How Security-Related Stress and Self-Efficacy Influence Actual Behavior: An Empirical Study

Seth Hastings[1], Tyler Moore[*1,2], Bradley Brummel[3], and Sal Aurigemma[2]

[1]*Tandy School of Computer Science, College of Engineering and Computer Science, The University of Tulsa, 800 S. Tucker Dr., Tulsa, OK, 74104, USA*
[2]*School of Cyber Studies, College of Engineering and Computer Science, The University of Tulsa, 800 S. Tucker Dr., Tulsa, OK, 74104, USA*
[3]*Department of Psychology, The University of Houston, Heyne Building Rm 126, 3695 Cullen Blvd, Houston, TX 77204, USA*

Abstract

Multi-factor authentication (MFA) is now a cornerstone of organizational cyber defense, yet its real-world effectiveness depends on how individuals experience and perform the required tasks. People naturally vary in technological proficiency and how they react to stress caused by struggles with security tasks. Typically, researchers investigate how such factors impact technology usage by collecting survey responses describing intended behaviors. Unfortunately, there can be a disconnect between what someone says they will do in response to a survey question and what actually transpires in practice. Hence, in this study, we describe an empirically-driven approach that combines survey responses about self-efficacy and stress with observational data describing actual security performance using MFA logs gathered from 109 users in a university setting. The exploratory analysis reveals significant relationships between stress, reduced authentication success, and increased time away following failed attempts. In addition to yielding new insights into the role of psychological constructs and authentication methods in shaping cybersecurity behaviors, the work offers a model in which future researchers could leverage a ubiquitous data source (authentication logs) to study how security performance varies for a range of technologies and user attributes.

[*]Corresponding author: tyler-moore@utulsa.edu.

Keywords: security-related stress; self-efficacy; authentication cost; MFA; user behavior; empirical analysis; behavioral cybersecurity; cybersecurity; security measurement

1 Introduction

Cyber defense has become more complex as responsibility for security is increasingly distributed across technology, individuals, and organizations. Consequently, people have been given increasing responsibility to take actions that safeguard their own cybersecurity and the organizations they work for. From selecting strong passwords to recognizing phishing attacks and utilizing more secure forms of authentication, individuals have become critical to effective cybersecurity practice, which are shaped by threats, policy requirements, and interface constraints.

This interdependence creates tension between usability and security, security controls intended to strengthen defensive posture often impose cognitive and temporal costs that can elevate frustration, fatigue, and breakdowns in task execution, thereby undermining the very protections they are designed to provide (Belk et al., 2017; Cram et al., 2021). Thus, behavioral cybersecurity research emphasizes that understanding the effectiveness of security controls requires studying not only what users intend to do, but how they actually perform when interacting with security mechanisms.

In this study, we adopt a theoretically grounded interpretation of how psychological antecedents shape observed security behaviors in the context of authentication identified self-efficacy as a critical requirement for any behavior change (Bandura, 1977). A recent meta-analysis examined the relationship between self-efficacy and secure behaviors (Borgert et al., 2024), identifying 174 studies conducted between 2010–2021. Research has consistently identified a positive correlation between self-efficacy and secure behavior intentions. A second important attribute is stress; completing security-related tasks can introduce stress in individuals, and when individuals are stressed, they may be less likely to adopt secure behaviors. One more modern and well studied construct that captures this concept is the Security Related Stress (SRS) measure developed by D'Arcy et al., 2014. This and subsequent works explore security related psychological constructs and how these individual differences influence self-reported secure behaviors such as compliance with information security policies (ISPs). For example, the meta-analysis conducted by Aggarwal and Dhurkari, 2023 reported that stress was associated with ISP non-compliance. Taken together, stress and self-

efficacy constructs have offered insight into how human factors might affect security behaviors.

However, a central challenge is that the vast majority of the literature only studies how these factors affect *intended* behavior, not actual behavior. This is because measuring actual security behaviors in a realistic setting is usually hard to do. Nonetheless, researchers have long recognized the value in studying observed behaviors directly. For example, D'Arcy et al., 2014, p. 307 observes that "...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". Hwang and Cha, 2018, p. 290 note that "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". To Warkentin and Mutchler, 2014, January, measuring actual security behavior is the "holy grail" of behavioral cybersecurity research. In other words, it is highly valuable but hard to obtain.

In this paper, we do measure actual security behavior through multi-factor authentication (MFA) activities of participants in an enterprise setting. For those same users, we collect measures of stress and self-efficacy and empirically evaluate how those measures impact these directly observed security behaviors. In doing so, we help close a gap in the existing literature, offering three primary contributions.

C1 We introduce a longitudinal dataset of 21,071 complete MFA events from 109 enterprise users, answering repeated calls for objective behavior data.

C2 We integrate SRS with dual efficacy constructs (NGSE, SRSE) to test direct and interactive effects on authentication *success*, *error*, and *lockout* rates.

C3 We reveal an inverted-U efficacy–performance pattern, moderate NGSE maximizes success, highlighting overconfidence as a neglected factor in secure-auth behavior.

2 Related Work

We first briefly review prior work on self-efficacy and stress as they relate to cybersecurity. Then, we discuss some of the existing research that has investigated actual behaviors. Finally, we contrast the behavioral models in the literature with the model adopted in this work.

There is a robust literature investigating the positive role self-efficacy plays in behavior change and performance under demanding conditions. For example, Wang et al., 2023 find that security self-efficacy is positively correlated with avoiding online threats, while Y. Chen and Zahedi, 2016 finds a similar relationship holds in populations in the United States and China. Furthermore, low security self-efficacy is associated with negative outcomes, including failure to adhere to information security policies (McLeod and Dolezel, 2022) interpreted through the lens of capitulation theory. A meta-analysis has confirmed the robustness of the positive relationship between self-efficacy and security behavioral intentions (Borgert et al., 2024).

In behavioral cybersecurity research, efficacy is conceptualized at multiple levels, making it important to distinguish between constructs conceptualizing self-efficacy in different contexts. New General Self-Efficacy (NGSE) reflects generalized beliefs about one's capability to succeed across novel and challenging situations (G. Chen et al., 2001). For this work, we have adapted Computer Self-Efficacy from Compeau and Higgins, 1995 by re-framing construct items with the security task domain, yielding a conception of Security-Related Self-Efficacy (SRSE) as a trait-like within-domain confidence in one's ability to perform security tasks in a computer environment.

Wang et al., 2023 discuss prior work around security-related self-efficacy in their paper examining the mediating role of security anxiety in internet threat avoidance behavior. They conceptualize it as "Internet users' belief in their ability to take protective measures to avoid internet security threats", and found it predicted avoidance of internet threats and moderates the effect of conceptual and procedural knowledge.

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 conceptualized SRS as a second-order construct, encapsulating dimensions of security-related overload,

uncertainty, and complexity. This construct effectively delineates the relationship between various stressors and the intention to violate information security policies (ISP). Their investigation illuminated key stressors, including security demands, overload, complexity, and uncertainty, which collectively contribute to the phenomenon of 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. 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 identified as mitigating factors.

Hwang and Cha, 2018 focused on the role of technostress creators and role stress in compliance with information security for employees. Their results indicated that technostress negatively impacts compliance by decreasing organizational commitment, with promotion focus moderating the relationship between technostress and role stress. 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 compliance, particularly the role of frustration. They found that frustration negatively impacts compliance, but this effect can be mitigated by providing autonomy to employees. Other work has further refined stress types (Cram et al., 2021; D'Arcy and Teh, 2019; Nasirpouri Shadbad and Biros, 2020) and personality traits (Maier et al., 2019), showing how each impact security compliance intentions. Several meta analyses have confirmed the relationship between

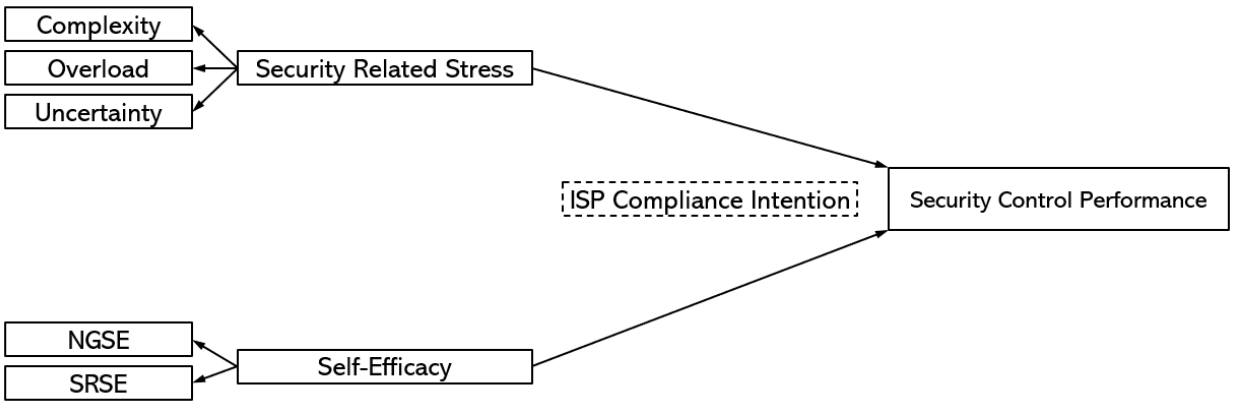

Figure 1: Behavioral model adopted by this study

various stress measures and security behavioral intention (Aggarwal and Dhurkari, 2023; Singh et al., 2023; Yuan et al., 2025).

While rare, a few studies have directly measured security behaviors. For example, Mattson et al., 2023 examines how psychological factors including self-efficacy impact observed password manager adoption and usage. Belanger and Crossler, 2019 demonstrates that users with higher self-efficacy are more likely to adjust mobile-phone settings to become more secure. Kwak et al., 2020 tasked participants with identifying phishing emails in a laboratory setting.

This 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 1. In Section 4 we compare self-reported stress and efficacy measures to observed security control performance without relying on self-reported compliance intentions. This enables a more direct assessment of any links between stress, self-efficacy, and security control performance.

3 Methodology

3.1 Psychological measures

3.1.1 Measures

We select three constructs that theory suggests may impact authentication performance: Security-Related Stress (SRS), New General Self-Efficacy (NGSE), and Security-Related Self-Efficacy (SRSE). SRS captures security-specific Overload, Complexity, and Uncertainty as articulated by Darcy et al D'Arcy et al., 2014; these dimensions represent perceived demands of security requirements that may impede an individual's ability to successfully configure and use authentication mechanisms over time. NGSE is defined as a broad, trait-like belief in one's capability to perform across novel and challenging situations, and is measured using the canonical scale (G. Chen et al., 2001). Computer Self-Efficacy (CSE) refers to confidence in one's ability to use computing technologies to accomplish tasks; we draw on the foundational scale by Compeau and Higgins, 1995 and adapt it to form SRSE, a domain-specific belief in one's capability to configure and use MFA for the studied accounts.

These measures are chosen as constructs well studied in complementary literature around security, while lacking examples of measured relationships to security task outcomes. Using established constructs allows us to test if there exists a measurable signal between these constructs and security outcomes, beyond already observed relationships to security intentions, which has been the primary variable of interest in other work. The efficacy measures complement SRS by capturing perceived capability rather than perceived demand: NGSE represents generalized self-belief, whereas SRSE is task and domain-anchored to MFA use. We treat NGSE as a stable, trait-like individual difference, and we conceptualize SRSE as a security-domain-specific efficacy that is sufficiently stable to predict future MFA configuration and use across study intervals. This conceptualization aligns with our broader assumption that SRS variables exhibit trait-like properties attributable to the individual, which is necessary for predicting authentication outcomes over time.

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, see the appendix for the full survey measure used, and associated item handling.

3.1.2 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 The University of Tulsa, 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

3.2.1 Data source

We collect authentication events data from the author's university between November 8, 2021, and December 31, 2022. Follow the methodology developed in (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.

Events are single row representations of complete interactions. Attributes are fairly intuitive, including the time spent authenticating, 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.

3.2.2 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[1].
- 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 authentica-

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

tion 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 at the start of Section 4.

3.2.3 Demographics

We filter for users who were active across all semesters and removed summer months, leaving 21,071 events from 109 users with an average of 16 events per user each month. Demographics are provided in Table 1, participants were majority male students near the age of 20. While performing group comparisons in detail could be of interest, we believe the small overall participant pool and small group sizes would render those results unreliable. However, we provide the mean differences, standard deviations, Welches t statistic, and Cohen's d between groups across response variables in Tables 10, 11, and 12 in the appendix. Gender had a small impact on Success Rate, with Male participants outperforming their Female counterparts. Role and Age were largely overlapping, with younger and undergraduate participants enjoying higher Success Rates and lower Friction, while conversely experiencing more Days Locked Out.

Table 1: Participant Demographics Summary

Category	**Age**		**Gender**			**Role**			
Sub-Category	**18–22**	**23+**	**Male**	**Female**	**Non-Binary**	**Undergraduate**	**Graduate**	**Staff**	**Other**
Count	92	16	68	38	2	95	11	1	1

3.3 Research Hypotheses

As prior work examines relationships with compliance intention, rather than actual behavior, we approach this exploratory study by postulating exploratory hypotheses on the direction of potential relationships between these well-studied measures and our response variables. This aligns with our goal of exploring a potentially under-utilized signal, authentication logs, which may be a rich data source for future work addressing psychological and behavioral aspects of security performance. The anticipated relationships between psychological constructs and response variables are shown in Table 2. 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 not 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 do not necessarily expect someone with higher SRSE to be more likely to seek help when locked out.[2] 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.

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 will be associated with lower Success Rates due to the increased cognitive

[2] 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.

Table 2: 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: +		

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 do not 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.[3]

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.

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

As this is a first test between these constructs and the observed authentication metrics, the hypotheses function as a vehicle for our analysis. Should the analysis be confirmatory, all we can say is that we have observed that relationships do exist. Further studies and theory building are required to properly explain any such relationships and quantify their strength and method of action.

4 Analysis

4.1 Summary Statistics

As we transition into the analysis phase, we delineate our independent variables in Table 3 and response variables in Table 4. Examining our independent variables, we note a moderate positive correlation between SRSE and NGSE, and moderate inverse associations between SRSE and both Overload and Complexity. NGSE likewise shows inverse correlations with all three SRS constructs. Among the SRS constructs, Overload exhibits the strongest positive association with Complexity and a smaller positive correlation with Uncertainty, while Uncertainty and Complexity are weakly related.

In addition to these descriptive relationships, Table 3 provides metrics reflecting the validity of the constructs within our sample population responses. Internal consistency was acceptable to strong across measures as seen in Cronbach's alpha, with SRSE and NGSE demonstrating particularly high reliability. Convergent validity, assessed via average variance extracted (AVE), was supported for SRSE, NGSE, Overload, and Uncertainty, exceeding the conventional .50 threshold. Complexity, however, fell slightly below this benchmark (AVE = .45), suggesting comparatively weaker convergence, weakening the validity of results around Complexity compared to our other constructs. Discriminant validity was supported by both the Fornell–Larcker criterion and HTMT statistics: the square roots of AVE exceeded the corresponding inter-construct correlations for all constructs, and maximum HTMT values remained below a strict cutoff, indicating that the constructs are empirically distinguishable despite some overlap between Overload and Complexity.

When considering the validity of our predictive variables, we note the uncommon context of this study and our primary contribution focus. This study compares self-reported measures, using previously validated scales, against objective behavioral outcomes. We are not proposing or testing a theoretical model of relationships among latent constructs, but rather examining how these established survey instruments perform as predictors of observable security behaviors. Identifying response variables that are both objective and correlated with established psychological constructs allows researchers to move beyond testing security behavior theory with self-reported outcome or intent. Thus, we do not go into detail in the examination of our predictive variables, as scale development is distal to the focus of the study. [4]

	Mean	SD	AVE	Max HTMT	Correlations				
					1.	2.	3.	4.	5.
1. SRSE	7.79	2.24	.58	.36	**.92**				
2. NGSE	4.18	.58	.60	.36	.34	**.92**			
3. Overload	2.88	1.31	.70	.75	-.31	-.17	**.90**		
4. Uncertainty	3.89	1.13	.65	.31	-.03	-.04	.28	**.87**	
5. Complexity	3.51	1.09	.45	.75	-.30	-.22	.58	.07	**.78**

Table 3: Means, standard deviations, AVE, maximum HTMT, and correlations for independent variables. Diagonal entries (bold) are Cronbach's alpha.

	Mean	SD	Correlations					
			6.	7.	8.	9.	10.	11.
6. Success Rate	.94	.13	1					
7. Success Rank	48.46	26.59	.64	1				
8. Elapsed Time	36.21	76.06	-.12	-.12	1			
9. Days Locked Out	19.73	18.55	-.04	.01	.08	1		
10. Time Away(hrs)	32.00	45.93	-.19	-.26	.00	-.31	1	
11. Friction	.07	.15	-.65	-.40	.15	-.04	.22	1

Table 4: Means, standard deviations, correlations for monthly period response variables[5]

Moving to our response variables in Table 4, we first notice the large correlation we anticipated between Success Rate and Success Rank. Success Rate has a similarly large inverse correlation

[4]Complete construct data available upon request

	1. SRSE	2. NGSE	3. Overload	4. Uncertainty	5. Complexity
6. Success Rate	-.03	**-.12**	**-.13**	-.03	**-.07**
7. Success Rank	-.04	-.01	-.03	-.04	-.03
8. Elapsed Time	.03	.01	-.02	.04	.01
9. Days Locked Out	.03	.04	.02	.00	.06
10. Time Away(hrs)	**-.14**	-.07	.05	-.02	-.03
11. Friction	-.04	.01	**.09**	-.05	**.08**

Table 5: Correlation results between independent and response variables, correlations significant at the 0.05 level in bold font.

with Friction, as failures are 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. The response variables differ between each other in both range and mean value. For example, Success Rate is in the range 0-1 skewed towards 1, while Elapsed Time ranges from 0-n. These response variables represent observed behavior derived from log data, not theoretically constructed measures or estimates. For the purpose of this study, we are not concerned with nor have the analytical power to focus on refining the set of response variables or optimizing treatment; rather, we are performing a first test of the utility of such log data by verifying if measurable associations are present, offering a valuable new data source for future research.

Finally, we examine the correlations between independent and response variables in Table 5.[6] 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 generalized 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

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

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.

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.[7] Natural log transforms were used for both the construct averages and response variables, enabling an intuitive reading of each beta values as an elasticity.[8] Using hypothesis **H3c** in Table 6 as an example: a .80 beta value means a 10% increase in Security Related Overload is associated with in a 8.0% increase in mean Time Away after failure.

Simple regressions supported four of our twenty-one hypotheses. Two additional hypotheses were inversely related at or near significance: NGSE shows a significant negative relationship with Success Rate, and positive relationship with Days Locked Out at a p-value of .07. 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 10% increase in Uncertainty was associated with a 10.7% increase in Time Away after a failed event.[9][10]

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

[8] 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.

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

[10] 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 6: Hypotheses, support indicators, and regression statistics

Hypothesis	Construct	Metric	Supported	Beta	P-value
H1a	NGSE	Success Rate (+)	No	$-.19$	$< .01$
H1b	NGSE	Success Rank (+)	No	$-.34$	.07
H1c	NGSE	Time Away (-)	No	.77	.38
H1d	NGSE	Days Locked Out (-)	No	.36	.09
H2a	SRSE	Success Rate (+)	No	$-.02$	.36
H2b	SRSE	Success Rank (+)	No	$-.07$	.33
H2c	SRSE	Time Away (-)	No	$-.37$	.26
H2d	SRSE	Friction (-)	No	$-.07$	.78
H3a	Overload	Success Rate (-)	Yes	$-.06$	$< .01$
H3b	Overload	Success Rank (-)	Yes	$-.11$	.05
H3c	Overload	Time Away (+)	Yes	.80	$< .01$
H3d	Overload	Days Locked Out (+)	No	$-.07$	.27
H3e	Overload	Friction (+)	No	$-.22$	.30
H4a	Complexity	Success Rate (-)	No	$-.03$	.20
H4b	Complexity	Success Rank (-)	No	$-.10$	.24
H4c	Complexity	Time Away (+)	No	.16	.67
H4d	Complexity	Days Locked Out (+)	No	.06	.53
H5a	Uncertainty	Success Rate (-)	No	$-.03$	.13
H5b	Uncertainty	Success Rank (-)	No	$-.10$	.16
H5c	Uncertainty	Mean Elapsed (+)	No	$-.67$	.10
H5d	Uncertainty	Time Away (+)	Yes	1.07	$< .01$

Labeled as supported for relationships in the hypothesized direction with $p \leq 0.05$

4.3 Multiple Regression Analysis

Next, we move beyond single regression through incorporating three control variables into multiple predictor regressions. The control variable Period represents the month over which a users performance metrics were summarized, inclusion of this variable allows us to remove the influence of temporal or seasonal disturbances in user experience. Our second control variable is PrimaryMFA, which is an important moderator to the 2FA experience as 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.[11] 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 do not require interaction by the user. One common MFA

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

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. Finally, we add NumEvents, the number of events in a given period as a third control variable.[12] Multicollinearity diagnostics were assessed to be acceptable across models; the largest VIF was 1.80, and adjusted GVIF values for factor terms ranged from 1.02 to 1.07.

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 significant relationships using single regression, but controlling for users' reported SRS may help clarify these relationships. All construct superscores are included in the regressions with control variables Period, NumEvents and PrimaryMFA. We evaluate these regressions for each response variable, and present the results in Table 7.

Overload was significantly related to Success Rate with an effect size of -0.07; a much stronger relationship was observed with Time Away, with a 10% increase in Overload correlating with a 7.6% increase in time spent away after a failed authentication. This matches our intuition, as a highly overloaded user would likely have less resources to expend on regaining access, and a lower threshold for breakdowns in task execution. A negative relationship between Overload and Friction runs counter to our intuition, as overloaded users may be expected to make more mistakes, not less, but this result did not reach significance.

After controlling for the other constructs, Complexity remains without any statistically significant relationships to any of the response variables. As we note in summary statistics, Complexity showed relatively weaker convergent validity in our sample, so this result is unsurprising. Uncertainty retains a strong relationship with Time Away; a 10% increase in Uncertainty corresponds to a 7.7% increase in Time Away.

Moving on to our two self-efficacy constructs, we see no significant relationships with SRSE.

[12] Single regressions were recomputed with control variables added; results were consistent with Table 6.

Table 7: 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)	$-.07^{***}$	$-.09$	$-.36$	$.76^{*}$	$-.02$	$-.31$
ln(Complexity)	$-.01$	$-.08$	.86	$-.34$	$-.05$	.32
ln(Uncertainty)	.01	$-.03$	$-.41$	$.77^{*}$	$-.04$	$-.42$
ln(NGSE)	$-.19^{***}$	$-.31$	$-.26$	1.52	.08	-1.06
ln(SRSE)	$-.02$	$-.09$	$-.21$	$-.17$	$-.03$	$-.10$
App Notification	$-.00$	.03	$-.33$	.51	$-.03$	.28
OATH Code	$-.16^{***}$	$-.33^{***}$	-1.24^{*}	1.01^{*}	$-.15^{*}$	$.76^{*}$
Remembered Device	$-.03$	.15	-3.27^{***}	-1.17	$-.33^{***}$	$-.92^{*}$
numEvents	$-.00^{***}$	$-.00^{*}$	$.07^{***}$	$-.02^{*}$	$.04^{***}$	$.03^{***}$
Constant	$.34^{***}$	4.57^{***}	.56	2.23	1.68^{***}	-5.00^{***}
Period F-test	F=2.13*	F=5.03***	F=4.39***	F=1.79	F=27.75***	F=1.60
Observations	1,068	1,068	1,068	420	1,068	1,068
R^2	.09	.07	.25	.13	.76	.08
Adjusted R^2	.07	.05	.23	.09	.76	.06
F-statistic	5.06^{***}	3.94^{***}	17.24^{***}	2.97^{***}	169.08^{***}	4.51^{***}

Note: Period fixed effects included; Primary MFA uses Text Message MFA as reference level. *p<0.05; **p<0.01; ***p<0.001
Multicollinearity diagnostics were acceptable across models (largest VIF = 1.80), see the appendix for model statistics. 13

NGSE displays a highly significant inverse relationship to Success Rate with an effect size of -0.19. This negative relationship is puzzling and bears further investigation. If we assume that those with higher self-efficacy are more competent or capable, this suggests some users may be over confident in NGSE responses.

Finally, we look at the relationships with our control variables: Period, MFA Type, and NumEvents. The control variables are not natural log transformed; for these relationships, we exponentiate the beta value to find the percentage change in our response variable relative to the reference category. Period is an exception due to being a factor variable; here we report the F-test comparing regressions with and without inclusion of the Period variable, finding that across all measures, Period was significant. The type of second factor used in authentication was also significant in our analysis. Mobile App MFA had no significant relationships relative to the reference method Text Message second factor. OATH Code MFA is significant and inversely related to Success Rate, Rank, Elapsed Time, and Days Locked Out. Users of OATH Code MFA had failures more often, but spent less time authenticating, and fewer days locked out of their accounts. Positive relationships with Time Away and Friction complete the picture, showing these users spent more Time Away after failure. The positive correlation with friction reveals that OATH Code

users encountered many more errors compared to their Text MFA peers, which likely contributed to the lower success rate. Use of the "Remember My Device" option, resulting in MFA fulfilled by a "Remembered Device" is associated with dramatic reductions in Elapsed Time, Time Away, Days Locked Out, and Friction, as expected. These results underscore the benefits of adopting this feature for the user experience and emphasize the significant influence of the chosen MFA type.

4.3.1 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. This, along with other unexpected results, suggest that we may have an uncontrolled effect confounding our results. 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 2 shows the distribution of the NGSE measure, an average of NGSE item responses.

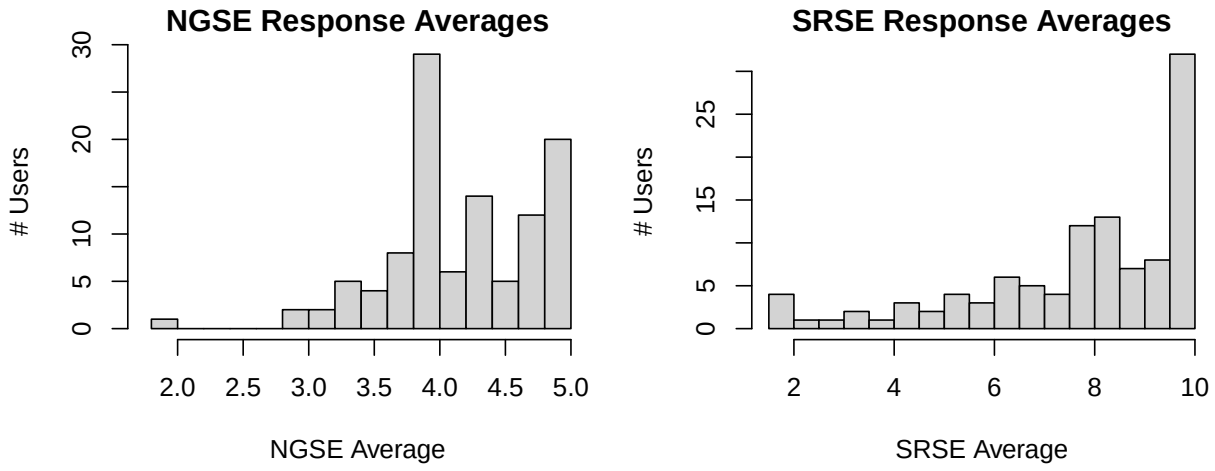

Figure 2: Self-Efficacy constructs superscore distribution

We break NGSE scores into three roughly equal sized categories based on the histogram, using breaks at the values 4.0 and 4.4.[13] As SRSE and NGSE were correlated in our analysis at 0.34 and are thematically similar, we repeat this process on SRSE, which has a similarly large rise in the

[13]Note that the graph shows user response average frequencies for 109 users, and the number of observations associated with each user depends on presence in the authentication dataset.

Table 8: Multiple regression with categorical efficacy variables

	Dependent variable:					
	ln(Success Rate)	ln(Success Rank)	ln(Elapsed Time)	ln(Time Away)	ln(Days Locked-Out)	ln(Friction)
ln(Overload)	-0.06^{**}	-0.06	-0.34	0.80^{*}	-0.02	-0.38
ln(Complexity)	-0.03	-0.19	0.91	-0.16	-0.05	0.41
ln(Uncertainty)	-0.003	-0.05	-0.46	0.84^{*}	-0.04	-0.50
LOW NGSE (n = 287)	-0.08^{***}	-0.23^{***}	0.28	0.02	0.04	0.59^{*}
HIGH NGSE (n = 362)	0.03	-0.02	0.28	-0.78^{*}	0.02	1.02^{***}
LOW SRSE (n = 326)	0.02	-0.01	-0.28	0.20	-0.04	-0.13
HIGH SRSE (n = 389)	0.05^{**}	0.13	-0.32	0.02	-0.06	-0.10
App Notification	0.01	0.07	-0.47	0.52	-0.06	0.11
OATH Code	-0.15^{***}	-0.30^{**}	-1.30^{**}	1.12^{*}	-0.15^{**}	0.63
Remembered Device	-0.02	0.18	-3.38^{***}	-1.02	-0.34^{***}	-1.11^{**}
numEvents	-0.00^{*}	-0.00	0.07^{***}	-0.02^{*}	0.04^{***}	0.03^{***}
Constant	0.05	4.10^{***}	-0.09	3.81^{***}	1.79^{***}	-6.91^{***}
Period F-test	F=1.95*	F=4.74***	F=4.40***	F=1.78	F=27.94***	F=1.54
Observations	1,068	1,068	1,068	420	1,068	1,068
R^2	0.10	0.08	0.25	0.14	0.76	0.09
Adjusted R^2	0.08	0.06	0.23	0.09	0.76	0.07
F Statistic	5.49^{***}	4.19^{***}	15.74^{***}	3.00^{***}	153.90^{***}	4.75^{***}

Primary MFA uses Text Message MFA as reference; NGSE and SRSE use MEDIUM *p<0.05; **p<0.01; ***p<0.001

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 2, yielding balanced 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 8 using the natural log transforms on each variable except for our new categorical variables. Starting with the Success Rate response variable and our new categorical variables, we find Low NGSE is negatively related to both Success Rate and Success Rank, relative to their Medium NGSE peers, having an 8% lower Success Rate. This contradicts the counter-intuitive results of earlier regressions, indicating that moderate NGSE responses are associated with higher success rates than the lowest NGSE users. However, we note that High NGSE users do 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. Low NGSE is also correlated with higher friction, with those in the low category having 80% higher incidence of errors than their medium NGSE peers. Interestingly, high NGSE users, while not having significantly different

Success Rates, also have much higher error incidence, at 177% higher Friction. We hypothesize that this result suggest high NGSE users engage in more complex or diverse authentication behaviors, which result in higher error rates, but balance that behavior with greater remediation skills, resulting in equivalent Success Rates. Similarly, High NGSE users were associated with 118% higher time away after failure. In summary, moderate NGSE responses were associated with the highest performance metrics, while high NGSE responses correlated with equivalent success, but higher error rates and more time spent away after failure. Low NGSE responders had the lowest authentication success, and elevated error rates.

SRSE, for which no significant relationship was found in the prior regressions, is significantly related to Success Rate. High SRSE users show elevated Success Rate with a 5% increase in absolute Success Rate over medium SRSE peers. Both High and Low SRSE users are associated with a reduction in Days Locked Out compared to Medium SRSE peers, at 17% and 36% fewer respectively, with corresponding decreases in Friction, though these results did not reach significance.

Overload retains its relationships from the prior analysis, and Complexity remains insignificant. Uncertainty retains its relationship with Time Away and has a slightly larger effect size. Unlike the Complexity measure, this relationship is accompanied by a large but statistically insignificant decrease in Elapsed Time and Friction. Taken together, these findings suggest that users with higher Uncertainty around the authentication experience may be more risk-averse, encountering fewer errors but demonstrating less resilience, as reflected in shorter authentication times and longer times away. If users achieve similar Success Rates despite spending less time authenticating, encountering fewer errors, and exhibiting more Time Away, it indicates that they may experience a higher cost for each error. These users might abandon the authentication process without attempting to resolve errors and spend more Time Away following a failed attempt.

In summary, controlling for NGSE and SRSE response levels clarifies the relationships between self-efficacy measures and performance metrics. Moderate NGSE users demonstrate the best overall performance, with the highest Success Rates, Success Ranks, and lowest error rates, while high NGSE users achieve equivalent Success Rates despite higher error rates and increased

time spent authenticating. Low NGSE users, however, exhibited the poorest performance metrics, including the lowest Success Rates and elevated Friction. SRSE, previously insignificant, becomes a significant predictor, with high SRSE users achieving higher Success Rates.. Overload and Uncertainty measures are significant, aligning with hypotheses regarding their influence on Success Rate and Time Away, respectively. Importantly, our control variable results reaffirm the significant role of second-factor authentication types across all metrics and underscore the complexity of authentication behaviors.

5 Discussion

Theoretical Implications of Model Analysis We now summarize the main findings of the analysis. The purpose of the study is to examine how self-efficacy, stress, and design choices influence multi-factor authentication (MFA) performance, using authentication logs collected over 13 months.

Prior work on self-efficacy has found a positive association with intention to avoid online threats and to adopt secure behaviors (Wang et al., 2023). Our findings are broadly consistent with this work, but the key difference is our focus on observed behaviors versus intentions. Additionally, the data suggests greater nuance on the relationship between general (NGSE) and security (SRSE) efficacy and user performance. In particular, moderate NGSE responses were associated with the best overall performance, suggesting that balanced self-efficacy fosters effective authentication behaviors. Low NGSE users faced the greatest challenges, including lower success rates and higher error rates. Interestingly, high NGSE users achieve comparable success rates to moderate users but encounter significantly more errors than both moderate and low NGSE peers. This suggests higher reported confidence levels may be associated with riskier or more complex authentication behaviors.

By contrast, we do not see the same dip among users with perceived high security self-efficacy. Instead, the relationship is more linear, consistent with expectations in the literature. High SRSE

users demonstrate increased success rates compared to their medium and low SRSE peers.

Stress-related constructs are also play significant roles, which is again consistent with prior research that has found a positive relationship between stress and poor security decisions such as intention to violate information security policies (ISPs) (D'Arcy et al., 2014). In particular, prior work has posited that technological overload induces stress that motivates ISP violations (Ament and Haag, 2016; Lee et al., 2016). Our findings quite clearly demonstrate evidence of this effect in user behavior following authentication failures. Overload correlates with longer Time Away after failed authentications, reflecting the cognitive burden placed on users. Uncertainty echos this relationship with an even larger effect size, suggesting that risk-averse behaviors may trade efficiency for cautious decision-making.

Control variables highlight key trends, such as variations in performance by MFA type and the high efficiency of the "Remember My Device" token option. Use of this option is associated with reduced error rate, fewer lockouts, and lower time costs, underscoring its value in improving user experience. These results provide insights into how design factors shape user interactions with authentication systems, and inform design considerations for reducing user friction. The findings are consistent with the recommendations made by those focusing on usability aspects of MFA technologies (Abbott and Patil, 2020; Das et al., 2018; Reese et al., 2019; Reynolds et al., 2020).

Practical Implications The model analysis confirms that it is feasible to directly link individual-level measures of self-efficacy and stress with observed security behaviors. This in turn has demonstrated the value of doing so, as it sheds light on how characteristics of the user population can lead to desirable behaviors such as comprehensive utilization of MFA as well as less desirable practices such as remaining logged out of enterprise resources for extended periods following authentication failure.

Scholars should design future experiments that collect actual security behavior, with greater confidence that it can yield valuable insights compared to relying upon behavioral intention alone. On the other hand, research focused on intended actions should also continue in cybersecurity

scholarship because the relationship between intention and action could very well be consistent, as we have found here.

Practitioners should pay closer attention to ways to improve self-efficacy and design their interfaces to minimize stress and the potential to overload users, since they may shut down in response. These findings emphasize the importance of confidence in security-related skills, particularly in high-stakes environments. Hence, targeted efforts to train users on the most crucial security technologies could be beneficial.

6 Limitations and Future Work

Limitations to Generalizability This analysis compares longitudinal authentication event data derived from Azure authentication logs starting in November 2021 with survey data collected in late 2020. Due to the separation in time, the temporal stability of the IVs influences the strength of observed associations, and current measurement would likely increase the strength of those relationships. Additionally, users' Security-Related Stress, New General Self-Efficacy, and Security-Related Self-Efficacy traits may have drifted in that time, and due to the time lag we do not capture any state-like effects. We believe this time lag between measurements contributes to the low confirmation rate of our exploratory hypotheses; further, we 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. Four of our five psychological measures had high validity when analyzed; Complexity had low convergent validity with an average variance extracted of .45, which may contribute to its lack of significant relationships in the analysis. Next, this study uses a sample of 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. With these considerations, we emphasize this study was a first test of the relationship between users reports about their own emo-

tion states and their measured performance later in the future. We find the presence of statistically significant relationships between these measures, despite a two-year gap between measurements, to be an exciting discovery that opens the door for future work. Hence, this study is generalizable in its findings of a significant relationship, but does not represent a thorough assessment of the effects themselves, or the formation of new theory.

Future Work The authors hope this work serves as a useful foundation for more studies on observed security behaviors using authentication logs. Future experiments could provide users with targeted training to see if measured performance improves, or if certain user groups (e.g., older adults, new employees) benefit disproportionately from such interventions. Similarly, capturing state-like affects through repeated measurements could further elucidate the relationship between psychological antecedents and MFA outcomes. Future longitudinal research can improve through a more diverse participant pool, psychological measures collected at multiple points in time, and exploration of fixed and mixed effect models.

When considering 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 may consider the possibility of corporate partnerships, where companies provide researchers with access to authentication logs, and researchers provide companies with data on the performance of their multi-factor authentication, such as time cost, the failure rates of each modality, or many other useful metrics and data that can by generated from a processed user authentication dataset derived from noisy and cumbersome sign-in logs.

7 Conclusion

This paper provides an empirical assessment of how security-related stress and self-efficacy relate to observed security control performance in the context of multi-factor authentication. This research advances our understanding of how psychological measures of self-efficacy and stress

influence security performance, analyzing real-world authentication behaviors rather than self-reported intentions. We construct a longitudinal, event-based dataset of 21,071 authentication interactions derived from enterpise sign-in logs, integrate those behavioral outcomes with survey-based pyschological measures – Security-Related Stress(SRS), New General Self-Efficacy(NGSE), and Security-Related Self-Efficacy(SRSE) – and evaluate their relationship with concrete performance and cost metrics using regression analyses while controlling for the monthly time period and MFA type.

Through these analyses, we identify multiple significant relationships between psychological measures and observed behavior, bridging the gap between intentions and actions, despite the year gap between survey measure collection and the first authentication log used as observed behavior. This both highlights the resilience of the trait-like aspects of the psychological measures across time, and prevents the analysis from capturing any state-like aspects of Overload, Uncertainty, and Complexity, the components of SRS. We find the ability to observe these relationships despite the time lag is an encouraging sign of the utility of the anonymized authentication log dataset as a measure of observed security behavior, offering a potentially more accessible data source for future work.

Our analyses indicate that within SRS, Overload and Uncertainty are meaningfully related to higher time costs following failed authentication, consistent with a resource depletion account of breakdowns in task recovery. Overload was also associated with an increase in the rate of authentication failure. Self-efficacy measures were also important predictors; NGSE exhibits a non-linear relationship with performance: moderate NGSE aligns with the most favorable overall outcomes, while very low NGSE corresponds to poorer success and higher error incidence, and very high NGSE corresponds to elevated errors with reduced post-failure time costs, consistent with an overconfidence-based interpretation. SRSE exhibited a more typical relationship, with the highest reporting users experiencing higher success rates than their low and medium peers.

This study highlights the practical importance of MFA design and configuration choices by showing that both the type of second factor used and use of remembered device tokens are sig-

nificant and have large effect sizes across user outcomes. In particular, "Remember My Device" is associated with substantial reductions in time and error costs, and significant differences were observed between SMS and OATH code modalities, underscoring that observed performance is shaped jointly by individual differences and the usability of the authentication mechanism. Taken together, our specific contribution is demonstrating that authentication logs contain a measurable behavioral signal suitable for evaluating security control effectiveness and user cost, and to provide a methodological foundation for future longitudinal work linking psychological antecedents to objective security outcomes.

Acknowledgments

The authors acknowledge support from Tulsa Innovation Labs via the Cyber Fellows initiative. The authors also thank the IT and IT Security teams at the University of Tulsa for supporting this research by sharing data and enabling our work.

References

Abbott, J., & Patil, S. (2020). How mandatory second factor affects the authentication user experience [https://dl.acm.org/doi/10.1145/3313831.3376457]. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3313831.3376457

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. *Thirty Seventh International Conference on Information Systems*.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215.

Belanger, F., & Crossler, R. E. (2019). Dealing with digital traces: Understanding protective behaviors on mobile devices. *J. Strateg. Inf. Syst.*, *28*(1), 34–49. https://doi.org/10.1016/j.jsis.2018.11.002

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 2024 CHI Conference on Human Factors in Computing Systems*. 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

Chen, Y., & Zahedi, F. M. (2016). Individuals' internet security perceptions and behaviors: Polycontextual contrasts between the united states and china. *MIS Q.*, *40*(1), 205–222. https://doi.org/10.25300/MISQ/2016/40.1.09

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

Das, S., Dingman, A., & Camp, L. J. (2018). Why johnny doesn't use two factor a two-phase usability study of the fido u2f security key. In S. Meiklejohn & K. Sako (Eds.), *Financial cryptography and data security* (pp. 160–179, Vol. 10957). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-58387-6_9

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

Mattson, T., Aurigemma, S., & Ren, J. (2023). Positively fearful: Activating the individual's HERO within to explain volitional security technology adoption. *Journal of the Association for Information Systems*, *24*(3), 664–699.

McLeod, A., & Dolezel, D. (2022). Information security policy non-compliance: Can capitulation theory explain user behaviors? *Comput. Secur.*, *112*(100). https://doi.org/10.1016/j.cose.2021.102526

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

Reese, K., Smith, T., Dutson, J., Armknecht, J., Cameron, J., & Seamons, K. (2019). A usability study of five two-factor authentication methods. *Proceedings of the Fifteenth USENIX Conference on Usable Privacy and Security*, 357–370.

Reynolds, J., Samarin, N., Barnes, J. D., Judd, T., Mason, J., Bailey, M., & Egelman, S. (2020). Empirical measurement of systemic 2FA usability. *USENIX Security Symposium*.

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

Wang, X., Li, Y., Khasraghi, H. J., & Trumbach, C. (2023). The mediating role of security anxiety in internet threat avoidance behavior. *Comput. Secur.*, *134*(100). https://doi.org/10.1016/j.cose.2023.103429

Warkentin, M., & Mutchler, L. (2014, January). Behavioral information security management. In *Computing handbook 3rd edition* (pp. 54.1–54.20). Taylor & Francis Group.

Yuan, Q., Kong, J., Liu, C., & Jiang, Y. (2025). Understanding the effects of specific techno-stressors on strain and job performance: A meta-analysis of the empirical evidence [https://doi.org/10.1108/ITP-08-2022-0639]. *Information Technology & People*, *38*(2). https://doi.org/10.1108/ITP-08-2022-0639

A Appendix

Table 9 below lists the survey items used to collect our construct scores. Construct scores were calculated as item means, and repeating the final regression using item sums yielded no qualitative changes. Two questions outside of the construct items were reverse coded as attention checks. All construct items were mandatory response questions; no construct items were dropped from the survey.

Group comparisons are provided for primary response variables in the tables following. Gender had a small difference in Success Rate where male participants outperformed their female counterparts. Age and Role are almost entirely overlapping, and show statistically meaningful effects across Success Rate, Friction, and Days Locked Out, with younger and undergraduate participants experiencing both higher Success Rates, lower Friction, and interestingly, more days locked out, though this effect was smaller and less significant.

Table 9: Survey Items by Construct

Item	Description
0-10	SRSE: "Regarding the use of 2FA for my TU accounts, I could configure and use 2FA . . . "
Q454	. . . if there was no one around to tell me what steps to follow.
Q456	. . . if I had never experienced using 2FA like it before.
Q458	. . . if I had access to written instructions for reference.
Q460	. . . if I had seen someone else using it before trying it myself.
Q462	. . . if I could contact someone for help if I got stuck.
Q464	. . . if someone else had helped me get started.
Q466	. . . if I had a lot of time to complete the job to set up and use TU 2FA.
Q468	. . . if I had just the TU help desk and website to use for assistance.
Q470	. . . if someone showed me how to do it first.
Q472	. . . if I had experienced a similar use of 2FA before.
1-5	NGSE:
NGSE1	I will be able to achieve most of the goals that I have set for myself.
NGSE2	I will be able to achieve most of the goals that I have set for myself.
NGSE3	In general, I think that I can obtain outcomes that are important to me.
NGSE4	I believe I can succeed at most any endeavor to which I set my mind.
NGSE5	I will be able to successfully overcome many challenges.
NGSE6	I am confident that I can perform effectively on many different tasks.
NGSE7	Compared to other people, I can do most tasks very well.
NGSE8	Even when things are tough, I can perform quite well.
1-7	Complexity:
SRSCX1	I sometimes feel pressure in school due to information security requirements.
SRSCX2	I find that other students often know more about information security than I do.
SRSCX3	I do not know enough about information security to comply with my TU's policies.
SRSCX4	I often find it difficult to understand my TU's information security policies.
SRSCX5	It takes me awhile to understand my TU's information security policies and procedures.
SRSCX6	I sometimes do not have time to comply with my TU's information security policies.
1-7	Overload:
SRSOL1	I am forced by information security policies and procedures to do more work than I can handle.
SRSOL2	TU's information security policies and procedures hinder my very tight time schedules.
SRSOL3	I have a higher workload due to increased information security requirements.
SRSOL4	I am forced to change my work habits to adapt to my TU's information security requirements.
1-7	Uncertainty:
SRSUC1	There are constant changes in information security policies and procedures at TU.
SRSUC2	There are frequent upgrades to information security procedures at TU.
SRSUC3	There are always new information security requirements at TU.
SRSUC4	There are constant changes in security-related technologies at TU.

Table 10: Outcome by Gender (Male vs Female). EG = Male mean, CG = Female mean, MD(SD) = mean difference (pooled standard deviation), t = Welch t statistic, d = Cohen's d.

Outcome	EG(n=68)	CG(n=38)	MD(SD)	t	d
SuccessRate	0.95	0.93	0.02(0.13)	2.61	0.19
MeanElapsed	34.44	39.32	-4.89(76.06)	-0.97	-0.06
LockoutDays	19.43	20.27	-0.84(18.55)	-0.67	-0.05
MeanTimeAway	2831.70	4866.90	-2035.20(12217.87)	-1.34	-0.17
Friction	0.07	0.08	-0.01(0.15)	-1.22	-0.08
SuccessRank	48.33	48.70	-0.38(26.61)	-0.22	-0.01

Two non-binary participants not included.

Table 11: Outcome by Age (18-22 vs 23+). EG = 18-22 mean, CG = 23+ mean, MD(SD) = mean difference (pooled standard deviation), t = Welch t statistic, d = Cohen's d.

Outcome	EG(n=92)	CG(n=16)	MD(SD)	t	d
SuccessRate	0.95	0.90	0.04(0.13)	2.82	0.35
MeanElapsed	36.40	35.14	1.26(76.10)	0.31	0.02
LockoutDays	20.17	17.26	2.91(18.53)	2.18	0.16
MeanTimeAway	3328.24	4618.01	-1289.77(12246.90)	-0.66	-0.11
Friction	0.06	0.13	-0.07(0.15)	-3.38	-0.46
SuccessRank	49.26	43.95	5.31(26.54)	2.25	0.20

Table 12: Outcome by Role (Undergrad vs Other). EG = Undergrad mean, CG = Other mean, MD(SD) = mean difference (pooled standard deviation), t = Welch t statistic, d = Cohen's d.

Outcome	EG(n=95)	CG(n=13)	MD(SD)	t	d
SuccessRate	0.95	0.89	0.06(0.13)	3.04	0.47
MeanElapsed	36.63	33.07	3.56(76.09)	0.82	0.05
LockoutDays	20.11	16.86	3.25(18.53)	2.34	0.18
MeanTimeAway	3311.34	5199.16	-1887.82(12240.27)	-0.68	-0.15
Friction	0.06	0.16	-0.10(0.15)	-3.98	-0.66
SuccessRank	48.93	44.91	4.02(26.58)	1.44	0.15

A.1 Multicollinearity Diagnostics for Table 7 Models

Table 13: Multicollinearity diagnostics for Table 7 regression models

Predictor	Models 1, 2, 3, 5, 6	Model 4 (Time Away)
VIF: $\ln(\text{Overload})$	1.80	1.72
VIF: $\ln(\text{Complexity})$	1.66	1.64
VIF: $\ln(\text{Uncertainty})$	1.13	1.10
VIF: $\ln(\text{NGSE})$	1.18	1.22
VIF: $\ln(\text{SRSE})$	1.26	1.25
Adj. GVIF: Period fixed effects	1.02	1.02
Adj. GVIF: Primary MFA	1.07	1.06
VIF: numEvents	1.36	1.36
Largest 1-df VIF	1.80	1.72

Note: Entries for $\ln(\text{Overload})$, $\ln(\text{Complexity})$, $\ln(\text{Uncertainty})$, $\ln(\text{NGSE})$, $\ln(\text{SRSE})$, and numEvents are ordinary variance inflation factors (VIFs). Entries for Period fixed effects and Primary MFA are adjusted generalized variance inflation factors. All diagnostics were well below commonly cited concern thresholds.

To evaluate multicollinearity in the multiple regression models reported in Table 7, we computed variance inflation factors (VIFs) for all one-degree-of-freedom predictors and adjusted generalized variance inflation factors for the factor terms. The resulting diagnostics were uniformly low across model specifications. Continuous-predictor VIFs ranged from 1.10 to 1.80, and adjusted GVIF values for the factor terms ranged from 1.02 to 1.07, indicating no evidence of problematic multicollinearity.