Published: 29 November 2025

Citation in BibTeX format

OZCHI '25: 37th Australian Conference on Human-Computer Interaction
November 29 - December 3, 2025
Gadigal | Sydney, Australia

SHORT-PAPER

Understanding Users' Perceptions and Barriers to Mental Workload Self-Tracking

YUWEI CHEN, Delft University of Technology, Delft, Zuid-Holland, Netherlands

TEODORA MITREVSKA, Ludwig-Maximilians-University Munich, Munich, Bayern, Germany

TILMAN DINGLER, Delft University of Technology, Delft, Zuid-Holland, Netherlands

CHRISTINA SCHNEEGASS, Faculty of Industrial Design Engineering, TU Delft, Delft, Zuid-Holland, Netherlands

OZCHI '25: Proceedings of the 37th Australian Conference on Human-Computer Interaction (November 2025)
https://doi.org/10.1145/3764687.3769928
ISBN: 9798400720161

Understanding Users' Perceptions and Barriers to Mental Workload Self-Tracking

Yuwei Chen
Industrial Design Engineering
Delft University of Technology
Delft, Netherlands
joanne103084@gmail.com

Teodora Mitrevska
LMU Munich
Munich, Germany
teodora.mitrevska@ifi.lmu.de

Tilman Dingler
Industrial Design Engineering
Delft University of Technology
Delft, Netherlands
t.dingler@tudelft.nl

Christina Schneegass
IDE
TU Delft
Delft, Netherlands
c.schneegass@tudelft.nl

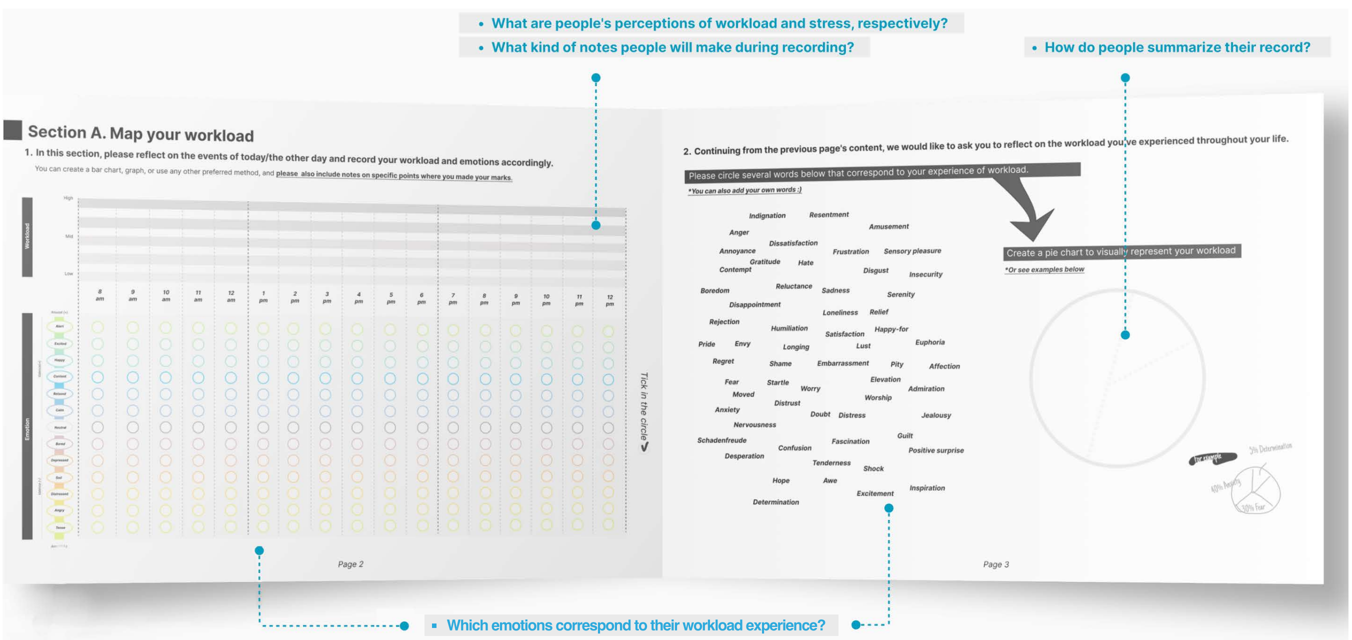

Figure 1: An excerpt of our experience sampling booklet used to understand users' workload during the day and how they relate their workload experience to emotions.

Abstract

Novel consumer neurotechnologies allow users to track their cognitive states and processes, such as attention and mental workload (MWL). However, data on these inherently complex, abstract, and invisible cognitive processes can be challenging to interpret, and little is known about how users make sense of their data. In this work, we explore how people understand and reflect on MWL through six semi-structured interviews and a follow-up experience sampling study. We examine how people conceptualize MWL, distinguish it from related concepts such as stress, what they consider high and low workload in their daily lives, and how they connect workload to emotional states. We discuss these user perceptions and identify barriers to MWL self-tracking, such as lack of trust in the data and ambiguity of the MWL concept, and propose five design guidelines to make cognitive tracking tools more intelligible and meaningful for users.

OZCHI '25, Sydney, Australia

ACM ISBN 979-8-4007-2016-1/25/11
https://doi.org/10.1145/3764687.3769928

CCS Concepts

• **Human-centered computing** → *Participatory design*; *Systems and tools for interaction design*; *Usability testing*.

Keywords

Mental Workload, Self-Tracking, User-Centered Design, Neurotechnology, Experience Sampling

ACM Reference Format:

Yuwei Chen, Teodora Mitrevska, Tilman Dingler, and Christina Schneegass. 2025. Understanding Users' Perceptions and Barriers to Mental Workload Self-Tracking. In *37th Australian Conference on Human-Computer Interaction (HCI) (OZCHI '25), November 29–December 03, 2025, Sydney, Australia.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3764687.3769928

1 Introduction

Since wearable cognitive tracking technologies have become available on the consumer market in recent years, users are able to capture more and more data on their (mental) health and well-being (e.g., with the Muse headband[1], Focus Calm[2], or Neurosity[3]). Tools that track personal data with the goal of monitoring and reflecting on them are summarized under the term Personal Informatics (PI). Since cognitive tracking is about collecting cognitive data, prior work has referred to it as Cognitive Personal Informatics (CPI) (cf. [27, 33]).

In contrast to a fitness tracker that counts steps, cognitive tracking aims to quantify intrinsically invisible and abstract cognitive processes and present them to users in a way that enables them to derive meaningful insights. This includes the monitoring of metrics such as users' focus time, stress, or mental workload (MWL). However, these concepts are complex, connected, and not always self-explanatory or understandable to non-expert users. Especially the intangible concept of MWL and its distinction from stress has already been discussed as a challenge in prior research [1, 2], with Alsuraykh et al. [1] highlighting that in HCI contexts, often the same physiological signal is used to estimate both MWL and stress levels. Yet, from a theoretical perspective, the concepts are distinct and need to be distinguished to properly interpret the data. To inform how CPI technologies could present such data in a way intelligible to layman users, we first need to explore people's perception and interpretation of the two constructs in more detail.

In this work, we conducted six semi-structured interviews to gain insights into people's definition of MWL, reported experiences of low and high MWL, and interpretations of visualizations of MWL and stress. Further, we performed an experience sampling study ($N = 10$) where participants had to sample their MWL, stress, and associated emotions for one day and reflect on their experience. Our results show that indeed (1) there is an overlap in peoples' understanding of MWL and particularly the concept of stress, (2) users doubt the accuracy of MWL data, (3) users prefer aggregated insights over granular data, and that (4) tracking MWL is more interesting in work-related contexts over personal well-being. We discuss our insights from a user-centered perspective and derive five guidelines for designing consumer CPI technologies that create meaningful tracking experiences.

[1]Muse Headband - https://choosemuse.com/, last accessed August 8, 2025
[2]FocusCalm - https://focuscalm.com/, last accessed August 8, 2025
[3]Neurosity - https://neurosity.co/, last accessed August 8, 2025

2 Related Work

2.1 Mental Workload vs. Cognitive Workload

Before we begin to explain the theoretical background of the workload concept, we first want to briefly distinguish between terms often used interchangeably but are not indeed synonyms - mental (work)load and cognitive (work)load. We want to acknowledge that the boundaries between mental and cognitive workload are not always clearly defined and applied in the literature, but for our work, we make the following distinction.

We predominantly use the term **mental workload**, which is often viewed from a broader psychological view and considered more subjective compared to **cognitive workload**, which is usually associated with specific information processing demands and often measured objectively [4]. The fact that the two concepts are measured differently, subjective vs. objective, already hints at one of the core research challenges we see for cognitive personal information systems – matching users' subjective perception of and experience with workload to objectively collected data from sensors.

2.2 Theoretical Foundation of Mental Workload

The concept of workload is founded in Baddeley's model of the working memory [5], describing our memory as a complex system, storing and processing information for the short term and being the foundation for the whole human thought processes. This model, but also many related models, assumes that working memory capacity is a limited resource. MWL describes the relationship between the working memory and attentional resources required by a task, the resources allocated to it, and the effect it has on the task performance [29]. Further, Wickens [30, 31] suggested that humans do not only have a single channel for processing information but multiple. Those can be used in parallel if the information does not conflict. For example, simultaneous processing is possible for visual and auditory information, focal and ambient presentation, spatially and verbally encoded information, or if the information requires manual spatial vs vocal verbal responses [30, 31]. This being a simplified description, it is important to note that perceived MWL is further affected by other factors, such as the number of tasks at hand, time of day, fatigue, or individual differences [22, 29].

2.3 Cognitive Personal Informatics and Mental Workload Tracking

Cognitive processes have long been evaluated only from an ergonomics perspective in safety-critical domains, such as air traffic control [8, 17] and shift work [16]. In HCI, established retrospective and subjective measures like the NASA-TLX or SWAT questionnaires are used more often than continuous sensing technologies [18]. With the advancements in wearable tracking technologies, cognitive tracking is on the verge of moving into our everyday lives [28, 32, 33]. Research has already explored how continuous sensing technologies such as EEG and fNIRS could estimate cognitive processes in the wild [11]. Here, it has been shown that consumer-grade devices or prototypes with a reduced electrode setup [24] can infer cognitive states with high accuracy.

In their recent work, Wilson et al. [34] discuss how mental workload could be positioned as personal data and integrated into our current view on personal informatics, phrasing it *cognitive personal informatics* [34]. Their work emphasizes the complexity of presenting such data and metrics to the user. Furthermore, in their later work, they explored users' lived experiences with workload, contexts in which they faced high or low load, and discussed users' interpretation of the concept itself [23].

Our work builds on this foundation but aims to dig deeper into the specific understanding of MWL and the distinction from related terms, especially the negatively connoted term of stress.

2.4 Mental Workload vs. Stress

From a theoretical perspective, workload and stress are distinct concepts, yet they can show similar physiological responses and can be measured with overlapping physiological signals (e.g., heart rate variability, fNIRS, EEG) [1, 2]. The model of Lazarus and Folkman [19] describes stress as a perceived imbalance between demands and resources. While mental workload is often perceived from the perspective of resources available for a task, stress can be seen as a task-unrelated and much broader emotional construct (cf. [14]). However, they are related [12], as stress does have a direct impact on workload, occupying working memory resources that are required for problem-solving and reasoning skills [3, 25]. In return, stress can lead to simple tasks requiring a higher workload, and a momentary high workload can also be perceived as stressful. Thus, these two concepts can be difficult to distinguish for the layperson, meaning that CPI systems need to be very clear about communicating their differences.

Research Gap: From a theoretical perspective, we already know that MWL is a complex and multifaceted construct that can be situational and subjective. We further learned from prior research that mental and emotional states often coincide to create one experience. What we don't know yet is how non-experts understand this complex concept. Thus, this work takes a user-centered approach and aims to look closer into how people perceive this concept of MWL and specifically how they distinguish it from stress.

3 Method

To investigate users' understanding of cognitive workload and its connection to stress, we conducted a two-part study. We started with six semi-structured interviews to gain an in-depth understanding of people's perception of MWL and a follow-up experience sampling study with ten participants to collect experiences of different workload levels in people's everyday lives.

3.1 Interviews

3.1.1 Sample. We recruited six participants, three students and three non-students, from the backgrounds of financial risk management, law, and engineering. Four identified as female, two as male. Two participants indicated prior experience with neurotechnologies; the other four did not. We will refer to them using the acronym A1-A6.

3.1.2 Procedure. We conducted the interviews in person and virtually via videoconferencing, each lasting around one hour, a time that participants volunteered and did not receive compensation for. For our study, we followed the ethical guidelines of our university

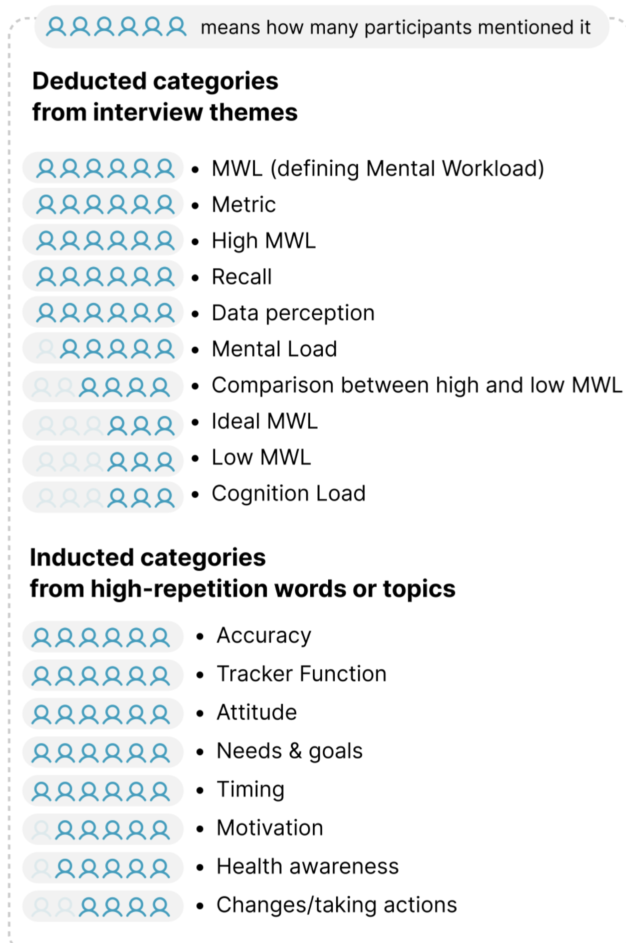

Figure 2: Overview of the 18 categories we generated from our interviews using deductive and inductive coding, the blue icons representing how many of our six participants mentioned the theme.

on human-subject research. After a short introduction to the interview topic and receiving the informed consent, our interview was structured into five parts: (1) discussing with participants their prior personal tracking experience (with neurotechnologies but also other health data trackers), (2) personal definition of MWL, (3) interpretation of visualizations of MWL data, (4) measuring MWL, and (5) reflecting on MWL. Finally, we debriefed our interviewees and thanked them for their participation.

3.1.3 Analysis. We performed an inductive thematic analysis [7, 9] using an iterative coding approach, resulting in 123 codes, categorized into 18 initial categories, ten deducted from the interview themes and an additional eight inducted from high-repetition words and topics (see Figure 2).

3.1.4 Results - Deducted Categories. We started by asking participants to **define MWL and distinguish it from related concepts**. Five participants associated MWL with stress or used stress as a way to explain high MWL, while three described it as a psychological limitation. When provided with the terms of **mental load** vs. **cognitive load**, participants found mental load more relatable, using words such as emotions or stress to explain it, such as A2 saying "*I am thinking.. [MWL] might be a kind of emotional accumulation. Like you may have negative emotions that keep accumulating over time. Otherwise, it could be that your mind is consistently in a stressed state.*" In contrast, people struggled more to explain the term cognitive load (3 out of 6 participants).

When asking for **understanding and examples of high, low, and optimal MWL**, we saw that each three participants were unable to report examples for low workload, articulate their ideal workload, or describe what could be considered low workload to begin with. For example, A2 mentioned that "*Low-level MWL might involve doing simple things.. It is a bit tricky to come up with an example.*" Each of the two participants described low MWL as being less focused or doing less thinking. In contrast, high MWL was again associated with being stressed (5/6 participants) or being fully attentive (3/6 participants). When it comes to telling the difference between high and low MWL, 3 participants referred to activities, such as reading or socializing as high MWL tasks, while doing laundry would be a low MWL task. Other participants differentiated between tasks requiring more or less thinking (2/6 participants) or classified MWL by whether they achieved their goal (2/6 participants). Furthermore, participants described their **ideal MWL** to be a task that is a bit challenging (2/6 participants), or that would create values and be of a duration that is adequate (2/6 participants).

In general, when asking participants to recall their mental workload during the day, their **data perception** predominantly focuses on peak values, such as high MWL (5 out of 6 participants), busiest times of the day (6/6 participants), accomplishing the most significant goal of the day (2/6 participants), emotionally intense moments (2/6 participants), or occurrences that do not fit they usual routine (4/6 participants).

3.1.5 Results - Inducted Categories. Regarding their **motivation to track cognitive data**, the majority of participants expressed interest in verifying their body's reaction (4/6 participants) and a need to understand the causes of negative cognitive states and stress (4/6 participants). Additionally, they mentioned interest in tracking for self-exploration (3/6 participants), efficiency and focus (3/6 participants), curiosity (2/6 participants), or health decline (1/6 participants).

All participants further stated that reflecting on their MWL after a task or experience would be their optimal **timing**. Four further noted that it would be distracting to do it during a task, with two even saying that it could further increase stress.

We further noticed that participants showed various **attitudes** towards CPI and MWL tracking in general. While some seemed rather indifferent (3/6 participants), others were rather reluctant to change their current practice (2/6) or did not expect to receive any new information from the data (4/6 participants).

Our participants also brought up the topic of **mental health awareness** and saw MWL as a relevant factor for it (4/6 participants), but none stated any current problems in that domain.

When discussing the data generated by cognitive tracking devices and their **accuracy**, many expressed doubts about the technology being able to distinguish properly among states (5/6 participants) and were not sure if the data would represent the situation well (4/6 participants).

Additionally, even when receiving data, participants reported uncertainty about what they would **change or what actions to take**. For example, A1 mentioned that "also, if it claims to measure stress or something like that, I personally feel that I don't have the capability to adjust in the moment. Even if you know you're under a lot of stress, you can't just stop doing what you need to do.."

What participants did consider helpful or mentioned as their **need**, was having a tool to check if one enters a defined target state (4/6 participants), suggesting a target state in the first place (2/6 participants), stabilizing stress (3/6 participants), helping with interpreting the data (2/6 participants), understanding the impact of it (4/6 participants) and providing suggestions on how to deal with the results (4/6 participants).

Lastly, linking back to our initial inquiry about intelligibility of the **metrics** of MWL, cognitive load, stress, etc., we saw that all participants would like to receive insights into their stress level (6/6 participants), focus (3/6 participants) and concentration (3/6 participants), engagement (2/6 participants), and exhaustion or prolonged stress (1/6 participants). However, multiple participants still expressed that they perceive ambiguity among the concepts (5/6 participants), and are uncertain about the measurement of metrics (2/6 participants).

3.1.6 Summary. We saw that participants struggled to articulate MWL and differentiate it from related concepts such as stress. Also, they reported trouble in describing situations of low MWL and used stress again as a synonym for high MWL. Overall, when being asked to report on their workload, participants focused strongly on extreme values. To gain deeper insights into the fluctuations of MWL during the day and which tasks and experiences create high and low workload, the following section will report on a follow-up study utilizing a one-day experience sampling approach. By instructing participants to note down their perceived MWL, stress, and associated emotions over one day, we aim to increase their awareness of the subtle differences among the concepts.

3.2 Experience Sampling Booklet

In the second part of our research, we wanted to find out more about how people experience and track mental workload on an average day and how they relate it to stress and emotions. Thus, we created a paper-based experience sampling booklet.

3.2.1 Sample. We recruited ten participants with no specific requirements through snowball sampling on campus. All ten were students of different programs and, in general, were interested in mental workload tracking. We will report on them using the acronym B1-B10.

3.2.2 Procedure. We conducted our survey offline with a printed booklet to sample participants' experience. Participants were asked to choose a normal day and record their mental workload, emotions, and stress between 8 am and midnight (details about the booklet in the following section). We asked them to enter data as often as they could for optimized recall, but did not specify a time interval. After one full day of sampling, participants returned the booklet.

3.2.3 Probe: Booklet. The experience sampling booklet consists of three main sections: (a) Daily record of workload and emotions, (b) daily record of stress, and (c) reflection on the recording of sections a and b. In the first two sections, participants are encouraged to express their thoughts and experiences about their perceived workload, emotions, and stress. Part C invites them to reflect on their data and provide suggestions in writing. In part (a), participants need to enter their perceived workload from 8 am to midnight, ranging from low to high (see Figure 1, top left). Data can be entered continuously (i.e., as a curve or line), or discretely (i.e., as [connected] points), focusing either on trends of specific events or occurrences that stand out to participants. The booklet shows one-hour intervals, but it is also possible to make multiple marks per hour. Participants are asked to annotate the data to provide additional insights. Below the workload recording (see Figure 1, bottom left), we use the same one-hour intervals to ask participants to enter their corresponding emotions. Here, we use the circumplex model of emotions [26], which includes 13 emotional states representing arousal (high/low) and valence (high, low) – three per combination plus a neutral state. On the second page (Figure 1, right side), we collect more detailed insights into the emotional experience independent of the specific events and times. For this purpose, we provide an additional list of 60 emotion words, 20 positive and 40 negative, based on the emotion typology by Fokkinga and Desmet [10] and give participants the opportunity to add their own words if needed. Lastly, we ask them to indicate in a pie chart how representative each emotion is for their perception of workload (e.g., frustration 20% In part (b) of the booklet, participants enter their stress in the same way they enter the workload, on a timeline from 8 am to 12 pm. To better understand stressful experiences, we ask them to illustrate three situations of any stress level and elaborate briefly on them. Lastly, they receive the same typology of 60 emotional states and are asked to create the pie chart. In the last section (c), participants reflect on their experience, especially indicating the perceived level of overlap of workload and stress levels (5-point Likert scale from no overlap to complete overlap, also represented as percent from 0% to 100% overlap), their subjective impressions of commonalities and differences among workload, emotions, and stress, and any additional thoughts.

3.2.4 Analysis. Due to the different types of data we collected, we performed both qualitative and quantitative analyses. We will descriptively report on people's reported MWL, stress, and emotions, and their perceived overlap, compute an exploratory linear regression between reported MWL and stress, and use a thematic analysis approach to analyze open comments. For the regression, we transformed the input to numerical values, the workload and stress reports into a scale of 1-10 (1 being very low workload and 10 being very high workload), and the emotions into a scale of 1-13 (1 being very low arousal, 10 very high arousal).

3.2.5 Results.

Overlap Workload and Stress. An exploratory multiple regression analysis was conducted to assess whether emotion and stress could predict perceived workload. The model indicates that stress significantly predicts workload ($\beta = .569$, $t = 8.30$, $p < .001$), while emotion does not ($\beta = -0.026$, $t = -0.44$, $p = .658$). Overall, stress explains 34.4% of the variance in workload ($R^2 = .344$). The perceived overlap was also subjectively reported by the participants and on average rated as $M = 3.5$ ($SD = 1.08$, 5 being a complete overlap). Figure 3 shows the MWL and stress curves based on participants' ratings ranked according to their perceived overlap rating.

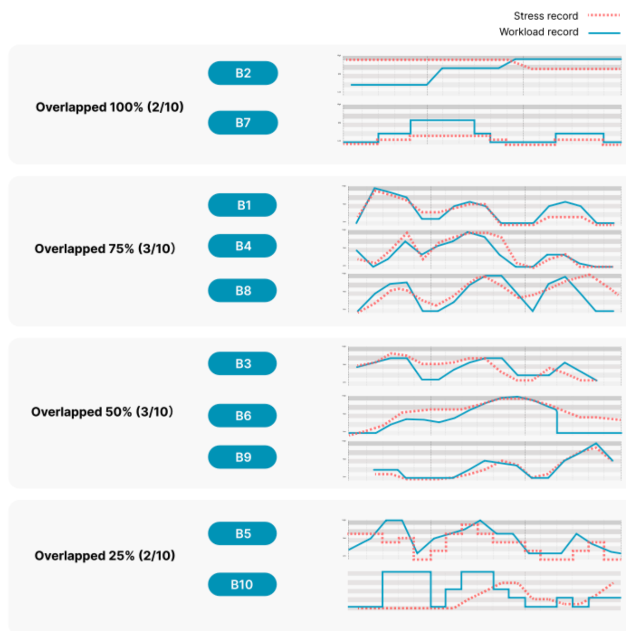

Figure 3: Visual Comparison of stress (red dotted line) and mental workload (blue line) record overlap among participants (B1-B10). Curves are grouped by subjective overlap rating from 25% to 100% perceived overlap.

Emotional Typology. Participants used positive words to describe mental workload, with a rate of 38% vs. 62% negative words. For stress, all words were negative. For both workload and stress, the most frequently used term by far was **Anxiety** (MWL 7/10 people, stress 8/10 people). It is also the emotion making up the largest proportion of the pie chart, with an overall average of 26% in MWL and 77% in stress. When describing workload, the terms frustration, dissatisfaction, worry, and self-hatred were also mentioned more than once. Figures 4 and 5 show the frequencies of the negative words used to describe workload and stress, respectively. Some participants also mention positive emotions when describing workload, such as satisfaction (5/10), but excitement is a term making up the highest percentage of the concept of MWL (11% in the pie chart).

Stress Experiences. After recording their stress level, we asked people to report on two or more memorable stress experiences, either through writing or drawing, that indicate the source or key factor contributing to their stress. We saw that the majority of

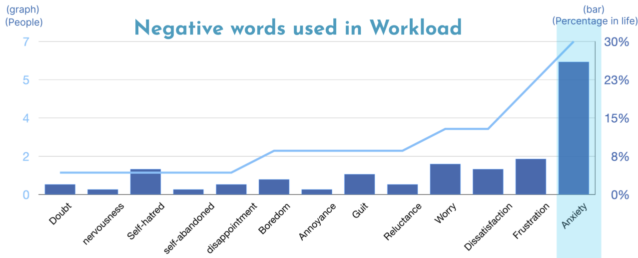

Figure 4: Frequency of negative words reported to explain mental workload.

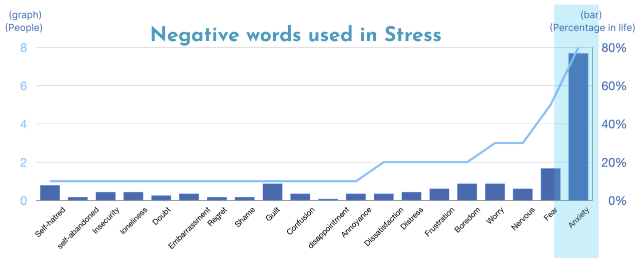

Figure 5: Frequency of negative words reported to explain stress.

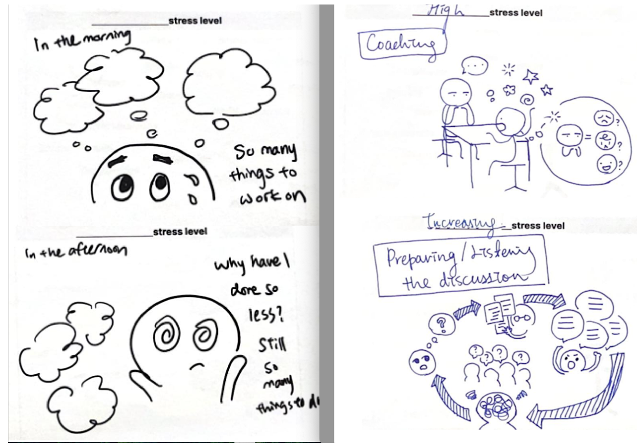

Figure 6: Two different participants' drawings of their stress experiences, on the left, the person distinguished by morning and afternoon (top vs. bottom), on the right, the person drew an instance of high and one instance of increasing stress.

participants still focused on reporting high-stress experiences (8/10). However, the way they described it varied – some differentiated between morning and afternoon (see Figure 6, left), while others focused on the magnitude (see Figure 6, right). We further saw that there was rarely a distinction between medium and low stress levels. In high-stress experiences, we noticed that the majority of participants described situations relating to planning (e.g., "time-management", "to-do's", "plan", or "efficiency") and communication (e.g., "group discussion", "communication", or "people").

4 Discussion

4.1 Reporting MWL

We noticed in our booklet that some people reported their MWL experience as a continuous line, while others drew segments of the same MWL level, while again others used marks at peak and valley points of different frequency (for example, B5 reported 15 events during the day, while B8 only plotted four). This raises the question of how granular people wish to report MWL and how they conceptualize it – as a dynamic fluctuation vs. a state that consists over a certain amount of time.

4.2 Ambiguity of Mental Workload

In both our interviews and experience sampling, we noticed that the concept of mental workload was not tangible for our participants. While the definition in theory is clear, described as the relationship between cognitive resources available and resources required by a task, they struggled to define it in their own words. Often, (negative) emotional terms were used to describe (high) MWL – most often the feeling of experiencing anxiety, followed by feeling frustrated, worried, or dissatisfied. This shows that cognitive processes and emotions are not distinct in everyday life and should not be treated as such. We see the need for a clear **mental model** of the concept MWL. In contrast to existing personal informatics metrics such as heart rate, step count, or sleep quality, MWL appears to be harder to observe, evaluate, and interpret for non-experts.

4.3 Differentiating MWL and Stress

A central challenge in interpreting MWL data is its common conflation with stress. The interviews in our study confirmed prior findings: users often describe MWL in terms of emotional tension or burnout rather than cognitive effort. This confusion undermines the ability of neurotechnology to provide differentiated insights.

4.4 Lack of trust in MWL data accuracy

In both the interviews and experience sampling, participants expressed concerns about the accuracy of wearable cognitive tracking technologies. Unlike, for example, a step counter, which can be evaluated by the user, reflecting on one's own workload during a cognitively demanding task is very challenging.

4.5 Limitations

We acknowledge that our work has several limitations. Firstly, our interview sample was small ($N = 6$) and homogeneous (mainly students). While we saw that many participants shared similar experiences, we can not assume full saturation as their perspectives, daily lives, and technology experiences are similar and do not represent a diverse population. Similarly, the results from our experience sampling study with 10 participants, specifically our quantitative regression approach, should be seen as exploratory and a stepping stone for future research, rather than generalizable conclusions.

5 Design Guidelines

To support users in interpreting MWL data, we propose the following design guidelines as starting points.

(1) **Use metaphors for making MWL understandable**
In our research, we have already tried to elicit colloquial phrases to describe situations of high and low MWL. Future work has to build on that and find terms understandable for the general public that describe the complex concepts of human cognition. This could be phrases such as "mental buffer" so indicate that MWL is a limited-capacity system. Prior work from related personal informatics domains has shown that designing metrics that support reflection on data is crucial to allow for meaningful interpretation [6, 35].

(2) **Provide comparative visualizations for stress and MWL**
Since MWL and stress appear to be difficult to distinguish for non-expert users, providing visualizations that make data on these two concepts comparable could be beneficial. For example, plotting both in one graph or displaying two graphs side-by-side can allow users to observe how fluctuations of one curve show in the other, such as a spike in MWL without the corresponding stress could indicate an enjoyable, flow-like state.

(3) **Present data in different levels of abstraction**
Offering different levels of data analysis, for example, providing a minute-to-minute breakdown of MWL complementary to a daily summary, can help users reflect better on their data. While some users want to just quickly reflect on their day, others might be curious to learn more about their MWL during specific times or activities.

(4) **Support self-reflection and contextual integration**
We could further enrich the data by asking users to provide annotations of their cognitive data on contextual factors, such as what activity they were doing, how they felt, or who they were with. Using questions such as "why do you think your MWL was so high at this point?", could help users with timely reflection on their data and create more meaningful data (cf. [20, 21]).

(5) **Allow users to provide feedback**
Related to (4), asking users for specific feedback on what high and low MWL mean to them would allow a personal configuration of the scale. We envision a human-in-the-loop approach where users correct, adjust, and confirm the data and its interpretation. By integrating user feedback, the system could become more personalized and adaptive to the user, providing a sense of control (cf. [13, 15]).

6 Conclusion & Future Work

With the advances of wearable cognitive sensing technologies, we see a need for ensuring that such devices present meaningful and intuitive data to users. This work explored how users perceive and reflect on their mental workload in everyday situations. We further looked into how they differentiated MWL from stress and which emotions related to both concepts. Our findings show that in users' perception, stress and MWL workload are closely tied together, making them hard to distinguish, and MWL itself is an ambiguous concept. Further, we see that users have trouble reflecting on cognitive data without context and show low trust in the accuracy of data from wearable cognitive sensing technologies. To address these challenges, we propose five design guidelines to help users derive meaningful insights from complex cognitive data.

We see the need for future work to further explore the usage of cognitive sensing technologies in long-term studies across different contexts (i.e., work and leisure), allowing users to engage deeply with their data. Furthermore, more research is needed to investigate different cognitive metrics and interfaces based on our recommendations to validate the effectiveness of our ideas.

References

[1] Norah H. Alsuraykh, Horia A. Maior, Max L. Wilson, Paul Tennent, and Sarah Sharples. 2018. How Stress Affects Functional Near-Infrared Spectroscopy (fNIRS) Measurements of Mental Workload. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3170427.3188646

[2] Norah H. Alsuraykh, Max L. Wilson, Paul Tennent, and Sarah Sharples. 2019. How Stress and Mental Workload are Connected. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare* (Trento, Italy) *(PervasiveHealth'19)*. Association for Computing Machinery, New York, NY, USA, 371–376. doi:10.1145/3329189.3329235

[3] Mark H. Ashcraft and Elizabeth P. Kirk. 2001. The relationship among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General* 130 (2001), 224–237.

[4] Ebrahim Babaei, Tilman Dingler, Benjamin Tag, and Eduardo Velloso. 2025. Should we use the NASA-TLX in HCI? A review of theoretical and methodological issues around Mental Workload Measurement. *International Journal of Human-Computer Studies* 201 (2025), 103515. doi:10.1016/j.ijhcs.2025.103515

[5] Alan Baddeley. 2003. Working memory: looking back and looking forward. *Nature reviews neuroscience* 4, 10 (2003), 829–839.

[6] Marit Bentvelzen, Jasmin Niess, and Paweł W Woźniak. 2023. Designing reflective derived metrics for fitness trackers. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–19.

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101.

[8] Gino Camp, Fred Paas, Remy Rikers, and Jeroen van Merrienboer. 2001. Dynamic problem selection in air traffic control training: A comparison between performance, mental effort and mental efficiency. *Computers in Human Behavior* 17, 5-6 (2001), 575–595.

[9] John Daly, Allan Kellehear, and Marjory Gliksman. 1997. *The Public Health Researcher: A Methodological Approach.* Oxford University Press, Melbourne, Australia.

[10] S.F. Fokkinga and P.M.A. Desmet. 2022. Emotion Typology. https://emotiontypology.com/. Delft University of Technology. Licensed under CC BY-NC-ND 4.0.

[11] Lex Fridman, Bryan Reimer, Bruce Mehler, and William T. Freeman. 2018. Cognitive Load Estimation in the Wild. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3173574.3174226

[12] Anthony WK Gaillard. 1993. Comparing the concepts of mental load and stress. *Ergonomics* 36, 9 (1993), 991–1005.

[13] Niloofar Hezarjaribi, Sepideh Mazrouee, Saied Hemati, Naomi S Chaytor, Martine Perrigue, and Hassan Ghasemzadeh. 2019. Human-in-the-loop learning for personalized diet monitoring from unstructured mobile data. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 9, 4 (2019), 1–24.

[14] Antonio R. Hidalgo-Muñoz, David Mouratille, Nadine Matton, Mickaël Causse, Yannick Rouillard, and Rachid El-Yagoubi. 2018. Cardiovascular correlates of emotional state, cognitive workload and time-on-task effect during a realistic flight simulation. *International Journal of Psychophysiology* 128 (2018), 62–69.

[15] Andreas Holzinger. 2016. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain informatics* 3, 2 (2016), 119–131.

[16] Reza Kazemi, Rashid Haidarimoghadam, Majid Motamedzadeh, Rostam Golmohamadi, Alireza Soltanian, and Mohamad Reza Zoghipaydar. 2016. Effects of shift work on cognitive performance, sleep quality, and sleepiness among petrochemical control room operators. *Journal of circadian rhythms* 14 (2016).

[17] Trevor Kistan, Alessandro Gardi, and Roberto Sabatini. 2018. Machine learning and cognitive ergonomics in air traffic management: Recent developments and considerations for certification. *Aerospace* 5, 4 (2018), 103.

[18] Thomas Kosch, Jakob Karolus, Johannes Zagermann, Harald Reiterer, Albrecht Schmidt, and Paweł W. Woźniak. 2023. A Survey on Measuring Cognitive Workload in Human-Computer Interaction. *ACM Comput. Surv.* 55, 13s, Article 283

(July 2023), 39 pages. doi:10.1145/3582272

[19] Richard S. Lazarus and Susan Folkman. 1984. *Stress, Appraisal, and Coping.* Springer Publishing Company.

[20] Ian Li. 2012. Personal informatics & context: Using context to reveal factors that affect behavior. *Journal of Ambient Intelligence and Smart Environments* 4, 1 (2012), 71–72.

[21] Zilu Liang, Bernd Ploderer, Wanyu Liu, Yukiko Nagata, James Bailey, Lars Kulik, and Yuxuan Li. 2016. SleepExplorer: a visualization tool to make sense of correlations between personal sleep data and contextual factors. *Personal and Ubiquitous Computing* 20 (2016), 985–1000.

[22] Neda Mahdavi, Leili Tapak, Ebrahim Darvishi, Amin Doosti-Irani, and Masoud Shafiee Motlagh. 2024. Unraveling the interplay between mental workload, occupational fatigue, physiological responses and cognitive performance in office workers. *Scientific Reports* 14, 1 (2024), 17866.

[23] Serena Midha, Max L Wilson, and Sarah Sharples. 2022. Lived Experiences of Mental Workload in Everyday Life. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–16. doi:10.1145/3491102.3517690

[24] Evangelos Niforatos, Tianhao He, Athanasios Vourvopoulos, and Michail Giannakos. 2024. Democratizing EEG: Embedding Electroencephalography in a Head-Mounted Display for Ubiquitous Brain-Computer Interfacing. *International Journal of Human–Computer Interaction* (2024), 1–25.

[25] Gerardo Ramirez and Sian L. Beilock. 2011. Writing about testing worries boosts exam performance in the classroom. *Science* 331 (2011), 211–213.

[26] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[27] Christina Schneegass, Max L Wilson, Horia A. Maior, Francesco Chiossi, Anna L Cox, and Jason Wiese. 2023. The Future of Cognitive Personal Informatics. In *Proceedings of the 25th International Conference on Mobile Human-Computer Interaction* (Athens, Greece) *(MobileHCI '23 Companion)*. Association for Computing Machinery, New York, NY, USA, Article 35, 5 pages. doi:10.1145/3565066.3609790

[28] Christina Schneegass, Max L Wilson, Jwan Shaban, Jasmin Niess, Francesco Chiossi, Teodora Mitrevska, and Paweł W Woźniak. 2024. Broadening the mind: how emerging neurotechnology is reshaping HCI and interactive system design. *i-com* 23, 2 (2024), 165–177.

[29] Sarah Sharples and Ted Megaw. 2015. Definition and measurement of human workload. In *Evaluation of Human Work*, John R. Wilson and Sarah Sharples (Eds.). CRC Press.

[30] Christopher D Wickens. 2002. Multiple resources and performance prediction. *Theoretical issues in ergonomics science* 3, 2 (2002), 159–177.

[31] Christopher D Wickens. 2014. The structure of attentional resources. In *Attention and performance VIII*. Psychology Press, 239–257.

[32] Max L Wilson, Serena Midha, Horia A. Maior, Anna L Cox, Lewis L Chuang, and Lachlan D Urquhart. 2022. SIG: Moving from Brain-Computer Interfaces to Personal Cognitive Informatics. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 163, 4 pages. doi:10.1145/3491101.3516402

[33] Max L Wilson, Jwan Shaban, Horia A. Maior, Christina Schneegass, and Anna L Cox. 2024. The CHI'24 Workshop on the Future of Cognitive Personal Informatics. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 491, 6 pages. doi:10.1145/3613905.3636296

[34] Max L Wilson, Natalia Sharon, Horia A Maior, Serena Midha, Michael P Craven, and Sarah Sharples. 2018. Mental workload as personal data: designing a cognitive activity tracker. In *Proceedings of the 3rd Symposium on Computing and Mental Health, Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA.

[35] Paweł W. Woźniak, Monika Zbytniewska, Francisco Kiss, and Jasmin Niess. 2021. Making Sense of Complex Running Metrics Using a Modified Running Shoe. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 50, 11 pages. doi:10.1145/3411764.3445506