The Journal of Systems and Software 210 (2024) 111887

Contents lists available at ScienceDirect

The Journal of Systems & Software

journal homepage: www.elsevier.com/locate/jss

Programming under the influence: On the effect of Heat, Noise, and Alcohol on novice programmers[☆]

Claus Brabrand[a], Nanna Inie[a], Paolo Tell[b,*]

[a] Center for Computing Education Research (CCER), IT University of Copenhagen, Denmark
[b] Department of Computer Science, IT University of Copenhagen, Denmark

ARTICLE INFO

Keywords:
Software engineering
Programming
Cognitive performance
Physical stressors
Environmental stressors

ABSTRACT

When humans are exposed to environmental and physical stressors, cognitive performance is degraded. Even though several studies have examined the effect of various stressors individually, there are limited studies comparing the impact of different types. This study examined the effects of Heat, Noise, and Alcohol on cognitive performance during two programming tasks to quantify the impact of stressors on novice programmers. The experiment enrolled $N = 100$ university student volunteers for a between-subjects experiment. Participants were randomly assigned to one of four conditions ($M = 25$): a room with at 38 °C (100°F), a room with conversational noise around 80 dBA, a blood alcohol content of 1.0‰, or a base condition. Two programming tasks were administered: one analysis task (reading programs) and one synthesis task (writing programs), taking about half an hour to complete in total. Short-term exposure to heat appears to not significantly affect neither reading nor writing programs; conversational noise significantly impacts analytical tasks but not synthesis tasks; while alcohol significantly worsens performance in both analytical and synthesis tasks. To provide a tangible summary for decision-makers able to influence conditions for novice programmers, an approximated comparison is provided, which "translates" negative cognitive effects of heat, noise, and alcohol to one another.

1. Introduction

High cognitive performance, such as the performance required in high-level programming and engineering tasks, depends on resilient cognitive control: the degree to which cognitive functions can withstand, or are resilient to, the effects of stress.

During the Covid-19 pandemic, most screen-based professionals were forced to work from home, and many continue to do so after the pandemic has officially ended, claiming that the home office makes it easier to balance work and personal life (Mitchell, 2023; Ford et al., 2021). At the same time, comfort of the work environment has been cited by software developers as one of the biggest challenges of working from home, especially attributing noise and physical comfort as distractors. 'Tuning out distractions' has been described as difficult in both home and open office environments (Ford et al., 2021).

Even though the effects of physical and environmental stressors on cognitive performance have been studied to a significant degree, commonly on military personnel and professional athletes (e.g., Martin et al., 2020; Taylor et al., 2016; Lieberman et al., 2002; Zhang et al., 2019; Hancock and Vasmatzidis, 2003; Martin et al., 2019), research on their effects on sedentary work is more limited. Such studies have, however, indicated detrimental effects to cognitive performance by, for instance, excessive sunlight (Jamrozik et al., 2019), non-organic interior design materials (Yin et al., 2018; Shen et al., 2020), exposure to high concentrations of CO_2 (Allen et al., 2016; Du et al., 2020) sleep deprivation (Fucci et al., 2020), as well as the combination of several of these factors (Liebl et al., 2012).

However, few studies quantify and *compare* the impact of different physical stressors to cognitive performance. To this end, this article presents a study of three different physical stressors against a baseline condition to begin to formulate a *scale* with which to compare different stressors. We specifically investigate the domain of *programming*. Rather than an arbitrary cognitive performance test, we administered assignments in *reading* programs (analysis) and *writing* programs (synthesis) to university students having taken CS1 (introductory programming).

[☆] Editor: Alexander Serebrenik.

[*] Corresponding author.
E-mail addresses: brabrand@itu.dk (C. Brabrand), nans@itu.dk (N. Inie), pate@itu.dk (P. Tell).

https://doi.org/10.1016/j.jss.2023.111887
Received 12 May 2023; Received in revised form 27 October 2023; Accepted 28 October 2023
Available online 8 November 2023
0164-1212/

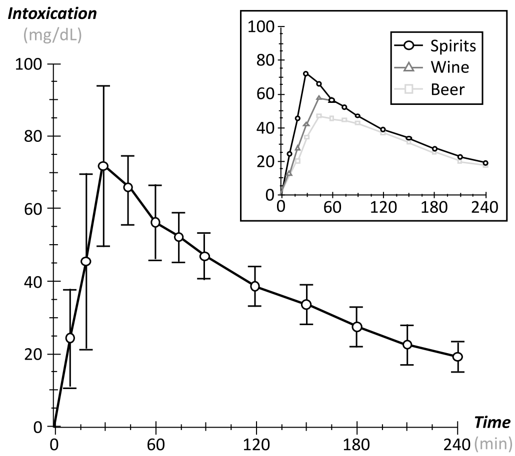


Fig. 1. Mean BAC (blood alcohol concentration) over time for spirits, wine, versus beer.
Source: Reproduced from Jr. et al. (2014), Fig. 1.

The results therefore contribute new knowledge about the impact of physical stressors to cognitive performance on novice programmers.

The study used *Heat* as a proxy for working in a warm office without air conditioning (A/C), *Noise* as a proxy for an open office space where people have conversations while others are working (or a home office where other household members are audible), and *Alcohol* for comparative purposes such that the potential effects on performance can be "translated" into something more tangible than abstract percentage decreases in performance. Compared to the baseline, we found statistically significant detrimental effects on the performance of programmers exposed to *Noise* and *Alcohol*, and weak evidence of detrimental effects on the performance of programmers under the *Heat* condition.

2. Background

When demands are placed on our performance that are greater than what can be carried out by means of automatic cognitive processes alone, we require *cognitive control* (Umemoto et al., 2019). Cognitive control is used to describe the processes, or capacity, by which individuals manage goal-oriented behaviors, and encompasses various behaviors and mental activities (Norman and Shallice, 1986; Mackie et al., 2013; Badre and Nee, 2018). The capacity to exert cognitive control is affected by multiple psychological and physiological factors. The three factors which were studied in this experiment are *heat*, *noise*, and *alcohol*:

2.1. Measuring and determining level of heat

Wet-Bulb Globe Temperature (WBGT) originated in the US military for reliably measuring personal heat stress under various conditions (Budd, 2008). Beyond air temperature (aka, Dry-Bulb Temperature) as measured in degrees Celsius (°C), degrees Fahrenheit (°F), or Kelvin (K), it also incorporates relative humidity percentage (RH%).

Research on the cognitive impact of heat exhaustion distinguishes between (superficial) sensory discomfort versus (internal) core body temperature overheating (Taylor et al., 2016). Our investigation considers the former (sensory discomfort). Recent research shows that a hot environment significantly reduces complex cognitive task performance even before any increase in core temperature (Gaoua et al., 2012). Arguably programming can be classified as a complex task.

The interval from 17 °C to 23 °C is designated as the *zone of preferred temperature* (Ramsey et al., 1983). For an overview of the impact of heat from 17 °C to 55 °C (after prolonged exposure or strenuous activity at 50% RH), we refer to Fig. A.5(a).

2.2. Measuring and determining level of noise

Prior research (especially from laboratory studies) show that noise exposure negatively impacts performance (Sorkin, 1988). According to a field study involving more than two thousand participants, more than half (54%) reported they were often bothered by noise, especially by people talking and telephones ringing (Sundstrom et al., 1994).

Research shows that human speech is much more detrimental than so-called "white noise", compared to a quiet environment (Jones et al., 2008) and that *reading* and *recollection* is negatively impacted by noise from nearby human speech, but not from non-speech noise (Salamé and Baddeley, 1982). Our investigation considers the impact from nearby human speech. Noise is measured in decibel (dB), but, since the sensitivity of the human ear depends on the frequency, an *adjusted* decibel scale is often used (dBA) for measuring noise impact on people.

For an idea of what noise levels from 30 dBA to 120 dBA correspond to, please consult Fig. A.5(b).

2.3. Measuring alcohol intoxication

Alcohol obviously impacts performance. Widmark's formula (Watson et al., 1981), serves to calculate (predict) how much alcohol each person needs to drink to attain a particular level of intoxication:

$$A = w \cdot c \cdot \rho$$

where A is the volume of pure *alcohol* to drink (in mL); w is the *weight* of the individual (in kg); c is the blood-alcohol *concentration* (in per mille) to be attained (in our case, 1.00‰ under our *Alcohol* condition); and ρ is a *gender-specific* constant known as the *distribution-volume* or *water-phase* in which the alcohol is diluted: $\rho_{♂} = 0.68$ L/kg for *men*, and $\rho_{♀} = 0.55$ L/kg for *women*. (Hence, men need to drink about 25% more than women in order to reach the same level of intoxication.) Since, we are using 40% vodka, we translate the amount of alcohol to drink from *volume* (in mL) to *weight* (in grams) by dividing with the density of `40%Vol` *volume percentage* EtOH (ethanol) which is 0.3447 mL/g at room temperature. Hence, in order for the first author (male, 80 kg) to attain 1.00‰ of intoxication (after $T_{max} = 36$ min on an empty stomach), he would need to drink the following (weight) amount of 40% Vodka:

$$(80 \text{ kg} \cdot 1.00 \text{ ‰} \cdot 0.68 \text{ L/kg}) \;/\; 0.3447 \text{ mL/g} \approx \underline{158} \text{ g}$$

In the name of Science, two of the authors tested the predictive power of the formula on themselves with the intent of reaching an intoxication of 1.00‰; they attained measured maximal concentrations of $C_{max} = 0.95$‰, respectively, 0.91‰, according to a conservatively under-approximating police-grade breathalyzer.

The alcohol concentration in exhaled air is an *indirect measure* for the blood alcohol concentration (BAC). The individual variation is 1800 to 3000 times lower in the exhaled air than the blood. For legal reasons, the Danish national conversion factor for breathalizers has deliberately been set to a conservatively low ratio of 2000 (T. D. M. of Justice, 2007).

The speed of absorption appears to depend on the type of alcoholic beverage: "The time to Cmax [maximum concentration] occurred significantly earlier ($p < 0.01$) after [20%] vodka/tonic (36 ± 10 min) compared to [$12\frac{1}{2}$%] wine (54 ± 14 min) or [5.1%] beer (62 ± 23 min)" (cf. Fig. 1) (Jr. et al., 2014). For comparability and since it is absorbed the fastest, we settled on using 20% vodka/tonic (i.e., half 40% vodka, half tonic). For an idea of the effects of alcohol intoxication from 0.0‰ to 1.5‰ BAC, see Fig. A.5(c).

3. Related work

Studies of environmental and physical stressors on novice programmers and software developers, specifically, are limited. Muller & Fritz's study from 2016 (Müller and Fritz, 2016) indicated that bio-metrics can be used to predict code quality, thus it is pertinent to identify and investigate different environmental factors seem to affect software engineers and to which degree.

Related research has studied factors such as the consequences of *affect* (Graziotin et al., 2017, 2018), stress (Liu et al., 2021), and sleep (Fucci et al., 2020) for software developers. The latter study found, for instance, that missing a single night's sleep led to a reduction in implementation quality of 50%. The study also uncovered some interesting insights about the specific behavioral effects of sleep deprivation on *software development*, indicating that sleep-deprived developers make more fixes to syntactic errors in the source code (Fucci et al., 2020). Related studies have shown that software developers often work under stress, and that simple interventions to their physical state (such as *breathing* exercises) can have positive effects to both attention awareness, well-being, perceived productivity, and self-efficacy (Penzenstadler et al., 2021).

As numerous studies have shown that the nature of the task performed and the familiarity of the individual with the task affects the degree impact with which cognitive performance is impaired (Taylor et al., 2016; Martin et al., 2019), the current study focuses directly on physical stressors on novice programmers (university students) performing programming tasks. This section presents related studies of cognitive performance under the three stressors: *heat*, *noise*, and *alcohol*.

3.1. A gap in research on moderate heat exposure

Yegeneh et al. conducted a literature review of 45 studies of ambient air temperature and cognitive performance published between 1980 and 2018 (Yeganeh et al., 2018). They unequivocally found that heat reduces cognitive performance – at a room temperature of 80°F (26.7 °C), they saw a reduction in performance of almost 8%. This literature review showed that heat stress causes the most significant decline in the most attention-demanding tasks, and that the estimated temperature-performance correlation follows a bell-shaped curve centered around the average control temperature.

Heat can also contribute to a decline in cognitive performance by dehydrating the body. A 2% body weight loss due to hydration (constituting "mild" dehydration Périard et al., 2021) can cause significant effects to both cognitive, physical, visuomotor, and psychomotor performance (Grandjean and Grandjean, 2007; Cian et al., 2001), and the dehydration effect can lead to prolonged exhaustion (Cian et al., 2001).

However, accurately quantifying and comparing the impact of heat on cognitive performance is difficult due to "the large number of factors that come into play, such as task type, exposure duration, skill, and acclimatization level of the individual and due to the absence of a concise theory on which experimental work can be based", according to a thorough review from 2003 by Hancock and Vasmatzidis (Hancock and Vasmatzidis, 2003). A later review pointed out that a large part of research on thermal environment impact has been focused on severe heat and cold exposure rather than *moderate* heat exposure in indoor environments, which is more common in most office environments (Zhang et al., 2019). This review also concludes that studies of moderate heat exposure in various fields have reported "inconsistent results", likely due to the vast amount of confounding factors, "methodological discrepancies between studies" and incompatible conceptual models (Zhang et al., 2019).

3.2. Noise impact depends on individual resilience

Environmental noise and its effects have also been studied in relation to office work, particularly incited by the proliferation of open-plan office environments. Studies have shown detrimental effects of noise to job satisfaction (Sundstrom et al., 1994), motivation (Evans and Johnson, 2000; Jahncke et al., 2011), tiredness (Jahncke et al., 2011), concentration (Banbury and Berry, 2005), productivity (Mak and Lui, 2012), and even long term sickness absence (Clausen et al., 2013).

All of the effects mentioned above, however, are based on self-appraisal, rather than physiological measurements; several of the studies found *no* physical impact of noise to their subjects (Evans and Johnson, 2000; Banbury and Berry, 2005; Jahncke et al., 2011), suggesting that noise impact is primarily cognitive and behavioral, and thus volatile to cognitive resilience, or ability to direct and focus attention.

Experimental studies of cognitive performance have unanimously shown adverse effects of noise to speed, accuracy, and memory (Schlittmeier et al., 2008; Jahncke, 2012; Brocolini et al., 2016; Jahncke and Hallman, 2020; Meng et al., 2021). *Speech* is one of the most disturbing types of background noise, as human cognition is exceptionally sensitive to communication: "Even if background speech is irrelevant and one intends to ignore it, the speech signal is automatically and obligatorily processed by the listener's auditory-perceptual and cognitive systems. Evidence suggests that this obligatory processing may include semantic analysis of background speech." (Schlittmeier, 2021).

Research in the area has been focused mainly on two aspects: exploring and comparing the impact of different types of noise (i.e. traffic noise, human speech intelligibility, nature sounds, air conditioning Ljungberg and Neely, 2007; Schlittmeier et al., 2008; Meng et al., 2021), and identifying mitigation of detrimental noise effects, for instance masking sounds and restoration periods (Brocolini et al., 2016; Jahncke, 2012). Most studies have also been centered around "office workers" generally, which is a fairly broad generalization. Several studies have shown that how much cognitive performance is impaired by noise varies with the cognitive processes required by the tasks, the individual is engaging in Jahncke (2012). We therefore find it highly relevant to investigate the influence of noise to performance in software development tasks, specifically.

3.3. Alcohol intoxication impairs executive control, but may increase experience of creative performance

Both research and the general public has more comprehensive knowledge of the effects of alcohol on cognitive performance—to the extent where descriptive scales of levels of intoxication have been developed (e.g. Fig. A.5(c)). Alcohol intoxication generally impairs executive control, which makes it detrimental to deep thinking and concentration (among other cognitive abilities such as verbal fluency, planning, memory, and complex motor control), e.g. Peterson et al. (1990), Giancola et al. (1996) and Dry et al. (2012).

Psychoactive substance use has been found to be prevalent in professional developers, with alcohol and cannabis shown to be the far most commonly used during completion of programming tasks (Endres et al., 2022; Newman et al., 2023). Developers in the survey believed that recreational use of the substances increased their programming performance and "got them in the zone". However, formalized studies of creativity under alcohol or cannabis intoxication have not confirmed this (Benedek et al., 2017; Benedek and Zöhrer, 2020; Kowal et al., 2015). Several studies have shown that mild intoxication increases a person's *joviality*, and makes them more likely to evaluate their own and others' ideas more favorably (Kowal et al., 2015; Heng et al., 2022). Some studies have even shown enhancement of creative output for individuals who *believed* they had consumed alcohol, whether or not they had actually done so (Lapp et al., 1994).

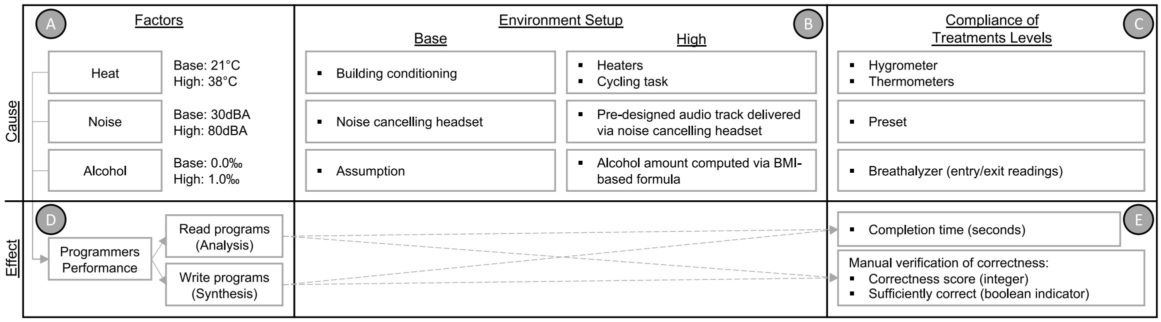


Fig. 2. Overview of the experimental design setting.

Although general trends in consumption of alcohol varied between countries during the Covid-19 lockdowns (Plata et al., 2022), many countries experienced a surge in alcohol sales during the pandemic. Such rises in sales have been explained by difficulty of coping with loneliness and isolation, but also by the "home mixology movement", leading people to buy more premium alcohol to replicate at home some of the experiences they were missing at bars and restaurants. Since most home offices do not carry a no-alcohol policy, it begs the question of whether more programming tasks are performed under the influence during remote work, and how this might influence the quality of the work performed.

Impairment of cognitive processes due to higher blood alcohol level (BAC) does not appear to be uniform across different tasks, neither does different doses of alcohol affect different processes equally (Dry et al., 2012). Some studies have indicated that performance in certain tasks (psychomotor, set-shifting, or working memory abilities) may not be affected by low doses of alcohol (Hoffman and Nixon, 2015). Alcohol intoxication has, however, most commonly been studied in either by anecdotal accounts or in controlled experiments using formalized cognitive performance batteries rather than realistic work assignments such as software development and programming.

4. Methodology

Following guidelines and recommendations by Juristo and Moreno (2013), Wohlin et al. (2012), and Kitchenham et al. (2002), we performed a quasi-experiment designed to investigate the effect of heat, noise, and alcohol on novice programmers. In this section, we present details regarding its design and implementation with regards to: research question, independent and dependent variables, hypotheses tested to investigate the research questions, participants, concrete design, and data collection and analysis procedures. A summary of the experimental setting is provided in Fig. 2.

4.1. Objective

Our objective is to collect quantitative evidence regarding adversarial effects of the three external factors (*heat*, *noise*, and *alcohol*) on novice programmers. The operational research question is:

> **RQ:** To what extent does *heat*, *noise*, and *alcohol* impact the ability of university student novice programmers to *read* programs (*analysis*) and *write* programs (*synthesis*)?

In the context of this experiment, performance has been defined as a combination of a task *correctness score* (along with a derived *sufficiently correct* boolean indicator) and task *completion time*. See the 'Cause and Effect Factors' sections of Fig. 2 for a visual overview.

4.2. Experimental setup

We had two aims for the experiment design: (1) it should be easily replicable, allowing parallelization and future reproduction, and (2) it should allow the tuning of each of the three physical stressors independently. In the following, we will describe for each factor: (i) which treatment levels were chosen and why (Fig. 2.A), (ii) how treatment levels were controlled in practice (Fig. 2.B), and (iii) the process and instruments used to verify the compliance of each experiment run with the treatment levels (Fig. 2.C).

Each independent variable (heat, noise, alcohol) was tested in isolation, hence, together with the base condition (i.e., the control condition), three treatment setups were used in which one independent variable would be set to high and the two remaining ones were equal to base level.

The design yields three *treatments*, each modifying the base (control) condition ⟨ 21 °C, 30 dBA, 0.0‰ ⟩:

- ⟨ **38 °C**, 30 dBA, 0.0‰ ⟩: **heat** condition;
- ⟨ 21 °C, **80 dBA**, 0.0‰ ⟩: **noise** condition; *and*
- ⟨ 21 °C, 30 dBA, **1.0‰** ⟩: **alcohol** condition.

The temperatures of 21 °C and 38 °C correspond to 70°F and 100°F, respectively.

4.2.1. Heat

Treatment levels: Levels for the heat factors were set to be maintained at the building standard temperature for the base condition (21 °C) and raised to 38 °C (100°F) at 50% relative humidity for the heat treatment. This level was chosen because it would create a sufficient raise in body temperature to impact cognitive capacity, yet not so high that subjects would risk acute heat disorders (above 39 °C) (Kjellstrom et al., 2009). It is a realistic temperature for an indoor (summer) work environment without air conditioning, such as the home office work environments many were compelled to work in during the Covid-19 lockdowns worldwide.

Environmental setup: Base conditions were ensured through the building ventilation system. To achieve the high heat treatment, an experimental room was insulated with rockwool over windows sealed with plastic sheets and equipped with two electric heaters (9 kW and 2 kW). Additionally, a stationary bicycle was placed in the room to increase the core temperature of the participants before the execution of the experimental tasks. The bicycle was used to slightly raise body core temperature, simulating the effects of longer term heat-exposure. Subjects were asked to use the stationary bicycle for 10 min, a compromise for raising body core temperature without risking dehydration or exhaustion (Maughan and Shirreffs, 2004). To further avoid dehydration, water was available to participants at all times.

Verification of compliance: This setup was piloted at the start of each period in which experiments were run to test whether the power of the heaters was enough to reach the desired temperature. The temperature in the high treatment level room was continuously monitored using three Dallas DS18B20 sensors hooked up to Arduino (Things Uno) boards via a 1-wire interface connection. The temperature was measured every 120 s, sent via a LoRaWAN connection to an InfluxDB and shown via a Grafana Web interface. Additionally, a hygrometer was placed in the room to measure humidity allowing the calculation and monitoring of the wet-bulb globe temperature (WBGT) (Carter et al., 2020).

4.2.2. Noise

Treatment levels: Following indications based on multiple studies reported by Maxwell (Maxwell, 2015) indicating open office workspaces levels between 42 dBA and 60 dBA, and in line with similar experiments using 42 dBA as the low threshold (Romano et al., 2018), we decided to ensure the base condition of the noise factor to be below 42 dBA. Similarly, following indications from the Directive 2003/10/EC of the European Parliament (Parliament, 2003) marking at 85 dBA the maximum noise level to ensure health safety in case of constant noise exposure and in line with similar experiments (Romano et al., 2018), we decided to ensure the high level of the noise factor to be below 85 dBA.

The noise level was delivered via a pre-designed audio track sampled from multiple controversial podcasts in the native language of the country in which the experiment was performed. This was chosen for several reasons: (i) human speech is a "major type of practical distractive noise" (Szalma and Hancock, 2011; Tang and Wong, 1998), (ii) Native language was chosen to ensure highest possible comprehension of the audio track or "meaningful speech" (Schlittmeier, 2021), (iii) diverse controversial content would most probably trigger the attention of the participants, mimicking interesting conversations at a work-place.

Environmental setup: To ensure the control of the noise factor (and avoid interference from random unexpected environmental sounds, such as sirens or building sites), we employed an active noise cancelling headset (model: Bose QC35 II) for each experimental room. The room allocated to deliver the high treatment level was always equipped with a cellphone connected to the headset to deliver the pre-designed audio track using tested settings.

Verification of compliance: We measured the sound pressure level (SPL) of the speech signal, played back from an iPhone 8 at the second-highest volume setting, through Bose QC35 II headphones with noise cancellation turned on at the highest setting, using a calibrated Norsonic 139 sound-level meter (SLM) coupled to a G.R.A.S. artificial ear (RA0039, IEC 60318-1). Only the left channel was recorded, as the right channel of the stereo signal was similar in playback level. The SLM measured the A-weighted SPL with a slow time-constant of one second for integration (i.e., calculating the short-term average level once every second).

4.2.3. Alcohol

Treatment levels: Assuming students and knowledge workers operate in a sober state, the base condition was set to 0.0‰ and the high condition to 1.0‰. This measure is above the legal limit in most of the United States, and double the legal value for driving in many European countries (including Denmark). Although 0.5‰ would most likely already cause negative effects when performing physical tasks including coordination, we chose a BAC of 1.0‰, which should be certain to provoke cognitive impairment and loss of judgment (see Fig. A.5(c) in the Appendix), even for individuals who may have developed a functional tolerance to alcohol due to regular consumption (chronic tolerance) (Tabakoff et al., 1986).

Environmental setup: Base conditions were not verified since "drinking in class [is] not a common practice" for Danish students (Ladekjær Larsen et al., 2016). All experiment runs for the high alcohol treatment were scheduled in the mornings and all subjects participating in those mornings were sent a preliminary experiment information including the request to attend the session fasting. The height and weight of subjects exposed the high treatment were measured, and the alcohol amount determined by the formula described in Section 2.3 was diluted in a solution of equal weight of tonic water to ease assimilation. The alcohol used was 40% *Absolut Vodka*®.

Verification of compliance: An ideal monitoring of the alcohol level of a participant would require constant analysis of blood samples extracted from a participant exposed to the alcohol. This would have been unfeasible for several reasons, primarily that it would have required interruptions throughout the executions of the tasks for sample gathering. We calculated the amount of alcohol based on sex and body mass index (BMI) that would be required to reach the 1.0‰ level and a breathalyzer for quantitative monitoring. In particular, after acquiring the height and weight to compute the amount, the solution was brewed and presented to the participant. A timer was started, and participants were asked – not enforced – to finish the solution within 5 min. At 25 min, participants were asked to use the breathalyzer to record the alcohol level ($M=25$, $\bar{x}=.60‰$). Once recorded, they were asked to start the experiment and, once finished ($M=25$, $\bar{x}=54'25''$), alcohol levels were measured again ($M=25$, $\bar{x}=.59‰$). Fig. 3(c) presents an overview of the alignment of the measurements with the ones predicted from reference.

4.3. Intervention: Programming tasks

Two programming tasks (listed in Fig. A.6 in the Appendix) were administered as one *analysis* task (T1) and one *synthesis* task (T2); i.e., the ability to *read* and *write* programs, respectively. These represent different skills required of software developers, as well as the higher levels in Bloom's learning taxonomy, e.g. Selby (2015). The tasks mimic simple but realistic programming tasks for both students and software development professionals. The difficulty level of the tasks was equal to what would be expected from any student who has completed a CS1 course; the level of the tasks were assessed by two individual teachers on one such course.

The analysis task (T1) involved predicting the correct output (nine numbers) produced by two nested `for`-loops wrapped around a conditional `if-else` statement for which either branch featured a `print`-statement outputting the result of an arithmetic expression. This task is similar (although slightly more intricate) to what the students have been exposed to during their CS1 course.

The synthesis task (T2) involved writing a program that counts the number of four-letter words ending in 'e'. In terms of difficulty, task T2 is similar to that of the FizzBuzz problem, often used by professional recruiters (Ghory, 2007). Canonical solutions to T2 as well as FizzBuzz both involve (a `for`-loop) *iteration* (through a string, respectively, through the numbers 1–100) wrapped around *conditional* `if-else` *statements* (testing for divisibility with 3 or 5, respectively, whether a word of length 4 is encountered). In addition, either problem features an *exceptional case* (when a particular number is encountered ($3 \times 5 = 15$), respectively, when the a particular letter ('e') is encountered at the end of a word). In fact, later editions of our CS1 course has featured FizzBuzz as a mandatory exercise; and it has even been used it as the experimental task of another controlled experiment (Kristiansen et al., 2023).

The analysis task was administered and solved on paper, while the synthesis task was administered on paper, but solved on subjects' own personal laptops (to make the setting as natural as possible).

4.4. Task assessment

The tasks were graded by two of the authors, thoroughly marking and counting different types of errors in continuous discussion. All student solutions were assessed from printed paper anonymized so as to hide (blind) the treatment exposure (heat, noise, alcohol, xor base) from the authors performing the assessment. For both tasks, the dependent variable, *performance* (of a participant), was measured by a *correctness score* and from that a derived *sufficiently correct* (boolean indicator) as well as task *completion time* (in seconds). Thus, the performance of a participant for each of the tasks was thus captured by a triplet:

$(correctness\ score, sufficiently\ correct, completion\ time)$.

Reading programs. For task 1, the *correctness score* (an integer from 0 to 10) was computed quantitatively as the number of correct outputs of the program as predicted by the student with an extra point for correctly predicting the number of outputs (nine). The correct answer was nine consecutive numbers: 2, 1, 0, 8, 5, 2, 20, 13, 6. In addition to the richer 0–10 *correctness score*, we also derived a simpler *sufficiently correct* (boolean indicator) which deems a solution "(sufficiently) correct" whenever a student predicts all-but-one of the numbers. (The reason behind the slight error tolerance in scoring comes from the Danish grade scale where the maximal grade tolerates a bit of imperfection;[1] see, e.g., Appendix A of Brabr and Dahl (2009).) A maximum task execution time (12 min) was imposed to represent the inability of a participant to finish the task.

Writing programs. For task 2, two of the authors assessed code solution printouts. The code solutions were graded qualitatively by looking closely at six distinct qualities a correct solution would have. Each of the six qualities were assigned a distinct color (were, for instance, *orange* corresponded to code checking for an ‘e’). The corresponding code lines were highlighted on the printouts for easy post-assessment score discussion and reconciliation:

loop through input and	(gray)
...count the number of	(green)
...words that have	(purple)
...exactly 4 characters	(blue)
...and end with	(yellow)
...an ‘e’.	(orange)

Typically incorrect solutions and how to score them were discussed and reconciled at a meeting between the two assessing authors, minor scoring differences were resolved by taking the average of the scores among the two authors. Each quality (color) that was addressed in the code was awarded one point, cumulatively giving rise to a score from 0 to 6 (with the possibility of half points due to averaging between the two authors). Task 2 was assessed qualitatively so as to provide a richer score than a quantitative (boolean) yes/no functional correctness score. A solution was deemed *sufficiently correct* if it incorporated *sufficient* aspects of a correct solution (a correctness score of at least 4[1]).

4.5. Subjects

Programming novice participants were recruited from the student population of the IT University of Copenhagen (ITU). To qualify for admission, students had to have successfully completed the CS1 (Introductory Programming) course, which is a mandatory first semester course of the Bachelor in Software Development. At ITU, CS1 is a 15 ECTS[2] first-semester course teaching the basics of imperative and object-oriented programming with a four-week group project in the end.

Neither gender, age, nor other distinctions were relevant. However, with respect to alcohol condition, as the formula used to calculate the amount of alcohol to administer to reach a given treatment level is based on the BMI level ranging from 18½ to 30 (Jr. et al., 2014). To avoid discrimination in recruitment, this cutoff was applied both during recruitment and when assigning treatments to participants. Eventually, no data point had to be dropped for this reason.

Table 1 shows demographic information about the participants.

Table 1
Demographic information.

Age[a]	# or %[b]
18–19	7
20–21	27
22–23	38
24–25	10
> 26	14
Unspecified	4

Gender (self-reported)	# or %[b]
Female	24
Male	76
Study programme	**# or %[b]**
Software Development (BSc)	80
Computer Science (MSc)	13
Other (IT-related)	7

[a] The legal age for purchasing and consuming alcohol in Denmark is 18 years.
[b] Since $N = 100$, each number of participants also corresponds to the percentage.

4.6. Ethical considerations

The study underwent review and received ethical approval by the university. All rooms had dedicated monitoring personnel during the experiments. All participation was based on volunteering. No extra credit or monetary reimbursement was offered, to ensure subjects were under no pressure to participate. All subjects were offered snacks and soft drinks after the experiment. No sensitive data was collected about the participants (except for weight, for those who participated in the alcohol condition, and this was only used to calculate required amount of intake), and all data was anonymized in compliance with GDPR regulations.

4.7. Design

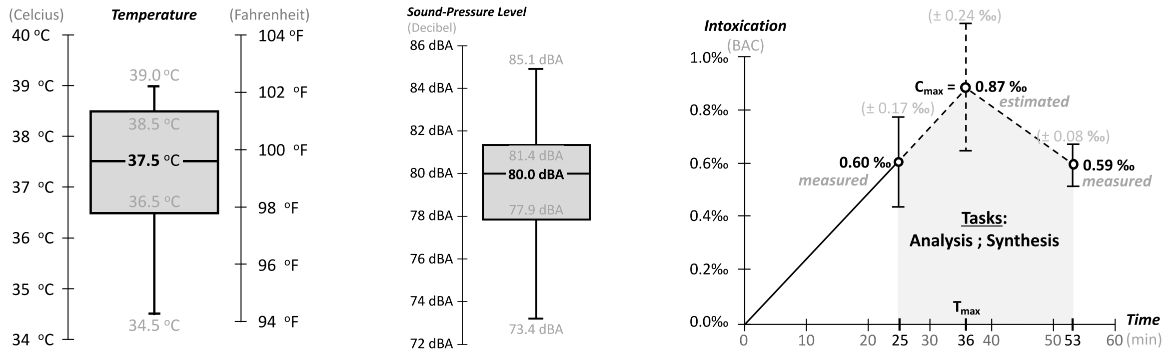
The experiment was run as a between-subject study. Participants first signed an informed consent form, where they agreed to potentially be exposed to one of the four conditions. Since the heat and alcohol conditions required some setup (the heat condition required significant preparation to insulate and heat up the room, and the alcohol condition required participants to show up fasting; therefore it was only scheduled in the morning), participants were assigned quasi-randomly to conditions based on their availability.

Upon arrival at their scheduled experiment date and place, participants in the alcohol condition were weighed to determine amount of alcohol to consume (as described in Section 4.2.3), and participants of the heat condition were asked to spend 10 min cycling on a stationary bike (as per the description in Section 4.2.1.) Participants of the alcohol condition were asked to consume the alcohol, wait 25 min (under supervision), and their BAC level was measured and noted.

The tasks were then administered, and a dedicated research assistant (in some instances, one of the authors) supervised the completion of the tasks, as well as the compliance of the relevant condition in the room. Participants were given a maximum of 12 min to complete each task.

Upon completion of the tasks or the elapsed 2 × 12 minutes, each participant was interviewed by one of the authors about their

[1] [Translated from Danish]: The (maximal) grade in the Danish scale (12) is awarded for an excellent performance which completely meets the course objectives, with no **or only a few insignificant weaknesses**.
[2] One academic year is 60 ECTS (European Credit Transfer and Accumulation System).



(a) *Heat*. (b) *Noise*. (c) *Alcohol*.

Fig. 3. Verification of compliance by quantitative metrics on *Heat*, *Noise*, and *Alcohol* as captured by a thermometer, sound-level meter, and alcometer in Celsius (°C), decibel (dBA), and BAC (blood-alcohol concentration) per mille (‰), respectively.

experience. All participants (especially relevant for those under the heat and alcohol condition) were invited into a supervised "recovery" room, where they were offered snacks and soft drinks. Participants in the alcohol condition were thoroughly monitored and were asked not to leave the recovery room before their BAC had dropped to the legal limit for driving or below, unless they signed an explicit waiver. This did not cause any form of conflict or friction, and virtually all participants were happy to "wait it out".

4.8. Data analysis

Whenever we compared normally distributed data vectors (the performance scores or the time spent on a task), we used a T-test; for non-parametric data we used the Mann–Whitney U-test. For ascertaining whether the data is normally distributed, we used the Kolmogorov–Smirnov Test. For comparing two population proportions (the ratio performing a task appropriately under a treatment as compared to the base condition), we will use the Z-test.

We used *one*-tailed tests when comparing an adversarial condition (heat, noise, or alcohol) versus the base condition, disregarding the possibility that people perform *better* under the physical stress conditions than under the base one—a reasonable assumption in line with recommendations for a study under *inferior* conditions (Kimmel, 1957). We adopt a 95% confidence interval with an α of 5%.

5. Results

We begin by reporting the results of the compliance verification, then we consider the results of reading programs (analysis) followed by that of writing programs (synthesis).

5.1. Compliance verification

For all conditions of the experiment, we used sensor equipment to measure the levels of *Heat*, *Noise*, and *Alcohol* in order to quantify compliance with the experimental setup.

Heat: Fig. 3(a) shows a boxplot of the measurements under the heat condition. The temperature measured was 37.5 ± 2.8 °C; the *lowest* spike in temperature was 34.5 °C (94°F; most likely in connection with the experiment room door opening to allow for a participant to enter or leave), the *highest* was 39.0 °C (104°F). The relative humidity (RH) stayed around 50% during the experiment and never went outside the range from 45% to 55%. In terms of *heat stress*, a temperature of 37.5 °C at 50% RH corresponds to a Wet-Bulb Globe Temperature (integrating temperature and humidity) of 28.7 °C WBGT (83.7°F WBGT). The *Heat* condition thus represents an exposure to approximately **37.5 °C** **(99.5°F)** (at 50% RH) corresponding to the experimental design.

Noise: The average SPL (Sound Pressure Level) over the first five minutes (300 s) of the signal was 79.6 dBA, with a median of 80.0 dBA, a standard deviation of ±2.8 dBA, and a maximum of 85.1 dBA (attained only once, for one second). Fig. 3(b) shows a boxplot of the noise readings in decibel (dBA) from the SLM. We note that we stayed clear of the EU regulatory maximum sound level for personal music players (Parliament, 2003) of 85 dBA, which is regarded as workplace safety-critical intensity (Welleschik, 1979). The *Noise* condition thus represents an exposure to approximately **80 dBA** as we aimed for.

Alcohol: Fig. 3(c) shows the blood-alcohol percentage (BAC) in per mille (‰) as measured by the police-grade Lion Alcometer®700 breathalyzer. According to our experiment protocol, we conducted a reading 25 min after ingesting the alcohol, just before the tasks; on average, the subjects attained a concentration of 0.60‰ with a standard deviation of ±0.17‰. Recall that the intoxication increases linearly until maximum concentration ($C_{\max}$) is reached $T_{\max}$ = 36 min after ingestion (Jr. et al., 2014). If we extrapolate from our data, we arrive at an *estimated maximum concentration* of 0.87 ± 0.24‰ which is fairly close to the *intended maximum concentration* of 1.0‰. We also did a measurement upon exit from the experiment room which occurred 53 ± 5 min after ingestion and produced a reading of 0.59 ± 0.08‰. According to the alcometer, the *average concentration* during the tasks was 0.73‰ (see the dashed line of Fig. 3(c)). If we instead calculate the *average intoxication* based on Widmark's formula, we get an *average concentration* of **0.85‰** during the tasks, close to the intended 1.00‰ goal.

We hypothesize that the lower BAC than anticipated might be due to *functional tolerance* of subjects. The participating students are probably accustomed to consuming alcohol on a regular basis. In Denmark, on-campus "Friday Bars" are normal, where many students (and faculty) go for a beer on Friday afternoons. Danes are generally among the heaviest drinkers in Europe (OECD, 2019), and 89% of young Danes between 15–25 drink alcohol at least once a month (19% drink alcohol at least twice a week) (Hansen et al., 2022). An experiment conducted on senior US college students that frequently exceed legal intoxication levels (Friel et al., 1995) proposes a revision of Widmark's constants to $\rho_{♂}$ = 0.71 and $\rho_{♀}$ = 0.65 as well as *delays* the time of maximum concentration to $T_{\max}$ = 39.6 minutes. Extrapolating from this *later* apex, gives our subjects an estimated maximum concentration of $C_{\max}$ = 0.96‰, matching our intended **1.00‰** target.

5.2. Reading programs (analysis)

The top rows of Table 2 shows the *average* and *normalized correctness score*, *completion time*, along with the *ratio* of *sufficiently correct* answers

Table 2
Results for reading (analysis task: T1) and writing programs (synthesis task: T2) under the influence of *Heat*, *Noise*, and *Alcohol* compared to the *Base* condition. Correctness score and completion time (for subjects completing the tasks within the 12' time limit) are analyzed using a T-test whenever the data is normally distributed, U-test otherwise (†); for sufficiently correct, a Z-test is used for comparing the ratios. Statistical significance ($p \leq 0.05$) is indicated with a star (⋆).

Task	*Treatment*	Correctness score			Sufficiently correct			Completion time			Interpretation
		Average	Normalized	*p*-value	Ratio	Normalized	*p*-value	Average	Normalized	*p*-value	
Analysis	*Base*	8.4	**100%**		20/25	100%		6’25”	100%		
	Heat	7.8	**92%**	0.43†	17/25	85%	0.17	5’12”	81%	0.12	inconclusive
	Noise	6.9	**82%**	0.25†	14/25	70%	0.034 ⋆	6’20”	99%	0.45	significantly worse
	Alcohol	7.0	**83%**	0.12†	14/25	70%	0.034 ⋆	7’03”	110%	0.20	significantly worse
Synthesis	*Base*	4.2	**100%**		17/25	100%		9’30”	100%		
	Heat	3.9	**92%**	0.25	17/25	100%	0.50	9’24”	99%	0.46	inconclusive
	Noise	3.7	**89%**	0.25†	14/25	82%	0.19	9’13”	97%	0.41	inconclusive
	Alcohol	3.4	**82%**	0.029 ⋆	10/25	59%	0.023 ⋆	9’20”	98%	0.45	significantly worse

for the analytical read program task under the four conditions: *Base*, *Heat*, *Noise*, and *Alcohol*.

Base: Among the M = 25 participants assigned to the *Base* condition, the average correctness score was $\mu = 8.4$ which we normalize to **100%** and use for subsequent comparison. The sufficiently correct ratio shows that 20 out of the 25 participants assigned to the base condition correctly predicted all-but-one of the outputs. Again, we normalize this to 100% for subsequent comparison. Finally, we see that the average time among the subjects that completed the task within the 12’ time limit was 6 min and 25 s (normalized as 100%).

Before we consider the three treatments, we duly note that they all – as expected – resulted in *inferior* scores; i.e., *below* the normalized baseline score of **100%**.

Heat: Under the *Heat* condition, we observe that the average correctness score drops to $\mu = 7.8$ corresponding to **92%** (compared to the *Base* condition). However, the drop in correctness score is *not* statistically significant ($p = 0.43$) based on a U-test (since the data is not normally distributed). The correctness ratio drops insignificantly ($p = 0.17$) to 17 out of 25 (which is 85%). Interestingly, participants are *faster* although not significantly so ($p = 0.12$) under the heat condition, taking only 5’12” to finish the task, on average. Adopting a conventional 95% confidence interval, this means that, statistically, the results are *inconclusive* in establishing that the *Heat* condition is any “worse” than that of *Base*. Considering we see a drop in the absolute score, it is possible that a larger group of participants would reveal significant results.

Noise: For the *Noise* condition, the average correctness score drops even further to $\mu = 6.9$ corresponding to **82%** of the *Base* score. However, the decrease is insignificant ($p = 0.25$ based on a U-test). Sufficient correctness drops from 20 to 14 out of 25 which is statistically significant ($p = 0.034$). The average time to complete the task is almost the same as that of the base condition. In conclusion, correctness under the noise condition is *significantly worse* than the base condition which is recorded to the far right in the top rows of Table 2.

Alcohol: Finally, the *Alcohol* condition yields an average correctness score of $\mu = 7.0$, corresponding to **83%** of the base condition. Again, the drop is insignificant ($p = 0.12$). Sufficient correctness is the same as that of noise and thus again statistically significantly worse than the base condition ($p = 0.034$). Students subjected to alcohol are generally *slower*, taking a bit more than seven minutes, on average, to complete the analysis task. (Although slower, the difference is not statistically significant.) In summary, correctness under the alcohol condition is *significantly worse* than the base condition.

5.3. Writing programs (synthesis)

The bottom rows of Table 2 presents the results of *writing* programs with the synthesis task of implementing (writing) a Java method based on a specification.

Base: Under the base condition, the average correctness score was $\mu = 4.2$ which, again, we normalize to **100%**. Of the 25 solutions, 17 were deemed sufficiently correct which we normalize to 100%. Among the subjects finished the task within the 12’ time limit, the average completion time was nine and a half minutes. Again, we normalize this to 100%.

Heat: Under the *Heat* condition, we see that the average score drops to $\mu = 3.9$ corresponding to **92%** of the base condition. However, the drop is not statistically significant ($p = 0.25$) compared to the base condition. Both correctness and time are comparable to that of the base condition. In conclusion, the results are *inconclusive* and do not establish that the heat is significantly worse than the base condition; at least, not for short-term exposure under the levels tested.

Noise: The average correctness score under noise was $\mu = 3.7$ corresponding to **89%** which is an insignificant drop ($p = 0.25$) compared to the base condition. Neither the decrease in the ratio of sufficiently correct programs from 17 to 14 out of 25 nor in time amounts to significant changes ($p = 0.19$, respectively, $p = 0.41$). In summary, the results are *inconclusive* in establishing that noise is worse than the base condition. Programmers exposed to noise appear to be more resilient in writing programs (synthesis) than reading programs (analysis), which showed a more significant drop in correctness.

Alcohol: Unsurprisingly, we see significant results under the *Alcohol* condition where average correctness score drops to $\mu = 3.4$ which amounts to **82%** ($p = 0.029$). Also for sufficient correctness, the drop from 17/25 to 10/25 is significant ($p = 0.023$). The time for writing programs, however, is not significantly different from that of the base condition ($p = 0.45$). Thus, writing a program under the *Alcohol* condition is *statistically significantly worse* than under the *Base* condition.

6. Discussion

We first interpret the results and consider the qualitative interviews; then, we show an attempt to compare the treatments; i.e., *Heat* vs *Noise* vs *Alcohol* in terms of their overall negative effects on performance.

6.1. Result interpretation

Heat: We saw limited reduction in absolute performance on both the analysis and the synthesis task. During the post-experiment interview, several participants (e.g., P20, P22, and P37) described their experience in terms of primarily *physical* factors, such as sweating, but did not mention cognitive challenges. One (P55) mentioned often working under more challenging conditions: *“It went well; I work here in the summer which was warmer”* (A/C is not common in Scandinavia).

Presumably, the reason for the modest effects under the *Heat* condition is that the experimental setup represents *short-term sensory discomfort* (as in *external* skin temperature increase) rather than *long-term heat exhaustion* (as in *internal* core body-temperature overheating). For instance: *“I didn’t feel the heat at first, but I slowly got more heated up* *which made me more mentally tired. My focus went from the assignment to the heat"* (P86). A contributing factor is that the immediate negative effects of the heat could have been mitigated by positive effects of the short-term aerobic exercise prior to the task solving, which has shown to improve learning and memory functions in young adults (Blomstr and Engvall, 2021).

We speculate that *prolonged* exposure would increase the negative effect on the participants, as has been found in previous studies (Yeganeh et al., 2018). In fact, a couple of days after the experiment, P86 reached out and commented: *"Half an hour after the experiment, I was completely exhausted; I wasn't even be able to look at a screen, and I didn't do anything productive for the rest of the day"*. (We, of course, apologize for this.) It would thus be interesting to study the effects of *longer-term* heat exposure, but considering the known risks of heat stress and dehydration, this may not be ethically justifiable.

Noise: The results suggest that reading a program (analysis) is more affected than writing a programs (synthesis) under the influence of *Noise*. Recall that prior research has established that specifically *reading* and *recollection* is negatively impacted by noise from nearby human speech (Salamé and Baddeley, 1982). Many participants mentioned that the difference in performance was due to the arithmetic and mathematics involved, as well as having to keep track of and remember numbers. For instance: *"I could feel an effect on me because of the noise. [..] The first [analysis] task was the hardest with the noise. Arithmetics in your head was complicated - holding on to numbers, getting bombarded with noise"* (P77); *"It was hard to focus - hard to write down numbers"* (P29). It is possible that the synthesis task relies on more coding *fluency* (i.e. Izu and Alexander (2018)), and therefore is less affected by external stress.

Several participants offered an alternative explanation. They simply got used to the noise over time and were less affected in the second (later) synthesis task than in the first (earlier) analysis task: *"I just had to get used to the voices. In the beginning, I was very distracted, but then I found my focus. [..] It was more distracting for the first task, because I got used to the noise [for the second task]"* (P61), *"It was difficult to concentrate because of the noise. At some point, I 'cancelled out' the noise, but it was still annoying to know that it was there"* (P76).

It could be true for all conditions that the first task offered a chance to "get used to" the physical stressor, and the performance during the second task would suffer less. Particularly since we see that the percentage scores for all three conditions are generally worse for the first (analysis) than for the second (synthesis) task. Temporal adaptation was, however, only mentioned directly by participants under the noise condition, and we hypothesize that *Noise* mitigation may depend more on cognitive resilience (Staal et al., 2008), since previous research has found no physiological impact of noise (Evans and Johnson, 2000; Banbury and Berry, 2005; Jahncke et al., 2011). It is interesting, that previous research has shown that performing a cognitively challenging task can significantly impair subsequent cognitive performance (Martin et al., 2021). However, since the scores for the analysis and the synthesis task are not directly comparable, it is more interesting that the *normalized* scores for the synthesis task are generally lower than the normalized scores for the analysis task.

It is likely that extended exposure to a noisy environment would cause performance to follow a diminishing trajectory as cognitive resilience is depleted. However, it is also likely that certain cognitive tasks suffer more from physical stressors than others (Martin et al., 2019), rendering the division in programming *analysis* vs *synthesis* highly important.

Alcohol: *Alcohol* is clearly detrimental to cognitive performance, in analysis as well as synthesis tasks.[3] Many participants described problems with *focus* and *concentration*: e.g., *"Hard to focus due to the alcohol"* (P73) and *"Harder to concentrate because of the alcohol"* (P33). Curiously, we also found repeated mentions of a *loss of sense of time*: *"My sense of time was completely gone"* (P03), *"My time perception was way off"* (P39), & *"I did not even realize time had passed"* (P24).

The cognitive effects of intoxication with alcohol are well-known, as is the approximate trajectory that they follow (Fig. A.5(c)). Interestingly, for the analytical task, the participants actually had a *lower* average correctness score under *Noise* than under the *Alcohol* condition, although the difference is insignificantly only one percentage point (82% vs 83%). Our results approach knowledge of how these effects "translate" to the domain of software engineering or programming, approaching a scale with which to compare various physical factors or stressors.

6.2. Towards a comparison of treatments

We now consider how the treatments could be compared in terms of their adversarial effect by presenting a model scale of cognitive performance impact (Fig. 4) , based on the correctness scores. The comparison is intended to facilitate and spark an informed conversation about the effects of work environments to cognitive performance by "translating" the effects of physical stressors into a more tangible scale (here, the effects of alcohol intoxication). Aside from being based on a fairly small number of data points, the comparison is based on a number of assumptions; in particular, *approximation*, *comparability* and *linearity* each of which will be discussed in the following before we present the results of the approximate comparison.

Approximation: The figure presents "best estimates" based on correctness scores of an experiment involving M=25 subjects per condition. It is an approximation, and results are expected to vary based on both internal and external conditions (cf. Section 7).

Comparability: The model assumes that these treatments can be compared only in terms of their *quantitative effect* on a subject in terms of task correctness. Qualitatively, the effects are obviously highly different. The model also does not consider individual differences in cognitive resilience, but approximates an average effect overall.

Linearity: Second, for simplicity, the model assumes that effect is *linear* in the treatment dosage and thus uses *interpolation*; estimating, for instance, *half treatment* to incur *half effect*. It is obviously likely that the actual function is polynomial, or even more of a threshold function. We hope future research will reveal more information on this.

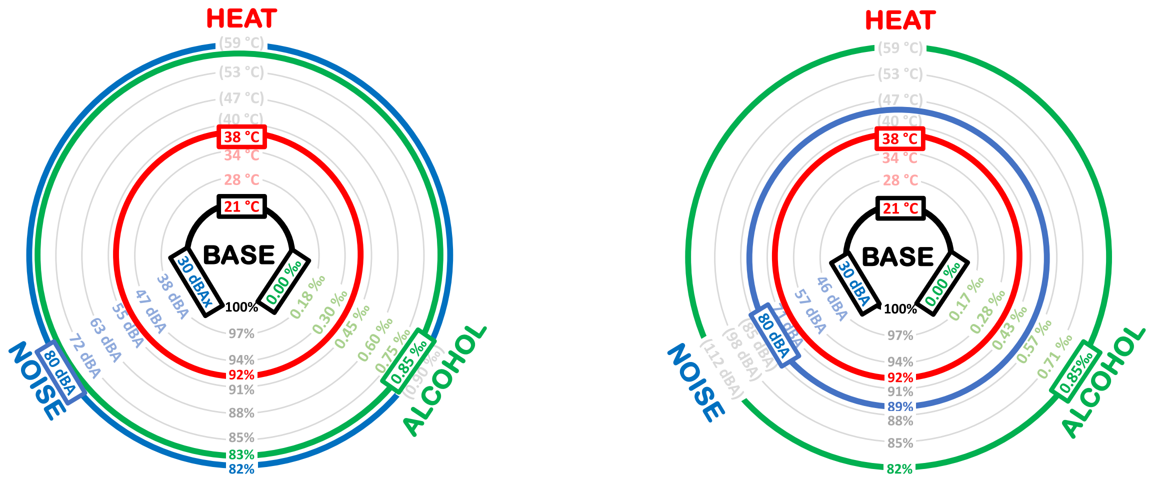
The severity and duration of exposure to a given stressor affects the degree to which performance is impaired, as does the complexity of the cognitive task and the skill or familiarity of the individual performing the task (Martin et al., 2019). The purpose of this model is to approach an *illustrative* scale of impact on cognitive performance on novice programmers in the domain of software engineering.

We infer from Fig. 4 that for the reading program task (*analysis*), the short-term effects of the environmental condition 'working in a 38°C room' could create a loss in performance equivalent to that of a BAC of 0.40‰ (or drinking around 2 alcoholic units for an average-weight male). An environmental noise level of 80 dBA could yield a loss in cognitive performance equal to a BAC of 0.90‰ (or consuming around 4 alcoholic units).

When *writing* programs, effects are smaller, but discernible. A heated room of 38°C causes a performance loss equal to that of a BAC level of 0.38‰ (or consuming around 11/2–2 alcoholic drinks for an average-weight male), and a noisy room of 80 dBA equals the performance loss of a BAC of 0.52‰ (around 21/2 alcoholic units).

Such comparisons, albeit simplified, may encourage reflection on the influence of environmental factors on the performance of workers outside the domains of physically demanding occupations. The model presented is an initial step towards formalized scales of cognitive impact; further studies in this area could help refine and expand the model.

[3] We note, that the prescribed BAC in this experiment did not reach the Ballmer Peak, thus we are unable to refute this hypothesis: https://xkcd.com/323/



(a) Reading programs (analysis).

(b) Writing programs (synthesis).

Fig. 4. Approximated comparison of programming under the influence of *Heat* vs *Noise* vs *Alcohol*. The figure is based on a number of simplifying assumptions: (i) *Approximation* – the effects are approximated based on the results of an experiment with M = 25 subjects in each condition; (ii) *Comparability* – the effects can be compared in terms of their overall performance impact; *and* (iii) *Linearity* – an increase/decrease in the exposure level will impact the effect *linearly*. (Interpolated values are written in washout colors; extrapolated values beyond the levels studied are in gray and parentheses.). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In light of recent research by Newman et al. (2023) showing many software developers occasionally perform their work under the influence of alcohol, and that their consumption of alcohol has *increased* during and after the Covid-19 pandemic, it is worth reflecting on the impact of this result. First, we note that the BAC induced in this experiment was higher than most limits for, e.g., safely operating a vehicle. It may be meaningful performing future studies with participants who have consumed less alcohol, equivalent to one or two drinks. Second, *productivity* (in the form of time to complete a task or the correctness of the solution) is only one aspect of performance enhancement. The experience of creativity is another. But a third aspect mentioned as a reason to program under the influence is simply *work enjoyment*. We are not attempting to advocate for casual intoxication testing or alcohol policies in home offices; simply that it is considering factors beyond productivity when interpreting the results of this experiment.

Also, when considering our results, it is important not to confound novice (student) programmers with professional software developers.

7. Threats to validity

First, we consider how the data was obtained and measured in the experimental setup (construct validity). Second, we investigate the threats involving the methodology and results (internal validity). Third, we examine the generalizability of our findings (external validity). Finally, we consider the threats to the conclusions derived from the study (conclusion validity).

7.1. Construct validity

Does the experimental setup sufficiently control the treatments? Failing on controlling the experiment setup would invalidate the study, therefore, the setup has been meticulously described in Section 4 and the compliance for each treatment has been verified as detailed in Section 5.1. Stronger controls could be set in place (e.g., internal thermometer for the Heat condition and blood samples for the Alcohol condition). However, the sophistication level of the instrumentation used was deemed sufficiently accurate for the measurements required for this experiment. Additionally, before-experiment pilot and during-experiment compliance verification (Section 5.1) demonstrated that the measurements obtained with the setup were as intended.

Appropriate level of treatments? The considerations for the design of the experiment were twofold: on one hand, a tradeoff between detectable effect (level should not be too low) and safety for experimentation with human participants (level should not be too high); on the other, being true to experienced working conditions. Therefore, we used: (a) *heat* of **38 °C** as a proxy for an office during the summer without air-conditioning (which is common in Scandinavia); (b) *noise* of **80 dBA** as a proxy for a discussion between co-workers in an open-office environment or noise in the home office environment; and (c) *alcohol* of **1.0‰** included for comparative and explorative purposes (see Section 6.2).

Was the task of appropriate difficulty? The tasks were designed based on the curriculum of the CS1 course, which all participants were required to have passed; were piloted on TAs which found them of appropriate difficulty; and, were able to differentiate the treatments from the base condition (i.e., average performance scores were below that of the base condition for all treatments).

Measuring heat exposure? As participants are exposed to the heat condition, the core body temperature rises only slowly due to the body's advanced mechanisms for thermal regulation (Nakamura, 2011). We attempted to control for this by having participants cycle for ten minutes prior to the experiment, which was chosen in an attempt to rise core temperature without causing physical exhaustion or adrenaline from physical exercise. A more invasive experiment, monitoring the internal core body temperature would have required sophisticated equipment (thermometer pills), medical personnel on stand-by, and a much more rigid ethical approval.

Measuring alcohol intoxication? The absorption (and elimination) of alcohol depends on prior food intake since alcohol predominantly enters the blood stream via the small intestine. Indeed, Widmark's formula assumes an empty stomach. To control for this, we asked participants not to eat anything before arriving for the experiment (aside from a small coffee), and each participant for the Alcohol condition was asked upon arrival to confirm that they had not eaten. However, we had no way of ensuring participant compliance with this stipulation.

7.2. Internal validity

Were the subjects stressed or intimidated? Since there is a risk of students being intimidated, supervising research assistants were instructed to observe, take notes, and not intervene. Participants were also informed that once the experiment had started, the supervising personnel could only monitor the room and not provide answers to questions. We employed research assistants to avoid any power relations between professors (PIs/authors) and students. For a few of the experiments, one of the authors acted as room supervision, but was never standing in physical proximity of the students and could not see how they solved the tasks.

In any experiment setup, there is a risk that simply being observed increases cognitive load, e.g., Behroozi et al. (2018, 2020). The tasks are therefore not necessarily directly comparable to realistic programming tasks performed in a professional setting. However, since all participants were part of the experiment and therefore all being observed, this effect should influence all participant groups equally. It is possible that it would increase detrimental effects for individuals with lower cognitive resilience. Further studies of higher ecological validity would have to be carried out to explore this effect.

Individual differences between subjects? The model in Fig. 4 assumes an average effect overall and does not consider individual differences in prior experience and cognitive status. People may respond differently to stressors depending on expertise, fatigue, and individual resilience. The experiment assigned participants *randomly* to one of the four conditions, hoping to mitigate effects attributable to individual differences. (After all, the Central Limit Theorem stipulates that the variance from individual differences will diminish as the number of subjects increase.) Future studies, however, could consider investigating whether certain factors, such as expertise, experience, or situational exhaustion made people more or less prone to detrimental effects of physical stressors (which is likely).

Choice of development environment? We used Java as the programming language for the tasks (reading and writing programs) since it is the language taught on the CS1 course. For the synthesis task, students used their own computer and development environment to make the task as realistic as possible and avoid performing poorly or slowly due to an unfamiliar computer or IDE.

Did the participants know what to do? To ensure that participants were not affected by different understandings of the setup and/or the tasks, the research assistant supervising each room was instructed to brief the participants via a sample of tasks that could have been expected. At this stage, participants were allowed to ask any question to clarify the experiment objective and setup. Very few students had questions about the tasks during this introduction.

Bias from circadian rhythm? Since participants obviously accepted experiment time slots depending on their availability, there could be a bias for a so-called "morning person" to get an earlier time slot versus a "night person" to get a later time slot. We expect this effect to be minimal since it should apply for all time slots (they were all based on voluntary sign-up).

Reading always preceded the writing task? The first task was always reading (analysis, T1), and the second task was always writing (synthesis, T2). For this reason, there could be "carry-over effects" from the first to the second task; either in the form of "getting used to the stressor" or "cognitive depletion from task 1 to task 2". Research has shown that performing a cognitively demanding task can significantly impair cognitive performance in subsequent tasks (Martin et al., 2021). Since the total time requirement to complete both tasks was relatively short (less than 30 min), we believe such interference to be minimal.

Sober students pre-experiment? We did not test that students showing up for the experiment (regardless of the condition to which they were randomly allocated) were, in fact, sober; i.e., with a BAC of 0.0‰. Although Danes are frequent drinkers, "drinking in class [is] not a common practice" (Ladekjær Larsen et al., 2016). Also, none of the students showed any signs of intoxication when they reported for the experiment.

7.3. External validity

Beyond students? The present work intends to reason about novice (university student) programmers reading and writing programs under adversarial conditions. It is hard to know to what extent the results generalize to professional developers. For noise, for instance, it is possible that professionals with long term exposure to a noisy work environment would develop a resilience by "getting used to" the noise. Similar reservations could apply to heat where individuals could "get used to" working in a hotter environment after longer term exposure. For alcohol, we expect *functional tolerance* will play more of a role (Tabakoff et al., 1986) for more seasoned, alcohol-accustomed developers. For all stressors considered, we hope future work will explore the generalizability of their effects on professional developers, including to what extent individuals are capable of developing "coping strategies" for dealing with the stressors.

Beyond lab setting? Obviously, the experiment was carried out in a controlled environment. However, experiments were performed in familiar surroundings for the subjects at the university—although in lecture rooms rather than group work rooms. This design was deemed optimal in terms of balancing internal/external validity (Curran and Wirth, 2004).

Beyond Denmark? The experiment was carried out in a Danish context, using Danish students. We expect the results to generalize to other populations with two reservations: (1) Danes might be less resilient to heat than someone growing up in a warmer climate; and (2) Danes might be more tolerant to alcohol than students from many other countries (cf.). In general, it thus may be that the effect of *heat* is even lower whereas the effect of *alcohol* is even higher.

Beyond artificial tasks? We deliberately tested both *reading* and *writing* programs. A more realistic programming or software development task is likely to be a combination of these two activities. To make the experiment feasible without too much time investment from the subjects, the programming tasks were fairly simplistic. Whether the results generalize to larger programs (which a worker may be more familiar with, if they have worked on the same project for extended time) is presumably a matter of tension between effects potentially increasing due to prolonged exposure (Martin et al., 2021) versus effects decreasing due to the subjects adapting to the conditions over time. The former might be applicable to the *heat* condition; the latter to that of *noise*.

7.4. Conclusion validity

Treatment comparison? The discussion surrounding Fig. 4 is based on a number of simplifying assumptions; in particular, *comparability* of the effects along with *linear* interpolation and extrapolation of the effects. Importantly, this is exclusively an issue (conclusion validity threat) for Fig. 4 and Section 6.2; the results in Section 5 are unaffected by these assumptions. Despite the validity issues, we retain Fig. 4 to spark a discussion about the negative impact of heat and noise on programmers and because the figure is the best approximation based on the data from our experiment with a limited number of students (N=4 × 25). Any *improved* future study would simply give rise to an equally *improved* comparison diagram (i.e., an *improved* version of Fig. 4).

8. Conclusion

Understanding the impact of environmental stressors is important. While physical stress has been studied to a larger extent in domains of e.g. military personnel, professional athletes, aircraft pilots, doctors, nurses, and long-haul drivers, studies of physical stressors to the cognitive performance of students and office workers is more limited. Additionally, few studies actively compare the impact of different stressors, rendering comparison of percentage metrics from different contexts somewhat meaningless.

In this article, we presented a study of $N = 100$ university student subjects under the influence of *Heat*, *Noise*, or *Alcohol* compared to a base condition. The subjects were assigned two programming tasks – one testing reading programs (*analytical* skills) and one writing programs (*synthesis skills*) – to mimic the constituents of common professional programming work. For each participant and task, we measured a *correctness score*, a derived *sufficiently correct* (boolean) indicator, and task *completion time*.

The results showed no statistically significant evidence of *Heat* (38°C, or 100°F) impacting neither the analysis nor the synthesis task. *Noise* (80 dBA) was detrimental to the analysis, but not the synthesis task (at least not to a statistically significant degree). *Alcohol* negatively impacted both the analysis as well as the synthesis task.

Under the *Noise* condition, the analysis task appears to be significantly more impacted than the synthesis task. This finding emphasized the importance of distinguishing between different types of tasks for future experiments measuring cognitive performance.

We presented an illustrative model for comparing the three influences on a scale. The model is based on a number of assumptions, most importantly of approximation based on a limited number of subjects, comparability between influences and subjects, as well as linearity of impact. We emphasize that this model is a first step towards a comparative scale, and that future studies are obviously needed to explore, modify, and refine this model. The goal of the model is to spark reflection on the relative impact of different environmental and physical stressors to the cognitive performance of programmers and software engineers. In particular, we direct to decision-makers with a say in the work environment for novice programmers (and, presumably to some degree, professional developers).

CRediT authorship contribution statement

Claus Brabrand: Conceptualization, Methodology, Formal analysis, Investigation, Visualization, Project administration, Resources, Writing – original draft, Writing – review & editing, Validation. **Nanna Inie:** Resources, Writing – original draft, Writing – review & editing, Conceptualization. **Paolo Tell:** Conceptualization, Methodology, Investigation, Resources, Data Curation, Visualization, Validation, Project administration, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We thank: Alma Freiesleben, Rebecca Sofie Brenøe, Stig Killendahl, Sebastian Mateos Nicolajsen, Amanda Bastrup, Troels Bergmann, Jesper Bernhardt, Andrew King, Sebastian Büttrich, Michael Bloch, and Dan Witzner Hansen for assistance with the experiment.

Appendix

See Figs. A.5 and A.6.

Heat (temperature)	Degree of discomfort (at 50% RH) after prolonged exposure or strenuous activity
(17°C – 23°C)	(Zone of preferred temperature)
21°C (70°F)	**Control: Comfortable temperature**
22°C – 28°C	No discomfort
30°C – 38°C	Some discomfort
38°C (100°F)	**Treatment: Uncomfortable temperature**
40°C – 45°C	Great discomfort (Warning: avoid exertion!)
47°C – 53°C	Dangerous
55°C –	Heat stroke imminent

(a) Heat (based on [94]).

Noise (decibel)	Corresponding noisy environment People	Things
30 dBA	**Quiet office**	**Leaves rustling**
40 dBA	Whisper	Refrigerator
50 dBA	Quiet speech	Washing machine
60 dBA	Conversation	Air conditioner
70 dBA	Classroom chatter	Vacuum cleaner
80 dBA	**Noisy restaurant**	**Hair dryer**
90 dBA	Shouting	Motorcycle
100 dBA	Symphony Orchestra	Powerdrill

(b) Noise (based on [95]).

Alcohol (BAC)	Intoxication effects Emotional Impact:	Physiological Impact:	Cognitive Impact:
0.0‰ – 0.1‰	**Unaffected.**	**Unaffected.**	**Unaffected.**
0.2‰ – 0.3‰	Slight relaxation, slightly altered mood.	Sense of warmth.	Slightly poorer judgments.
0.4‰ – 0.6‰	Exaggerated behavior and louder speech.	Slight loss of control of small muscles, such as the ability to focus the eyes, so vision will become blurry.	Minor impairment of reasoning and memory.
0.7‰ – 0.9‰	**Euphoria.**	**Increased loss of coordination, balance, and speech. Reaction times, and even hearing will get worse.**	**Reduced judgment and self-control. Caution, reason, and memory impaired. Over-estimated functioning.**
1.0‰ – 1.2‰	**Euphoria.**	**Reaction time and control will be reduced, speech will be slurred, and the ability to coordinate arms and legs is poor.**	**Thinking and reasoning are slower, cognitive functioning is reduced.**
1.3‰ – 1.5‰	Dysphoria begins to appear (anxiety, restlessness).	Gross motor impairment. Lack of physical control. Blurred vision. Major loss of balance. Perception severely impaired.	Judgment severely impaired.

(c) Alcohol (based on [96]).

Fig. A.5. Heat, Noise, and Alcohol at increasing intensities. (Base and experimental conditions are highlighted in bold face; conditions beyond the intensity levels tested are included in gray for context.) (see Rotronic, 2023; Chepesiuk, 2005; Anon, 2022).

TASK 1 (READ PROGRAM):

Task 1 (Read Program):

```
static void task1() {
    for (int x=2; x<9; x=x*2) {
        for (int y=4; y>1; y=y-1) {
            if (y <= x) {
                System.out.println((x-1)*(y-1))-1);
            } else {
                System.out.println(y-x);
            }
        }
    }
}
```

OUTPUT (sequence of numbers):

TASK 2 (WRITE PROGRAM):

Task 2 (Write Program):

```
Write a static method 'task2' that takes one String
argument and counts the number of words that have
exactly four characters and end with an 'e'.

        static int task2(String s) { ... }

Note that you are allowed to assume that all words end
in either ' ' (space) or '.' (dot).

If, for instance, called with the String argument
"Once upon a time.", it should return 2.

RESTRICTIONS:
You are only allowed to use the String methods:
  - char charAt(int index); // character at index #i
  - int length();           // length of the string

You are not allowed to use other String methods:
  - String[] split()
  - boolean matches(String regexp)
```

OUTPUT (PROGRAM):

Program it on your own computer in either Java or C# (your choice).

When finished, send the source code to [EMAIL] and let the assistant take a photo of the program on your screen.

Fig. A.6. The two programming tasks: *left* (analysis/read programs) & *right* (synthesis/write programs).

References

Allen, J.G., MacNaughton, P., Satish, U., Santanam, S., Vallarino, J., Spengler, J.D., 2016. Associations of cognitive function scores with carbon dioxide, ventilation, and volatile organic compound exposures in office workers: a controlled exposure study of green and conventional office environments. Environ. Health Perspect. 124 (6), 805–812.

Anon, 2022. Blood alcohol level & effects on the body. https://www.alcohol.org/effects/blood-alcohol-concentration/. (Accessed 3 May 2022).

Badre, D., Nee, D.E., 2018. Frontal cortex and the hierarchical control of behavior. Trends Cogn. Sci.

Banbury, S.P., Berry, D.C., 2005. Office noise and employee concentration: Identifying causes of disruption and potential improvements. Ergonomics 48 (1), 25–37.

Behroozi, M., Lui, A., Moore, I., Ford, D., Parnin, C., 2018. Dazed: measuring the cognitive load of solving technical interview problems at the whiteboard. In: Proceedings of the 40th International Conference on Software Engineering: New Ideas and Emerging Results. pp. 93–96.

Behroozi, M., Shirolkar, S., Barik, T., Parnin, C., 2020. Does stress impact technical interview performance? In: Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. pp. 481–492.

Benedek, M., Panzierer, L., Jauk, E., Neubauer, A.C., 2017. Creativity on tap? Effects of alcohol intoxication on creative cognition. Conscious. Cogn. 56, 128–134.

Benedek, M., Zöhrer, L., 2020. Creativity on tap 2: Investigating dose effects of alcohol on cognitive control and creative cognition. Conscious. Cogn. 83, 102972.

Blomstr, P., Engvall, J., 2021. Effects of a single exercise workout on memory and learning functions in young adults—a systematic review. Transl. Sports Med. 4 (1), 115–127.

Brabr, C., Dahl, B., 2009. Using the solo taxonomy to analyze competence progression of university science curricula. Higher Educ. 58 (10), 531–549.

Brocolini, L., Parizet, E., Chevret, P., 2016. Effect of masking noise on cognitive performance and annoyance in open plan offices. Appl. Acoust. 114, 44–55.

Budd, G.M., 2008. Wet-bulb globe temperature (wbgt)—its history and its limitations. J. Sci. Med. Sport 11 (1), 20–32, heat Stress in Sport.

Carter, A.W., Zaitchik, B.F., Gohlke, J.M., Wang, S., Richardson, M.B., 2020. Methods for estimating wet bulb globe temperature from remote and low-cost data: A comparative study in central alabama. Geohealth 4 (5), e2019GH000231.

Chepesiuk, R., 2005. Decibel hell: the effects of living in a noisy world.

Cian, C., Barraud, P., Melin, B., Raphel, C., 2001. Effects of fluid ingestion on cognitive function after heat stress or exercise-induced dehydration. Int. J. Psychophysiol.

Clausen, T., Kristiansen, J., Hansen, J.V., Pejtersen, J.H., Burr, H., 2013. Exposure to disturbing noise and risk of long-term sickness absence among office workers: A prospective analysis of register-based outcomes. Int. Arch. Occup. Environ. Health 86 (7), 729–734.

Curran, P.J., Wirth, R., 2004. Interindividual differences in intraindividual variation: Balancing internal and external validity. Measurement: Interdiscip. Res. Perspect.

Dry, M.J., Burns, N.R., Nettelbeck, T., Farquharson, A.L., White, J.M., 2012. Dose-related effects of alcohol on cognitive functioning. PLoS One 7 (11), e50977.

Du, B., Tandoc, M.C., Mack, M.L., Siegel, J.A., 2020. Indoor co2 concentrations and cognitive function: a critical review. Indoor Air 30 (6), 1067–1082.

Endres, M., Boehnke, K., Weimer, W., 2022. Hashing it out: a survey of programmers' cannabis usage, perception, and motivation. In: Proceedings of the 44th International Conference on Software Engineering. pp. 1107–1119.

Evans, G.W., Johnson, D., 2000. Stress and open-office noise. J. Appl. Psychol. 85 (5), 779.

Ford, D., Storey, M.-A., Zimmermann, T., Bird, C., Jaffe, S., Maddila, C., Butler, J.L., Houck, B., Nagappan, N., 2021. A tale of two cities: Software developers working from home during the covid-19 pandemic. ACM Trans. Softw. Eng. Methodol. (TOSEM) 31 (2), 1–37.

Friel, P.N., Baer, J.S., Logan, B.K., 1995. Variability of ethanol absorption and breath concentrations during a large-scale alcohol administration study. Alcohol. Clin. Exp. Res. 19, 1055–1060.

Fucci, D., Scanniello, G., Romano, S., Juristo, N., 2020. Need for sleep: The impact of a night of sleep deprivation on novice developers' performance. IEEE Trans. Softw. Eng.

Gaoua, N., Grantham, J., Racinais, S., El Massioui, F., 2012. Sensory displeasure reduces complex cognitive performance in the heat. J. Environ. Psychol. 32 (2), 158–163.

Ghory, I., 2007. Using fizzbuzz to find developers who grok coding. https://imranontech.com/2007/01/24/using-fizzbuzz-to-find-developers-who-grok-coding/. (Accessed 13 January 2023).

Giancola, P.R., Zeichner, A., Yarnell, J.E., Dickson, K.E., 1996. Relation between executive cognitive functioning and the adverse consequences of alcohol use in social drinkers. Alcohol. Clin. Exp. Res. 20 (6), 1094–1098.

Grandjean, A.C., Grandjean, N.R., 2007. Dehydration and cognitive performance. J. Am. Coll. Nutr.

Graziotin, D., Fagerholm, F., Wang, X., Abrahamsson, P., 2017. Consequences of unhappiness while developing software. In: 2017 IEEE/ACM 2nd Int. Workshop on Emotion Awareness in Software Engineering. SEmotion, IEEE.

Graziotin, D., Fagerholm, F., Wang, X., Abrahamsson, P., 2018. What happens when software developers are (un) happy. J. Syst. Softw. 140, 32–47.

Hancock, P.A., Vasmatzidis, I., 2003. Effects of heat stress on cognitive performance: the current state of knowledge. Int. J. Hyperth. 19 (3), 355–372.

Hansen, S.E.R., Lundgaard, P.B., Christensen, A.S.P., Hansen, S., Lundgaard, P., Christensen, A., Bekæmpelse, K., 2022. Unges alkoholvaner i danmark 2021.

Heng, Y.T., Barnes, C.M., Yam, K.C., 2022. Cannabis use does not increase actual creativity but biases evaluations of creativity. J. Appl. Psychol.

Hoffman, L.A., Nixon, S.J., 2015. Alcohol doesn't always compromise cognitive function: Exploring moderate doses in young adults. J. Stud. Alcohol Drugs.

Izu, C., Alexander, B., 2018. Using unstructured practice plus reflection to develop programming/problem-solving fluency. In: Proceedings of the 20th Australasian Computing Education Conf.
Jahncke, H., 2012. Cognitive Performance and Restoration in Open-Plan Office Noise (Ph.D. dissertation). Luleå tekniska universitet.
Jahncke, H., Hallman, D.M., 2020. Objective measures of cognitive performance in activity based workplaces and traditional office types. J. Environ. Psychol. 72, 101503.
Jahncke, H., Hygge, S., Halin, N., Green, A.M., Dimberg, K., 2011. Open-plan office noise: Cognitive performance and restoration. J. Environ. Psychol. 31 (4), 373–382.
Jamrozik, A., Clements, N., Hasan, S.S., Zhao, J., Zhang, R., Campanella, C., Loftness, V., Porter, P., Ly, S., Wang others, S., 2019. Access to daylight and view in an office improves cognitive performance and satisfaction and reduces eyestrain: A controlled crossover study. Build. Environ. 165, 106379.
Jones, D., Hughes, R., Marsh, J., Macken, W., 2008. Varieties of auditory distraction.
Jr., M.C.M., Teigen, E.L., Ramchandani, V.A., 2014. Absorption and peak blood alcohol concentration after drinking beer, wine, or spirits. Alcohol. Clin. Exp. Res. 38.
Juristo, N., Moreno, A.M., 2013. Basics of Software Engineering Experimentation. Springer Science & Business Media.
Kimmel, H.D., 1957. Three criteria for the use of one-tailed tests. Psychol. Bull. 54, 351–353.
Kitchenham, B.A., Pfleeger, S.L., Pickard, L.M., Jones, P.W., Hoaglin, D.C., El Emam, K., Rosenberg, J., 2002. Preliminary guidelines for empirical research in software engineering. IEEE Trans. Softw. Eng. 28 (8).
Kjellstrom, T., Holmer, I., Lemke, B., 2009. Workplace heat stress, health and productivity–an increasing challenge for low and middle-income countries during climate change. Glob. Health Action 2 (1), 2047.
Kowal, M.A., Hazekamp, A., Colzato, L.S., van Steenbergen, H., van der Wee, N.J., Durieux, J., Manai, M., Hommel, B., 2015. Cannabis and creativity: highly potent cannabis impairs divergent thinking in regular cannabis users. Psychopharmacology 232, 1123–1134.
Kristiansen, N.G., Nicolajsen, S.M., Brabrand, C., 2023. Feedback on student programming assignments: Teaching assistants vs automated assessment tool. In: Koli Calling 2023. (in press).
Ladekjær Larsen, E., Smorawski, G., Lund Kragbak, K., Stock, C., 2016. Students' drinking behavior and perceptions towards introducing alcohol policies on university campus in Denmark: a focus group study. Subst. Abuse Treat. Prev. Policy 11 (17).
Lapp, W.M., Collins, R.L., Izzo, C.V., 1994. On the enhancement of creativity by alcohol: Pharmacology or expectation? Am. J. Psychol. 173–206.
Lieberman, H.R., Tharion, W.J., Shukitt-Hale, B., Speckman, K.L., Tulley, R., 2002. Ffects of caffeine, sleep loss and stress on cognitive performance and mood during us navy seal training. Psychopharmacology 164 (3), 250–261.
Liebl, A., Haller, J., Jödicke, B., Baumgartner, H., Schlittmeier, S., Hellbrück, J., 2012. Combined effects of acoustic and visual distraction on cognitive performance and well-being. Appl. Ergon. 43 (2), 424–434.
Liu, M., Balamurugan, S., Seetharam, T.G., 2021. Impact of stress on software developers by moderating the relationship through emotional intelligence in a work environment. Aggress. Violent Behav..
Ljungberg, J.K., Neely, G., 2007. Stress, subjective experience and cognitive performance during exposure to noise and vibration. J. Environ. Psychol. 27 (1), 44–54.
Mackie, M.-A., Van Dam, N.T., Fan, J., 2013. Cognitive control and attentional functions. Brain Cogn.
Mak, C.M., Lui, Y., 2012. The effect of sound on office productivity. Build. Serv. Eng. Res. Technol.
Martin, K., Flood, A., Pyne, D.B., Périard, J.D., Keegan, R., Rattray, B., 2021. The impact of cognitive, physical, and psychological stressors on subsequent cognitive performance. Hum. Factors.
Martin, K., McLeod, E., Périard, J., Rattray, B., Keegan, R., Pyne, D.B., 2019. The impact of environmental stress on cognitive performance: a systematic review. Hum. Factors.
Martin, K., Périard, J., Rattray, B., Pyne, D.B., 2020. Physiological factors which influence cognitive performance in military personnel. Hum. Factors 62 (1), 93–123.
Maughan, R., Shirreffs, S., 2004. Exercise in the heat: challenges and opportunities. J. Sports Sci.
Maxwell, L.E., 2015. Noise in the office workplace. Facil. Plan. Manag. Notes 1 (11).
Meng, Q., An, Y., Yang, D., 2021. Effects of acoustic environment on design work performance based on multitask visual cognitive performance in office space. Build. Environ.
Mitchell, T., 2023. COVID-19 Pandemic Continues to Reshape Work in America — pewresearch.org. https://www.pewresearch.org/social-trends/2022/02/16/covid-19-pandemic-continues-to-reshape-work-in-america/. (Accessed 26 July 2023).
Müller, S.C., Fritz, T., 2016. Using (bio) metrics to predict code quality online. In: 2016 IEEE/ACM 38th International Conference on Software Engineering. ICSE, IEEE, pp. 452–463.
Nakamura, K., 2011. Central circuitries for body temperature regulation and fever. Am. J. Physiol.-Regul., Integr. Comp. Physiol. 301 (5), 1207–1228.
Newman, K., Endres, M., Weimer, W., Johnson, B., 2023. From organizations to individuals: Psychoactive substance use by professional programmers. In: 2023 IEEE/ACM 45th International Conference on Software Engineering. ICSE, IEEE, pp. 665–677.
Norman, D.A., Shallice, T., 1986. Attention to Action, in Consciousness and Self-Regulation. Springer, pp. 1–18.
OECD, 2019. E. O. on Health Systems, and Policies. In: Denmark: Country Health Profile 2019. [Online]. Available: https://www.oecd-ilibrary.org/content/publication/5eede1c6-en.
Parliament, E., 2003. On the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (noise).
Penzenstadler, B., Torkar, R., Montes, C.M., 2021. Take a deep breath. Benefits of neuroplasticity practices for software developers and computer workers in a family of experiments. ArXiv preprint arXiv:2109.07285.
Périard, J.D., Eijsvogels, T.M., Daanen, H.A., 2021. Exercise under heat stress: Thermoregulation, hydration, performance implications, and mitigation strategies. Physiol. Rev.
Peterson, J.B., Rothfleisch, J., Zelazo, P.D., Pihl, R.O., 1990. Acute alcohol intoxication and cognitive functioning. J. Stud. Alcohol 51 (2), 114–122.
Plata, A., Motoki, K., Spence, C., Velasco, C., 2022. Trends in alcohol consumption in relation to the covid-19 pandemic: A cross-country analysis. Int. J. Gastron. Food Sci. 27, 100397.
Ramsey, J.D., Burford, C.L., Beshir, M.Y., Jensen, R.C., 1983. Effects of workplace thermal conditions on safe work behavior. J. Saf. Res. 14 (3), 105–114.
Romano, S., Scanniello, G., Fucci, D., Juristo, N., Turhan, B., 2018. The effect of noise on software engineers' performance. In: Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. ESEM '18, ACM.
Rotronic, 2023. Humidex monitoring. https://www.rotronic.com/en-ca/humidity_measurement-feuchtemessung-mesure_de_l_humidite/humidex-humidity-index-mr. (Accessed 20 December 2023).
Salamé, P., Baddeley, A., 1982. Disruption of short-term memory by unattended speech: Implications for the structure of working memory. J. Verb. Learn. Verb. Beh.
Schlittmeier, S., 2021. Review of research on the effects of noise on cognitive performance 2017–2021.
Schlittmeier, S., Hellbrück, J., Thaden, R., Vorländer, M., 2008. The impact of background speech varying in intelligibility: Effects on cognitive performance and perceived disturbance. Ergonomics 51 (5), 719–736.
Selby, C.C., 2015. Relationships: computational thinking, pedagogy of programming, and bloom's taxonomy. In: Proceedings of the Workshop in Primary and Secondary Computing Education.
Shen, J., Zhang, X., Lian, Z., 2020. Impact of wooden versus nonwooden interior designs on office workers' cognitive performance. Percept. Mot. Skills 127 (1), 36–51.
Sorkin, R.D., 1988. Am. J. Psychol. 101 (2), 290–293.
Staal, M.A., Bolton, A.E., Yaroush, R.A., Bourne Jr., L., 2008. Cognitive performance and resilience to stress. In: Biobehavioral Resilience to Stress, vol. 19.
Sundstrom, E., Town, J.P., Rice, R.W., Osborn, D.P., Brill, M., 1994. Office noise, satisfaction, and performance. Environ. Behav. 26 (2), 195–222.
Szalma, J.L., Hancock, P.A., 2011. Noise effects on human performance: a meta-analytic synthesis. Psychol. Bull..
T. D. M. of Justice, 2007. Parliamentary Report on Alcohol in Exhaled Air and a Zero-Tolerance for Euforic Drugs etc. Copenhagen.
Tabakoff, B., Cornell, N., Hoffman, P.L., 1986. Alcohol tolerance. Ann. Emerg. Med. 15 (9), 1005–1012.
Tang, S.K., Wong, C., 1998. Performance of noise indices in office environment dominated by noise from human speech. Appl. Acoust. 55 (4), 293–305.
Taylor, L., Watkins, S.L., Marshall, H., Dascombe, B.J., Foster, J., 2016. The impact of different environmental conditions on cognitive function: a focused review. Front. Physiol.
Umemoto, A., Inzlicht, M., Holroyd, C.B., 2019. Electrophysiological indices of anterior cingulate cortex function reveal changing levels of cognitive effort and reward valuation that sustain task performance. Neuropsychologia 123, 67–76.
Watson, P.E., Watson, I.D., Batt, R.D., 1981. Prediction of blood alcohol concentrations in human subjects, updating the widmark equation. J. Stud. Alcohol 42 (7), 547–556, pMID: 7289599. [Online]. Available: https://doi.org/10.15288/jsa.1981.42.547.
Welleschik, B., 1979. The effect of the noise level on the occupational hearing loss. observations carried cut in 25, 544 industrial workers (author's transl). Laryngol. Rhinol. Otol.
Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A., 2012. Experimentation in Software Engineering. Springer Science & Business Media.
Yeganeh, A.J., Reichard, G., McCoy, A.P., Bulbul, T., Jazizadeh, F., 2018. Correlation of ambient air temperature and cognitive performance: A systematic review and meta-analysis. Build. Environ. 143, 701–716.
Yin, J., Zhu, S., MacNaughton, P., Allen, J.G., Spengler, J.D., 2018. Physiological and cognitive performance of exposure to biophilic indoor environment. Build. Environ.
Zhang, F., de Dear, R., Hancock, P., 2019. Effects of moderate thermal environments on cognitive performance: A multidisciplinary review. Appl. Energy 236, 760–777.

Claus Brabrand Ph.D. in Computer Science from the BRICS International Research Center, Aarhus University (2003). Writer, Director, and co-Producer of the award-winning educational short-film "Teaching Teaching & Understanding Understanding" (2006). Full Professor at the IT University of Copenhagen. Worked on and of with Educational Development and Teaching Teachers to Teach (since 2007). Worked on Programming Languages, Program Analysis, and Software Product Lines (until 2019)). Now working exclusively within Computing Education Research. Head of the Center for Computing Education Research (CCER), IT University of Copenhagen. Recipient of the Danish National Teaching Award 2020 (awarded to 2 : 18,000 university teachers in all of Denmark).

Nanna Inie holds a PhD in Digital Design from Aarhus University and is appointed postdoctoral VILLUM Fellow at the IT University of Copenhagen. Her research focuses on interaction with generative AI, red teaming, ideation with digital tools, and in how people learn with computers. She has won over $1.5M in external funding from sources in the USA and Denmark. Dr. Inie has previously also held positions at UC San Diego and University of Washington. She is a strong advocate of equity in academia, particularly focusing on women and neurodivergence. nans@itu.dk.

Paolo Tell is an Associate Professor of Software Engineering in the Computer Science Department at the IT University of Copenhagen, 2300 Copenhagen, Denmark. His research and teaching are centered on industry-academia collaborations. His research interests include software processes and software process improvement, teamwork, computer supported cooperative work, and (globally) distributed software engineering. Paolo Tell received a Ph.D. in software engineering from the IT University of Copenhagen. Contact him at pate@itu.dk.