (-)**	**Yes**	**-0.04**
H5b	Uncertainty	Success Rank (-)	No	−0.12
H5c	Uncertainty	Mean Elapsed (+)	No	−0.52
H5d	**Uncertainty**	**Time Away (+)**	**Yes**	**1.09**

Note: Bold font indicates significance at the 0.05 level

from Table 5 remained, with no qualitative changes to effect sizes or significance.

Our second control variable is PrimaryMFA, which is an important moderator to the 2FA experience. Users may experience different issues depending on the type of second factor used. In our dataset, PrimaryMFA includes three second factor types: SMS, App Notification, and OATH code[12]. These forms of 2FA events do not include instances where no second factor presentation is required due to fulfillment by session token, or similar temporary credential, which don't require interaction by the user. One common MFA feature, where the user can choose to "Remember my Device", enables the user's device to serve as the second factor confirmation. This type of authentication is included when the authentications are interactive through password entry or similar. We re-examine our analysis using multiple regression, including both Period and PrimaryMFA as

[12]Phone Call MFA was also present, but removed due to having only 19 associated observations

Table 6: Regression Results

	Dependent variable (larger or smaller values "better" indicated below):					
	ln(Success Rate) *Larger*	ln(Success Rank) *Larger*	ln(Elapsed Time)	ln(Time Away) *Smaller*	ln(Days Locked Out) *Smaller*	ln(Friction) *Smaller*
ln(Overload)	-0.07^{***}	-0.10	0.01	0.80^{*}	0.03	-0.21
ln(Complexity)	-0.01	-0.08	0.64	-0.47	0.04	0.40
ln(Uncertainty)	0.01	-0.03	-0.71	0.71	-0.07	-0.48
ln(NGSE)	-0.20^{***}	-0.28	0.23	1.50	0.29	-0.85
ln(SRSE)	-0.03	-0.12	-0.09	-0.39	0.01	-0.07
Period	-0.00	-0.06^{***}	-0.14^{***}	0.08^{*}	-0.06^{***}	0.01
App Notification MFA	0.00	0.04	-0.69^{*}	0.49	-0.23^{***}	0.16
OATH Code MFA	-0.14^{***}	-0.25^{*}	-1.64^{**}	0.90^{*}	-0.44^{***}	0.52
Remembered Device MFA	-0.00	-0.02	-5.33^{***}	-1.65^{**}	-1.19^{***}	-1.84^{***}
Constant	0.36^{***}	4.86^{***}	2.62	2.85	2.72^{***}	-4.35^{***}
Observations	1,132	1,132	1,132	434	1,132	1,132
Adjusted R^2	0.05	0.07	0.13	0.07	0.19	0.03

Note: Primary MFA uses Text Message MFA as reference level; *p<0.05; **p<0.01; ***p<0.001

control variables.[13]

Overload, Uncertainty, and Stress are sub constructs of the Security Related Stress (SRS) second-order construct; we expect them to only increase the significance of our observed relationships when included, as they are designed to capture orthogonal variance. Of our two efficacy constructs, only NGSE has any significant relationships using single regression, but controlling for users' reported SRS may help clarify these relationships. All construct superscores are added to the regression with control variables Period and PrimaryMFA. We evaluate these regressions for each response variable, and present the results in Table 6.

Overload was significantly related to Success Rate, with an effect size of -0.07, or 7%. The relationship with Success Rank was insignificant after controlling for other constructs, with the p-value dropping to 0.14. Overload was also significantly related to Time Away. A negative relationship between Overload and success metrics makes sense, as overloaded users could make more frequent mistakes. It also seems reasonable that overloaded users are more likely to stay away following an authentication failure. A doubling in a user's Overload score is expected to result in an 80% increase in the amount of time a user spends away from their device after failing to authenticate.

After controlling for the other constructs, neither Uncertainty nor Complexity have statistically

[13]As a sanity check, we first re-ran all single regressions with just these control variables added. Results were consistent with Table 5.

significant relationships with any of the response variables. Moving on to our two self-efficacy constructs, we see no significant relationships with SRSE. NGSE had a highly significant inverse relationship to Success Rate with an effect size of -0.20. This negative relationship is puzzling and bears further investigation; it suggests some users may over-estimate their ability in the NGSE responses.

Finally, we look at the relationships with our control variables: Period, and MFA Type. Period is negatively associated with Success Rank, Elapsed Time, and Days Locked Out; Period is positively associated with Time Away. These relationships indicate that over time, users' spent fewer days locked out, and less time authenticating; conversely, they spent a bit more time away from their accounts after failed authentication events, and failed more often. The type of second factor used in authentication was also significant in our analysis. Mobile App MFA was significant and negatively associated with Elapsed Time and Days Locked Out compared to the reference method of Text Message second factor. OATH Code MFA was also significant, and inversely related to Success Rate and Rank, Elapsed Time, and Days Locked Out, while positively related to Time Away. This indicates that users who use OATH Code MFA fail more often, resulting in more time away, but spend less time authenticating, and experience lockout more rarely than their Text MFA peers. Use of the "Remember My Device" option, resulting in MFA fulfilled by a "Remembered Device" was associated with significant reductions in Elapsed Time, Time Away, Days Locked Out, and Friction. These results highlight how beneficial adoption of this feature can be to the user experience, and the significant impact that the type of MFA used can have. Finally, we note that for Days Locked Out and Friction, only time and authentication method are significant in explaining variation.

Regressions with Self-Efficacy Categorical Variables Throughout the analysis we observe a negative relationship between NGSE and users' Success Rate, and Success Rank, which is their performance relative to their peers for a given period. We posit this may be due to poorly informed users rating their NGSE too highly; as we close out our analysis, we briefly investigate this result.

The left plot in Figure 3 shows the distribution of the NGSE measure, an average of NGSE item responses.

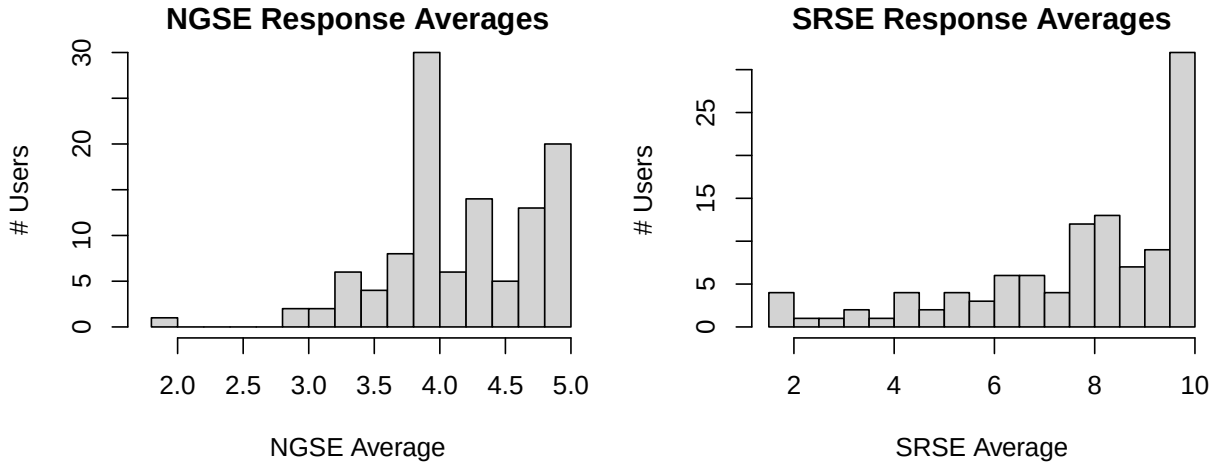

Figure 3: Self-Efficacy Constructs Superscore Distribution

We break NGSE scores into three roughly equal sized categories based on the histogram, using breaks at the values 4.0 and 4.4, yielding 316, 445, and 394 observations for the Low, Medium, and High categories respectively.[14] As SRSE and NGSE are known to correlate and are thematically similar, we repeat this process on SRSE, which has a similarly large rise in the distribution of response averages near the ceiling. We split the SRSE superscores at 7.5 and 9 after consulting the second distribution plotted in Figure 3, yielding relatively equal groups. Replacing the NGSE and SRSE superscores with response categories allows us to control for this potential non-linear correlation with performance metrics. In the regressions, we use the medium score ranges as baseline, so we can look at how low and high-scoring individuals perform relative to those in between.

The regression results controlling for both SRSE and NGSE response levels are shown in Table 7, again using the natural log transforms on each variable. The categorical variables are not natural log transformed; for these relationships, we exponentiate the beta value to get a percentage change in our response variable relative to the reference category, known as an elasticity.

[14]Note that the graph shows user response average frequencies for 111 users, and the number of observations associated with each user depends on presence in the authentication dataset.

Table 7: Multiple Regression with Categorical Efficacy Variables

	Dependent variable: (larger or smaller values "better" indicated below)					
	ln(Success Rate) *Larger*	ln(Success Rank) *Larger*	ln(Elapsed Time)	ln(Time Away) *Smaller*	ln(Days Locked Out) *Smaller*	ln(Friction) *Smaller*
ln(Overload)	-0.06^{***}	-0.08	0.01	0.87^{**}	0.03	-0.25
ln(Complexity)	-0.03	-0.15	0.63	-0.31	0.00	0.44
ln(Uncertainty)	-0.00	-0.04	-0.80	0.81^{*}	-0.10	-0.57
Low NGSE	-0.08^{***}	-0.20^{**}	0.51	-0.03	0.15^{*}	0.63^{*}
High NGSE	0.03	0.00	0.08	-0.58	-0.10	0.84^{**}
Low SRSE	0.03	0.06	-0.68	0.43	-0.31^{***}	-0.32
High SRSE	0.06^{***}	0.17^{*}	-0.63	0.29	-0.27^{***}	-0.26
Period	-0.00	-0.06^{***}	-0.14^{***}	0.09^{*}	-0.06^{***}	0.01
App Notification MFA	0.02	0.08	-0.84^{*}	0.48	-0.26^{***}	-0.02
OATH Code MFA	-0.13^{***}	-0.22^{*}	-1.69^{**}	1.07^{*}	-0.44^{***}	0.39
Remembered Device MFA	0.01	0.00	-5.42^{***}	-1.50^{*}	-1.21^{***}	-1.98^{***}
Constant	0.04	4.28^{***}	3.21^{***}	3.68^{***}	3.42^{***}	-5.73^{***}
Observations	1,132	1,132	1,132	434	1,132	1,132
Adjusted R^2	0.07	0.08	0.13	0.07	0.21	0.04

Note: NGSE, SRSE use Medium as reference level, Primary MFA uses Text Message MFA; *p<0.05; **p<0.01; ***p<0.001

Starting with the Success Rate response variable, we now see Low NGSE is negatively related to both Success Rate and Success Rank, relative to their Medium NGSE peers. This contradicts the counter-intuitive results of earlier regressions, indicating that moderate NGSE responses are assocaited with higher success rates than the lowest NGSE users. However, we note that High NGSE users did not have significantly higher success than their Medium NGSE peers, suggesting some users may be over-confident in their responses. More work is needed to investigate this result.

Moving on to our other response variables, we see newly significant relationships with Days Locked Out and Friction. Low NGSE users experience 16% more days Locked Out from digital resources than their Medium NGSE peers, and while High NGSE users are associated with fewer Days Locked Out, this result failed to reach significance. Taken together, this shows an inverse relationship between NGSE and Days Locked Out. Finally, both Low and Medium NGSE users experienced more friction than their Medium NGSE peers, experiencing an 88% and 232% increase in errors per event respectively. In summary, moderate NGSE responses were associated with the highest Success Rates and Ranks, fewest Days Locked Out, and least errors per authentication event.

SRSE, in which no significant relationship was found in the prior regressions, is significantly related to success and Days Locked Out in the categorical regressions. High SRSE users show

elevated Success Rate and Rank, with a 6% increase in absolute Success Rate over Medium SRSE users and a 19% increase in ranking compared to Medium SRSE peers.

Both High and Low SRSE users are associated with a reduction in Days Locked Out compared to Medium SRSE peers, at 31% and 36% fewer respectively. This result is intriguing, and suggests that Low and High SRSE users may use different coping strategies compared to Medium SRSE users, but more work is needed to investigate this result.

Overload and Complexity had no meaningful changes to their relationships; the effect size and significance associated with the relationship between Overload and Time Away increased slightly. A newly significant relationship emerged between Uncertainty and Time Away after controlling for self-efficacy response levels. A doubling in a user's Uncertainty is associated with an 81% increase in Time Away. This makes intuitive sense, and agrees with our hypothesis H5d.

In summary, we confirmed that biased responses were skewing the observed relationships between efficacy measures and performance metrics, specifically yielding negative relationships between NGSE and performance metrics. Controlling for the level of NGSE and SRSE responses revealed that Low NGSE users outperform their Medium NGSE peers, with no significant difference between Medium and High NGSE success metrics. SRSE became significant in several relationships, and Uncertainty emerged as positively related to Time Away. Forms of second factor were significant for every response variable, and only raw Success Rate and Friction were unaffected by time.

5 Discussion

We now summarize the main findings of the analysis after controlling for both time and type of second factor used, and breaking NGSE and SRSE supserscores into response level categorical variables to control for response biases. Table 8 lists our original twenty-one hypotheses with additional sub-hypotheses to account for NGSE response categories, self-efficacy hypotheses are only broken into their constituent categories where necessary to discuss significant hypotheses.

Significant results are in bold font, and labeled supported when both significant and the effect size is in the predicted direction.

Table 8: Hypotheses, Support Indicators, and Regression Statistics

Hypothesis	Construct	Metric	Supported	Beta
H1a1	**NGSE Low**	**Success Rate (+)**	**No**	**0.08**
H1a2	NGSE High	Success Rate (+)	No	0.03
H1b1	**NGSE Low**	**Success Rank (+)**	**No**	**0.20**
H1b2	NGSE High	Success Rank (+)	No	0.00
H1c	NGSE	Time Away (-)	No	−0.58
H1d1	**NGSE Low**	**Days Locked Out (-)**	**Yes**	**0.15**
H1d2	NGSE High	Days Locked Out (-)	No	**-**0.10
H2a1	SRSE Low	Success Rate (+)	No	0.03
H2a2	**SRSE High**	**Success Rate (+)**	**Yes**	**0.06**
H2b1	SRSE Low	Success Rank (+)	No	0.06
H2b2	**SRSE High**	**Success Rank (+)**	**Yes**	**0.17**
H2c	SRSE	Time Away (-)	No	0.42
H2d	SRSE	Friction (-)	No	−0.32
H3a	**Overload**	**Success Rate (-)**	**Yes**	**-0.06**
H3b	Overload	Success Rank (-)	No	−0.08
H3c	**Overload**	**Time Away (+)**	**Yes**	**0.87**
H3d	Overload	Days Locked Out (+)	No	0.03
H3e	Overload	Friction (+)	No	−0.25
H4a	Complexity	Success Rate (-)	No	−0.03
H4b	Complexity	Success Rank (-)	No	−0.15
H4c	Complexity	Time Away (+)	No	−0.31
H4d	Complexity	Days Locked Out (+)	No	0.00
H5a	Uncertainty	Success Rate (-)	No	0.00
H5b	Uncertainty	Success Rank (-)	No	−0.04
H5c	Uncertainty	Elapsed Time (+)	No	−0.80
H5d	**Uncertainty**	**Time Away (+)**	**Yes**	**0.81**

Note: Bold font indicates significance at the 0.05 level

First, we note that Low NGSE is inversely related to both Success Rate and Rank, showing significantly higher scores relative to Medium NGSE users. Interestingly, High NGSE users do not have significantly different results compared to their Medium NGSE peers. These results both run counter to our hypotheses, and should be investigated in future work with more participants to further clarify these relationships. Next, we see first supported hypothesis H1d1 indicates that Low NGSE users experience more Days Locked Out than their Medium NGSE peers, suggesting

an inverse relationship between NGSE and Days Locked Out as hypothesized.

Moving on to SRSE relationships, we see two newly significant hypotheses after controlling for self-efficacy response levels. High SRSE is associated with improvements in both Success Rate and Success Rank compared to Medium SRSE users, supporting our hypothesized positive relationship. Closing with Security Related Stress constructs, we find two significant hypotheses with Overload and a single significant hypothesis with Uncertainty. A doubling in Overload is associated with a 6% decrease in absolute Success Rate, and an 87% increase in Time Away. Lastly, a 100% increase in Uncertainty was associated with an 81% increase in Time Away following a failed authentication event.

There are two un-hypothesized relationships that emerged as significant in our final regressions: SRSE with Days Locked Out, and NGSE with Friction. Both High and Low SRSE are associated with large reductions in Days Locked Out compared to their Medium SRSE Peers, with a 31% and 36% reduction respectively. Secondly, both Low and High NGSE are associated with increases in Frction, at 88% and 232% respectively, compared to their Medium NGSE peers.

Control variables are also significant in many relationships. Our population had more difficultly over time, as indicated by increased Time Away after failures. Conversely, the frequency of users becoming Locked Out decreased over time, all of which may indicate improved problem solving of the user, improvement in IT help structures, or changes in the associated interfaces. The form of second factor was also significant in many relationships. Using Text Message based MFA as a baseline, App Notification MFA was associated with lower authentication times and fewer days Locked Out. OATH Code MFA was associated with lower rates of success, even relative to peers, and shorter authentication times with fewer instances of lock-out, but more time away from digital resources after a failed attempt. Finally, the use of the "Remember My Device" token option was associated with far shorter authentication times, much less time away from resources after failure, many fewer instances of lock-out, and reduced encounter of errors.

6 Limitations and Future Work

Limitations to Generalizability It is important to remind the reader that this analysis compares longitudinal authentication event data derived from Azure authentication logs starting in November 2021 with survey data collected in late 2020. Users' Security Related Stress, New General Self-Efficacy, and Security Related Self-Efficacy may have changed in that time, and due to the time lag we don't capture any state-like effects. We also note that the Uncertainty and Complexity constructs may capture more state-like situational information than Overload or the efficacy measures, contributing to their lack of significance in the analysis. Next, this study uses a convenience sample of 111 students and faculty associated with several business and technology courses at the authors' university. In this academic context, effect sizes for measures of time cost to the user and organization such as Time Away and Days Locked Out may be inflated compared to a population with more rigid time restrictions on work.

Future Work Collecting the independent and response variables at the same time could help clarify these relationships by capturing state-like effects. Acquiring real-world security control performance data is exceedingly difficult, and the ground truth dataset used in this study was the result of a healthy, mutually beneficial relationship with University IT and IT-SEC. Researchers should consider developing mutually beneficial relationships with organizations' IT and IT-SEC to bring academic insights to industry and develop industry datasets for research.

Future work could extend this research through replication in a corporate population, where the variables capturing time cost are more meaningful. Additionally, the constructs included in our survey measure were chosen prior to the design and collection of the objective authentication event data. Future work should explore additional constructs for inclusion; we recommend considering field dependency vs. independence as a promising addition for predicting performance with digital interfaces often used for security controls(Belk et al., 2017). Lastly, future work should consider recruiting a more diverse or representative group of participants, and specifically investigate the puzzling non-linear relationships between self-efficacy and performance metrics as observed in

our analysis.

7 Conclusion

We have investigated the relationship between psychological measures of individual efficacy and observed security performance using authentication logs collected over 17 months. We responded to the call of existing work to measure actual security control behavior rather than self-reported compliance intention. We hope that our approach can be more widely adopted by others researchers in our field.

This work offers a first test of the ability for the Security Related Self Efficacy, New General Self-Efficacy, and Security-Related Efficacy measures to predict security control performance beyond adoption. We used multi-factor authentication as a representative security control task to investigate the relationships between stress, self-efficacy, and multi-factor authentication performance. Higher levels of security related stress are associated with significant increases in the time spent away from resources following a failed authentication event, and lower success rates. Security Related Self-Efficacy is associated with increased success in authentication, and moderate SRSE is associated with fewer Days Locked Out relative to high or low responses. New General Self-Efficacy is associated with decreased rates of errors for moderate responses, and lower success authenticating with more Days Locked Out for Low NGSE users. Second factor type is significant for every response variable except Friction, which measures the number of errors per authentication event.

References

Aggarwal, A., & Dhurkari, R. K. (2023). Association between stress and information security policy non-compliance behavior: A meta-analysis. *Computers & Security*, *124*, 102991. https://doi.org/10.1016/j.cose.2022.102991

Ament, C., & Haag, S. (2016). How information security requirements stress employees. *International Conference on Interaction Sciences*. https://api.semanticscholar.org/CorpusID:39315788

Belk, M., Fidas, C., Germanakos, P., & Samaras, G. (2017). The interplay between humans, technology and user authentication: A cognitive processing perspective. *Computers in Human Behavior*, *76*, 184–200. https://doi.org/10.1016/j.chb.2017.06.042

Borgert, N., Jansen, L., Böse, I., Friedauer, J., Sasse, M. A., & Elson, M. (2024). Self-efficacy and security behavior: Results from a systematic review of research methods. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–32. https://doi.org/10.1145/3613904.3642432

Chen, G., Gully, S., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods - ORGAN RES METHODS*, *4*. https://doi.org/10.1177/109442810141004

Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, *19*(2), 189–211. https://doi.org/10.2307/249688

Cram, W. A., Proudfoot, J. G., & D'Arcy, J. (2021). When enough is enough: Investigating the antecedents and consequences of information security fatigue. *Information Systems Journal*, *31*(4), 521–549. https://doi.org/10.1111/isj.12319

D'Arcy, J., Herath, T., & Shoss, M. (2014). Understanding employee responses to stressful information security requirements: A coping perspective. *Journal of Management Information Systems*, *31*, 285–318. https://doi.org/10.2753/MIS0742-1222310210

D'Arcy, J., & Teh, P.-L. (2019). Predicting employee information security policy compliance on a daily basis: The interplay of security-related stress, emotions, and neutralization. *Information & Management*, *56*(7), 103151. https://doi.org/10.1016/j.im.2019.02.006

Hastings, S., Bolger, C., Shumway, P., & Moore, T. (2024). Transforming raw authentication logs into interpretable events. *Workshop on SOC Operations and Construction (WOSOC 2024)*. https://dx.doi.org/10.14722/wosoc.2024.23003

Hwang, I., & Cha, O. (2018). Examining technostress creators and role stress as potential threats to employees' information security compliance. *Computers in Human Behavior*, *81*, 282–293. https://doi.org/10.1016/j.chb.2017.12.022

Jeon, S., Son, I., & Han, J. (2023). Understanding employee's emotional reactions to issp compliance: Focus on frustration from security requirements. *Behaviour & Information Technology*, *42*(13), 2093–2110. https://doi.org/10.1080/0144929X.2022.2109512

Kim, S. Y., Park, H., Kim, H., Kim, J., & Seo, K. (2022). Technostress causes cognitive overload in high-stress people: Eye tracking analysis in a virtual kiosk test. *Information Processing & Management*, *59*(6), 103093. https://doi.org/10.1016/j.ipm.2022.103093

Kwak, Y., Lee, S., Damiano, A., & Vishwanath, A. (2020). Why do users not report spear phishing emails? *Telematics and Informatics*, *48*, 101343. https://doi.org/10.1016/j.tele.2020.101343

Lee, C., Lee, C. C., & Kim, S. (2016). Understanding information security stress: Focusing on the type of information security compliance activity. *Computers & Security*, *59*, 60–70. https://doi.org/https://doi.org/10.1016/j.cose.2016.02.004

Maier, C., Laumer, S., Wirth, J., & Weitzel, T. (2019). Technostress and the hierarchical levels of personality: A two-wave study with multiple data samples. *European Journal of Information Systems*, *28*(5), 496–522. https://doi.org/10.1080/0960085X.2019.1614739

Moody, G. D., & Galletta, D. F. (2015). Lost in cyberspace: The impact of information scent and time constraints on stress, performance, and attitudes online. *Journal of Management Information Systems*, *32*(1), 192–224. https://www.jstor.org/stable/26613982

Nasirpouri Shadbad, F., & Biros, D. (2020). Technostress and its influence on employee information security policy compliance. *Information Technology & People*, *35*(1), 119–141. https://doi.org/10.1108/ITP-09-2020-0610

Singh, T., Johnston, A. C., D'Arcy, J., & Harms, P. D. (2023). Stress in the cybersecurity profession: A systematic review of related literature and opportunities for future research. *Or-*

ganizational Cybersecurity Journal: Practice, Process and People, *3*(2), 100–126. https://doi.org/10.1108/OCJ-06-2022-0012

Warkentin, M., & Mutchler, L. (2014, January). Behavioral information security management.

Yuan, Q., Kong, J., Liu, C., & Jiang, Y. (2023). Understanding the effects of specific techno-stressors on strain and job performance: A meta-analysis of the empirical evidence. *Information Technology & People*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/ITP-08-2022-0639