Research Spotlight

Does Peoples' Keyboard Typing Reflect Their Stress Level?

An Exploratory Study

Paul Freihaut and Anja S. Göritz

Department of Occupational and Consumer Psychology, University of Freiburg, Germany

Abstract: Keyboard-typing tracking offers a convenient behavioral data collection method in web-based study settings. This paper investigated the feasibility of utilizing keyboard-typing for stress measurement. We present data from two experiments: a laboratory study with $N = 53$ participants and an online study with $N = 924$ participants. In both studies, participants typed standardized text sequences during a high-stress or low-stress condition. The manipulation checks revealed consistent differences in participant's stress levels according to experimental conditions. The analysis of 11 typing features with frequentist and machine learning methods revealed a few isolated links between stress and keyboard typing, but the results were inconsistent across both studies and the analysis methods. To foster replication, critical discussion, and new developments, we follow the open science principles of open data, source, and methodology.

Keywords: keyboard typing, tracking, behavior, stress, measurement

In web-based studies, researchers are usually limited to use subjective self-report when assessing stress levels or other affective states. Collecting physiological data such as heart rate is difficult because their measurement requires specific instruments and the expertise to use them. In the present paper, we examine keyboard-typing tracking as an alternative technique to measure stress. Tracking participant's use of their computer input device offers an objective, convenient, and unobtrusive behavioral data collection method, thus potentially providing a useful addition to existing stress measurement approaches (for an overview of stress measurement approaches, see Alberdi et al., 2016).

The rationale for an effect of stress on keyboard typing is that affective states manifest in a behavioral response, for example, a characteristic facial expression (Zimmermann et al., 2003). Indeed, evidence suggests that the stress reaction involves changes in psychomotor functions likely required for typing behavior such as muscular activity and motor control (Van Gemmert & Van Galen, 1997) or attention and working memory (Eysenck & Derakshan, 2011), which seem to play an important role in the planning and execution of sensorimotor actions (Gallivan et al., 2018).

Empirical evidence on the topic, however, is sparse. Vizer et al. (2009) investigated the potential to predict different stress states from keyboard typing behavior. The authors let 24 participants write fictitious emails in a physical stress, cognitive stress, and a neutral condition and captured 42 keystroke (e.g., the number of backspace presses)

and linguistic (e.g., the number of unique words) typing parameters. Using machine learning classification, they predicted the condition in which the text was written at 75% accuracy for the cognitive stress versus neutral classification and 62.5% for the physical stress versus neutral classification. Lee et al. (2015) studied the effect of valence and arousal on a standardized typing task. The authors let 41 participants listen to an emotional sound before typing "748596132" in several trials. Their analysis revealed an effect of arousal on keystroke dwell-time (i.e., the time between pressing and releasing a key), and keystroke latency (i.e., the time between releasing a key and pressing the next key), but not on typing accuracy (i.e., the percentage of correctly typed trials). They did not find an effect of valence on any of the three typing features and no Valence by Arousal interaction.

In sum, there is some theoretical reasoning for the effect of stress on keyboard typing, and the first studies support the idea. However, theory and empirics provide only tentative evidence, and the research area lacks a solid foundation. In line with Vizer et al. (2009) and Lee et al. (2015), we, therefore, refrained from testing hypotheses about specific effects of stress on keyboard typing behavior and followed an exploratory approach. Across two experimental studies, we searched for meaningful evidence in favor of a relationship between typing a standardized text and the stress level during the typing to foster a better understanding of the underlying processes and the potential use of

keyboard typing data for stress measurement. Experiment 1 was a web-delivered laboratory (lab) study to maximize internal validity. Experiment 2 was an online study with a focus on external validity. In accordance with open science, our data as well as the code of our experiments and data analyses are available at https://doi.org/10.5281/zenodo.4445197.

Method

We report both experiments in parallel because they follow the same logic. The studies were programmed as single-page web apps using React.js and Firebase as a backend. Demos of the experiments can be viewed at https://freihaut.github.io/Experiments-Live-Demo/.

Participants

Fifty-three participants (M_{age} = 21.80, SD_{age} = 3.40; 40 female, 13 male, 92.5% students) took part in the lab study. Participants in the online study were recruited via WiSoPanel (Göritz, 2009). One thousand ninety-one participants completed the study. We excluded data from 167 participants (15.31%) because they showed signs of careless responding or technical difficulties, which left N_{final} = 924 participants (M_{age} = 53.69, SD_{age} = 13.03; 488 females, 436 males).

Design

The lab study had a within-subject design. Participants typed in standardized text in a high-stress and low-stress condition. The online study had a between-subject design and additionally included a baseline measurement of typing behavior. Participants typed in standardized text in a neutral baseline condition and subsequently in a high-stress or low-stress condition.

Stress Manipulation

The stress manipulation in both studies consisted of a threatening versus neutral framing of the condition to induce social-evaluative stress combined with a clearly versus mildly challenging stress manipulation task to induce mental stress. In the lab study, the high-stress condition was framed as a monitored performance test, which requires additional testing with an experimenter if performance is too weak. The low-stress condition was framed as an exercise without performance monitoring or evaluation. The stress manipulation task in the lab study was a mental arithmetic task (Figure 1; Pruessner et al., 1999), in which participants had to solve five trials of additions and subtractions within a time limit of 7 s each. A score indicated the total performance. In the high-stress condition, the trials were difficult to solve within the time limit, had a ticking sound to increase time pressure, and a sound of failure upon wrong answers. In the low-stress condition, the trials were easier and without sounds.

In the online study, the high-stress condition was framed as an intelligence test. The low-stress condition was framed as an exercise that teaches skills for working on computerized tasks. The stress manipulation task in the online study was a self-developed counting task (Figure 1). The rationale for using a self-developed task was that the stress manipulation task in a web-based study should require little practice and be as standardized as possible to maximize internal validity. In seven trials of 5 s each, participants saw a varying number of three geometrical shapes (i.e., squares and two differently rotated hexagons) and needed to count the squares. A loading bar visualized the remaining time during each trial. At the end of the task, participants had 10 s to type the total number of counted squares during all trials into an input field. In the high-stress condition, participants needed to count more squares (287 vs. 115) and saw more distracting hexagons (798 vs. 319).

Typing Task

The typing task in both studies required participants to type standardized text sequences similar to a password into an input field in several trials (Figure 2). A trial ended as soon the text sequence was correctly typed in. Participants received feedback upon mistakes and had to correct them.

In the lab study, the typing task consisted of seven trials of 6-digit-number sequences (e.g., 257187). There was a time limit of 60 s to complete all trials. All participants finished within the time limit. The typing task was identical in the high-stress and low-stress conditions.

In the online study, the typing task consisted of eight trials of 6-digit-letter-number sequences (e.g., Tz3j98). There was no time limit. We excluded 88 participants (factored in N_{final} = 924) who paused the task for longer than 10 s or had an outlier task time because they likely did not work on the task as intended. The baseline typing task was identical to the typing task in the high-stress and low-stress conditions.

Measures

Keyboard typing behavior was captured via a JavaScript script, which was embedded in the experiment web apps of both studies. The script logged each keystroke event, that is, a key-down event when a key was pushed and a key-up

Figure 1. Screenshots of the (translated) stress manipulation: Mental arithmetic task (left) and counting task (right).
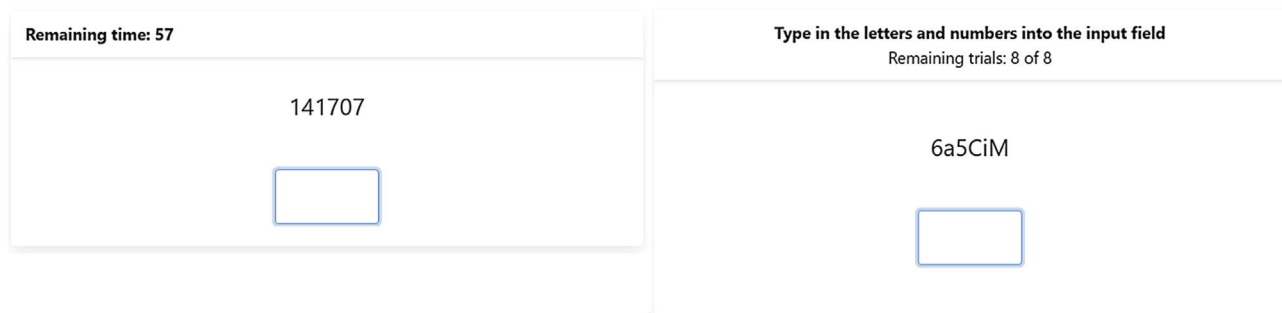

Figure 2. Screenshots of the (translated) typing task in the laboratory study (left) and online study (right).

event when a key was released, with the corresponding keycode (e.g., backspace key) and a timestamp. From the data, we calculated 11 typing features in both studies representing typing accuracy (e.g., number of backspace presses) and typing speed (e.g., average time between pressing and releasing a key). For an overview of all features, see the supplementary material at http://dx.doi.org/10.23668/psycharchives.5024.

Stress was assessed with multiple measures. In both studies, participants reported their emotional state's valence and arousal on the Self-Assessment-Manikin (SAM; Bradley, & Lang, 1994), their mood, rest, and alertness on the German Multidimensional Mood Questionnaire (MDBF; Steyer et al., 1997) as well as their stress and nostalgia. The purpose of the nostalgia question was to assess the specificity of the stress manipulation, as the stress manipulation was expected to affect stress level but to *not* affect nostalgia (Van Tilburg et al., 2019). In the lab study, we additionally captured participant's heart rate (BPM) and electrodermal activity (EDA) during the typing task.

Procedure

The lab study started with the introduction and the practice of all study tasks. The study tasks included the keyboard

typing task, the mental arithmetic stress manipulation task, and four computer mouse usage tasks, which are not part of the present paper and are discussed elsewhere (Freihaut & Göritz, 2021). Next, participants were randomly assigned to start with the high-stress or low-stress condition. Each condition had an introduction page, which included the stress manipulation framing. Then, participants worked on the stress manipulation task and, without a pause, completed the typing task (or one of the mouse usage tasks) until they finished all tasks. After the typing task, participants rated their emotional state's valence and arousal on the SAM. At the end of the condition, participants filled in the MDBF and rated their stress and nostalgia level. Next, participants completed the missing condition. The lab study ended with a debriefing. An experimenter was present during the entire experiment. Participants had to type with their right hand because the electrodes to capture EDA were attached to their left hand. Four participants reported being left-handed. We included them in further data analysis as they were familiar with typing with both hands.

The online study started with the introduction and the practice of all study tasks and also included four mouse usage tasks in addition to the keyboard typing task (for a discussion of the mouse usage tasks, see Freihaut et al., 2021). The practice of the typing task (and all mouse usage tasks) was followed by the baseline measurement of the

typing task as well the baseline measurement of valence and arousal on the SAM. After completing all practice and baseline tasks, participants filled in the MDBF and rated their baseline stress and nostalgia level. Next, participants were randomly assigned to the high-stress (n = 457) or low-stress condition (n = 467). The condition started with an introduction page, which included the stress manipulation framing. Next, participants worked on the counting stress manipulation task and, without a pause, completed the typing task (or one of the mouse usage tasks) until they finished all tasks. After the typing task, participants rated their valence and arousal on the SAM. At the end of the condition, participants filled in the MDBF and rated their stress and nostalgia level. The online study ended with a debriefing.

Results

In the following, we summarized the most relevant results.

Manipulation Check

In the lab study, paired t-tests revealed significant differences between the high-stress and low-stress conditions on all stress measures (all $p < .05$, $0.17 \leq d \leq 0.78$) except for the alertness versus tiredness MDBF subscale, $p = .15$. As expected, there was no significant difference in nostalgia, $p = .41$.

In the online study, we conducted mixed analysis of variance (ANOVAs) with conditions (high-stress vs. low-stress) as between-subjects and experimental phase (baseline vs. condition) as the within-subjects factors. The Condition × Phase interaction, was significant for all MDBF subscales, the stress rating, and valence (all $p < .05$, $.005 \leq \eta^2_p \leq .018$). The stress-related condition-baseline change scores were greater in the high-stress condition, and post hoc comparisons of the stress level during the condition phase revealed significantly higher stress levels for participants in the high-stress condition as compared to the low-stress condition (all $p < .05$, $0.14 \leq$ Hedges' $g \leq 0.18$). The interaction effect for arousal was not significant, $F(1, 922) = 3.06$, $p = .080$, $\eta^2_p = .003$. There was no significant interaction effect for nostalgia, $p = .17$.

Exploring the Effect of Stress on Keyboard Typing

We used three data analysis procedures:

(1) For each keyboard typing feature, we tested individually if it differed between the high-stress and low-stress conditions. To account for multiple testing, we controlled for a false discovery rate of 5% and adjusted the p-values accordingly (Benjamini & Hochberg, 1995).

In the lab study, paired t-tests revealed a significant difference between the conditions for the mean typing latency, M_{HS} = 589.09 ms, M_{LS} = 569.72 ms, $t(52) = 2.97$, $p = .0495$, $d = 0.18$. There were no significant differences for any of the 10 other typing features.

In the online study, mixed ANOVAs revealed a significant Condition × Phase interaction for the standard deviation of the dwell times, $F(1, 922) = 8.18$, $p = .0473$, $\eta^2_p = .088$. The condition-baseline change was greater in the low-stress condition (Δlow-stress = −29.90 ms) than in the high-stress condition (Δhigh-stress = −7.04 ms). A post hoc test revealed no significant difference during the condition phase, M_{HS} = 257.07 ms, M_{LS} = 254.00 ms, $p = .08$, Hedges' $g = 0.024$. There were no significant interaction effects for any of the 10 other typing features.

(2) We used machine learning classification to test globally if the stress condition (high-stress vs. low-stress) can be predicted from the keyboard typing behavior (similar to Vizer et al., 2009). The machine-learning algorithm was a 3-nearest neighbor's classifier. Prediction performance (i.e., the percentage of correct predictions) was assessed with 5-times repeated 5-fold cross-validation (Kuhn & Johnson, 2013). A permutation test assessed the significance of the prediction performance (Figure 3; Ojala & Garriga, 2010). We trained two machine learning models: the first used the original keyboard feature values as the model input and the second the keyboard feature's difference scores between the high-stress and low-stress conditions (lab study) and condition and baseline (online study) as the input. The purpose of the latter approach was to account for individual differences in typing behavior. In both studies, no model predicted the condition significantly better than chance.

(3) Our experimental design assumed the existence of two groups with a dichotomous stress level high versus low. However, stress is continuous, and stress manipulation affects participants differently. To account for this, we collapsed the groups and tested if participants' typing behavior can predict their stress level during the typing task regardless of which experimental group they belonged to. Specifically, we used machine learning regression to test if participants' BPM and EDA (lab study) as well as their valence and arousal ratings (both studies) can be predicted from their keyboard typing behavior. The machine-learning algorithm was a 3-nearest neighbor's regressor. The prediction performance (i.e., coefficient of determination) was assessed with 5-times repeated 5-fold cross-validation. The model predicts better than chance if $R^2 > 0$. Again, we trained one model with the original keyboard features and another model with the difference score keyboard
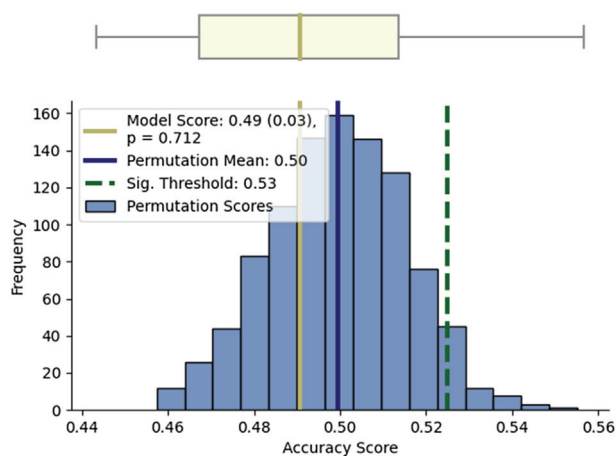
Figure 3. Results of a permutation test for condition prediction. The bars represent the distribution of the 1,000 prediction scores on permutated condition labels. The model with the real condition labels predicts significantly better than chance if the prediction score is better than 95% of the prediction scores on permuted labels (53% accuracy). The boxplot visualizes the distribution of the predictions of the five times repeated 5-fold cross-validation.

features as the model input. In both studies, no model predicted any dependent variable at $R^2 > 0$.

Discussion

In sum, our analyses did not reveal a clear link between stress and keyboard typing. Although isolated keyboard features differed between the high-stress and low-stress conditions, there are no consistent patterns in the data across both experiments. Therefore, the results question the hypothesis that stress has a characteristic effect on typing a standardized text, and hence the validity of a standardized typing task for stress measurement. Compared to Vizer et al. (2009) and Lee et al. (2015), we interpret our results more cautiously. Both other studies found mixed rather than clear evidence for an effect of affective states on keyboard typing, and both had small sample sizes. More generally, the absence of empirical evidence on this topic might signify publication bias that indirectly supports our interpretation. For example, Vizer et al. (2009) additionally let participants write standardized texts but only published the results about the free text writing.

As to limitations, the by-and-large null results of both studies do not warrant the conclusion that stress does not affect typing behavior since failure to find an effect in any given study cannot prove the non-existence of the effect in question. As regards the machine learning analysis, other models (e.g., with a different machine learning algorithm) might also fit the data. However, the goal of the analysis was to get a (conservative) estimate of the effect

of stress on typing behavior rather than to find the best fitting model for our datasets. If there was a substantive effect in the present datasets, it is likely that our approach would have found it. The manipulation check indicated differing stress levels between the conditions, but especially in the online study, the effects were small, and we had no physiological data to back up the self-report. For this reason, our experiment involuntary highlighted the difficulty of studying stress in a web-based setting and, conversely, the need for valid stress manipulation protocols and measurements. In this regard, we hope that the present paper provides a starting point for replication, critical discussion, and new developments.

References

Alberdi, A., Aztiria, A., & Basarab, A. (2016). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. *Journal of Biomedical Informatics, 59*, 49–75. https://doi.org/10.1016/j.jbi.2015.11.007

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *Series B, 57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Eysenck, M., & Derakshan, N. (2011). New perspectives in attentional control theory. *Personality and Individual Differences, 50*(7), 955–960. https://doi.org/10.1016/j.paid.2010.08.019

Freihaut, P., & Göritz, A. S. (2021). Using the computer mouse for stress measurement – An empirical investigation and critical review. *International Journal of Human-Computer Studies, 145*, Article 102520. https://doi.org/10.1016/j.ijhcs.2020.102520

Freihaut, P., Göritz, A. S., Rockstroh, C., & Blum, J. (2021). Tracking stress via the computer mouse? Promises and challenges of a potential behavioral stress marker. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-021-01568-8

Gallivan, J. P., Chapman, C. S., Wolpert, D. M., & Flanagan, J. R. (2018). Decision-making in sensorimotor control. *Nature Reviews Neuroscience, 19*(9), 519–534. https://doi.org/10.1038/s41583-018-0045-9

Göritz, A. S. (2009). Building and managing an online panel with phpPanelAdmin. *Behavior Research Methods, 41*, 1177–1182. https://doi.org/10.3758/BRM.41.4.1177

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer. https://doi.org/10.1007/978–1–4614–6849-3

Lee, P. M., Tsui, W. H., & Hsiao, T. C. (2015). The influence of emotion on keyboard typing: An experimental study using auditory stimuli. *PLoS One, 10*(6), Article e0129056. https://doi.org/10.1371/journal.pone.0129056

Ojala, M., & Garriga, G. C. (2010). Permutation tests for studying classifier performance. *Journal of Machine Learning Research, 11*, 1833–1863. https://doi.org/10.1109/ICDM.2009.108

Pruessner, J. C., Hellhammer, D. H., & Kirschbaum, C. (1999). Low self-esteem, induced failure and the adrenocortical stress response. *Personality and Individual Differences, 27*(3), 477–489. https://doi.org/10.1016/S0191-8869(98)00256-6

Steyer, R., Schwenkmezger, P., Notz, P., & Eid, M. (1997). *Der Mehrdimensionale Befindlichkeitsfragebogen (MDBF). Handanweisung* [The Multidimensional Mood Questionnaire]. Hogrefe.

Van Gemmert, A. W., & Van Galen, G. P. (1997). Stress, neuromotor noise, and human performance: A theoretical perspective. *Journal of Experimental Psychology: Human Perception and Performance, 23*(5), 1299–1313. https://doi.org/10.1037/0096-1523.23.5.1299

Van Tilburg, W. A. P., Bruder, M., Wildschut, T., Sedikides, C., & Göritz, A. S. (2019). An appraisal profile of nostalgia. *Emotion, 19*(1), 21–36. https://doi.org/10.1037/emo0000417

Vizer, L. M., Zhou, L., & Sears, A. (2009). Automated stress detection using keystroke and linguistic features: An exploratory study. *International Journal of Human-Computer Studies, 67*(10), 870–886. https://doi.org/10.1016/j.ijhcs.2009.07.005

Zimmermann, P., Guttormsen, S., Danuser, B., & Gomez, P. (2003). Affective computing – A rationale for measuring mood with mouse and keyboard. *International Journal of Occupational Safety and Ergonomics, 9*(4), 539–551. https://doi.org/10.1080/10803548.2003.11076589

Paul Freihaut

Department of Occupational and Consumer Psychology
Albert-Ludwigs-Universität Freiburg
Engelbergerstr. 41
79106 Freiburg
Germany
paul.freihaut@psychologie.uni-freiburg.de