electronics

Article

Evaluating a Camera-Based Approach to Assess Cognitive Load During Manufacturing Computer Tasks

Nicola Vasta [1], Noor Jajo [1], Frida Graf [2], Legolas Zhang [1] and Francesco N. Biondi [1,3,*]

[1] Human Systems Lab, Department of Kinesiology, University of Windsor, Windsor, ON N9B 2Z5, Canada; nvasta@uwindsor.ca (N.V.); jajon@uwindsor.ca (N.J.); legolas.zhang733@student.publicboard.ca (L.Z.)
[2] Atlas Copco Inc., 13154 Stockholm, Sweden; frida.graf@atlascopco.com
[3] Applied Cognition Lab, Department of Psychology, University of Utah, Salt Lake City, UT 84112, USA
* Correspondence: francesco.biondi@uwindsor.ca

Abstract: Suboptimal levels of cognitive load have been shown to lead to distractions, stress, and physical injuries in work environments. Yet, traditional methods for measuring cognitive load present known logistical and methodological issues: while self-reported measures suffer from poor construct validity, physiological measures often require expensive instruments and time-consuming calibration. In recent years, research has linked blink rate (i.e., the number of eye blinks per minute) with cognitive load, showing a higher blink rate with increased load. Despite this, scientific-grade eye trackers are usually expensive and invasive, making them unsuitable for work environments. In this study, we aimed to evaluate the accuracy of a camera-based approach to measure blink rate using a widely available generic webcam. To test this, we employed two tasks that resemble computer tasks common in office and manufacturing settings. Our results showed that the camera-based approach measured cognitive load as accurately as a scientific-grade eye tracker. These findings are crucial as they provide an affordable alternative to expensive and invasive instruments for measuring cognitive load in the workplace.

Keywords: eye blink; cognitive load; computer tasks; office; manufacturing

check for updates

Academic Editor: Luca Mesin

Received: 11 December 2024
Revised: 14 January 2025
Accepted: 17 January 2025
Published: 24 January 2025

Citation: Vasta, N.; Jajo, N.; Graf, F.; Zhang, L.; Biondi, F.N. Evaluating a Camera-Based Approach to Assess Cognitive Load During Manufacturing Computer Tasks. *Electronics* **2025**, *14*, 467. https://doi.org/10.3390/electronics14030467

1. Introduction

Cognitive load refers to the mental resources required to perform a task, with easy tasks, such as assembling a basic object, requiring fewer mental resources, and harder tasks, such as assembling a complex object, demanding a higher amount of mental resources [1,2]. Accurately assessing cognitive load is crucial for ensuring the operator's optimal performance in the workplace. In manufacturing settings, both low and high cognitive loads can impair performance, as errors may arise from drowsiness in cases of low cognitive load or from mental distress when the cognitive load is high [3,4]. Recent research indicated that higher levels of cognitive load are associated with a 30% increase in workplace injuries [5], highlighting the need to develop efficient tools for assessing cognitive load. In fact, measuring cognitive load could help prevent conditions such as burnout [6,7] and physical injuries due to both excessive and insufficient loads [5]. Consequently, in the past years, research has increasingly focused on developing simple and effective methods to monitor cognitive load, especially in environments like manufacturing, where prolonged mental and physical demands are common.

Measuring cognitive load, however, comes with logistical challenges, especially when performing real-world tasks. Many current assessment tools rely on subjective ratings, requiring workers to self-assess their level of cognitive load. The NASA Task Load Index

(NASA-TLX) questionnaire [8], for example, is widely used due to its quick and easy applicability, and it is often considered the gold standard for measuring changes in cognitive load in workplace environments. Nevertheless, several researchers have argued that there are no clear reasons to prefer this tool over others, and they suggested incorporating additional measures when assessing cognitive load [9,10]. Moreover, recent studies have shown that subjective measures of cognitive load often diverge from other indicators, such as physiological and behavioral measures [11]. This discrepancy suggests that self-reported measures may either assess a different construct or lack sufficient validity for testing cognitive load [11,12], indicating that more objective measures, such as physiological ones, should be preferable.

Eye metrics, for instance, have been found to accurately measure dynamic changes in cognitive load in workplace environments. Ref. [4] observed an increased blink rate (measured in blinks per minute) when the mental demand of a manufacturing task simulation increased in a laboratory setting by adding a mental counting task. Similarly, Ref. [13] asked participants to perform an auditory task requiring them to pay attention to a sequence of tones, finding a higher blink rate during a demanding phase compared to a resting phase. Blink rate is usually measured via equipment that requires an initial time-consuming mounting and calibration and laborious data processing afterward. Two of the most used methods to measure blink rate are electrooculography (EOG) and optical scientific-grade eye trackers. EOG requires the use of contact sensors placed on the user's eye muscles to record changes in voltage. Optical eye trackers use infrared light emitted by infrared beams, which reflects off the pupil to estimate eye closure, and they require an initial time-consuming calibration. While these methods are quite accurate in detecting eye blinks, they require expensive and labor-intensive equipment, making them unsuitable for everyday workplace settings, where time and resources are limited.

To address this issue, recent research has explored novel non-invasive methods for recording eye blinks using widely available camera technology. For example, Ref. [14] estimated eye aperture by extracting facial landmarks from generic video footage and then compared this information to a set threshold to detect eye blinks. Similarly, Ref. [15] applied a machine learning algorithm called Dlib-ml [16] to video footage to estimate eye aperture based on six facial landmarks. A threshold, known as the eye aspect ratio, was calculated, with a blink being detected whenever the eye aperture fell below this threshold. Consistent with previous research, Ref. [15] observed changes in blink rate in conditions of cognitive underload during driving (for similar results, see also [17,18]). However, it is important to note that the studies above failed to validate these approaches when compared to more established methods for measuring eye blinks.

Though useful, the studies above were conducted on an extremely limited participant sample, wherein only a handful of brief video recordings underwent processing. For this reason, Ref. [19] reasoned that such studies lacked the use of controlled and reliable metrics for validation and attempted to validate the same method presented by [15] in a controlled laboratory context. In their study, participants were instructed to perform a mental task wherein they had to memorize and repeat strings of numbers with increasing lengths. Eye blinks were recorded via both a generic webcam and a scientific-grade eye tracker. Their results showed that blink rate increased as a function of task difficulty and that the camera-based approach was as effective as the eye tracker in detecting these changes.

The Present Study

Building on prior work, the current study aims to adopt a similar procedure to that used by [19] but with participants completing more realistic tasks. While the previous study adopted experimental tasks whose use is common in applied cognition research, here, we wish to extend the validity of the camera-based approach to measuring changes in cognitive load during the completion of computer tasks common in manufacturing and office settings. This is key to establishing the effectiveness and reliability of the proposed methodology within real working environments.

There are two objectives investigated in this study:

Objective 1 aims to assess the accuracy of the camera-based system in detecting eye blinks during realistic computer tasks. If this approach is accurate, we expect that the blink rate measured by the camera-based approach will not differ from those measured by a scientific-grade eye tracker.

Objective 2 aims to determine whether the camera-based system can accurately detect differences in cognitive load between two computer tasks of varying difficulty levels. Given the relationship between blink rate and cognitive load, we expect a higher blink rate during trials with greater cognitive demand. Task cognitive load was also tested using self-report measures for comparison.

To achieve these objectives, participants performed two cognitive tasks on a desktop computer (block-building and point-and-click tasks) with increasing levels of difficulty. The block-building task required participants to assemble blocks to match a target array, while the point-and-click task involved pointing and clicking on targets following a predetermined order. These two tasks were selected because they closely resemble typical desktop and office activities being performed in diverse workplace environments (for a study using similar tasks, see [20]). For instance, in manufacturing settings, they can simulate stationary activities like inspecting products for quality control, assembling electronic components, or managing inventory. The generic video feed from a ubiquitous camera was processed using the same threshold-based approach used in [19] to estimate eye blinks. Its output was validated using the blink rate from a scientific-grade eye tracker, which was manually inspected by a research assistant. Cognitive load was measured through both self-reported measures and blink rate during the two computer tasks.

2. Method

2.1. Participants

Twenty-five volunteers (12 men, 13 women) were recruited from the University of Windsor student population and received a CAD 10 Amazon gift card in exchange for their participation. Their age ranged between 18 and 30 years old (M = 22, SD = 3.3). They all had normal or corrected-to-normal hearing and sight. The research was approved by the University of Windsor Research Ethics Board (#19–045).

2.2. Design

A 2 (method: camera-based detection vs. eye tracker detection) by 2 (type of task: block-building vs. point-and-click) by 2 (difficulty: easy vs. hard) within-subject design was used in this experiment. Eye blinks were recorded throughout the entire experiment using both a camera-based system and a scientific-grade eye tracker. Blinks were detected using the dlib algorithm with the camera-based system and were visually inspected by an experimenter when using the eye tracker (see Section 2.5 for details). Participants completed two computer tasks: one task involved replicating a series of arrays displayed on the screen (i.e., the block-building task), while the other involved clicking on numbers in sequential order according to the displayed templates (i.e., the point-and-click task). Each task had two levels of difficulty, which varied by the number of colors in the array for the block-building task and by the length of the sequence in the point-and-click task (see below for further details).

To test our hypotheses, two dependent variables were used: the blink rate (in blinks per minute), produced by either the camera-based system or manually produced using the eye tracker; the self-reported cognitive load rating measured via the NASA Task Load Index (TLX) questionnaire [8].

2.3. Equipment and Materials

2.3.1. Camera

For this study we used a generic webcam, available at office supply stores and online. The NexiGo N660P (NexiGo Inc., Seattle, WA, USA) has a resolution of 1080 pixels and a sampling rate of 30 frames per second. Importantly, this sampling rate is sufficient to detect any eye blink, as blinks typically last no less than 100 milliseconds (cf., [21]). The webcam was positioned atop a 27-inch Acer ET322QR (Acer Inc., Taipei, Taiwan) monitor with a resolution of 1920 × 1080, which was connected to a PC running Windows 11. Participants were seated in an office chair approximately 50 cm from the screen, with the webcam directed at their faces. Video footage recorded during the experiment was later processed for eye blink detection (see Section 2.5).

2.3.2. Eye Tracker

A remote Gazepoint GP3 eye tracker (Gazepoint Inc., Vancouver, BC, Canada) with a data collection frequency of 60 Hz was used. Previous research has demonstrated that this device is a reliable tool for desktop-based eye tracking (e.g., [19,22]). The eye tracker features a user-friendly graphical interface, which was utilized for the calibration process.

2.3.3. Computer Tasks

Participants were required to complete two distinct computer tasks, each consisting of two easy trials and two hard trials.

Block-building task: Participants were instructed to replicate arrays displayed on their screens using the TinkerCad block-building website. The screen was arranged in split-view, with the template shown at the bottom and the TinkerCad interface at the top. The arrays were screenshots of a 7 × 7 plate, displaying either combinations of two colors for easy trials or six colors for hard trials (see Figure 1). Participants had to meticulously recreate the arrays' color, shape, and design by dragging colored blocks with the mouse. During instructions, they received training on how to manipulate the blocks, including placement, deletion, movement, and color changes.

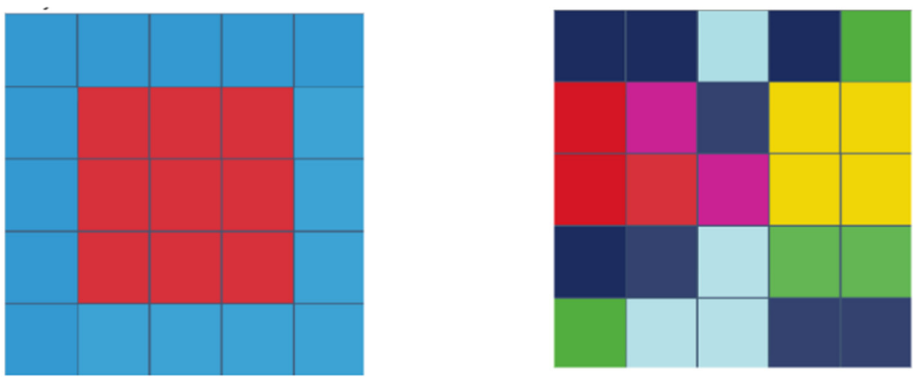

Figure 1. Participants were instructed to recreate the target arrays in the easy (**left**) and hard (**right**) conditions.

Point-and-click Task: This task required participants to click on numbers in sequential order as per the displayed template. Participants viewed two split-screen images: one with empty circles and another with numbered circles arranged vertically to minimize head movement. The easy condition featured 12 numbers, while the hard condition included 21 numbers (see Figures 2 and 3). An error (clicking the wrong number) necessitated starting over from number one. Instructions on task mechanics and consequences of errors were provided before the trials began.

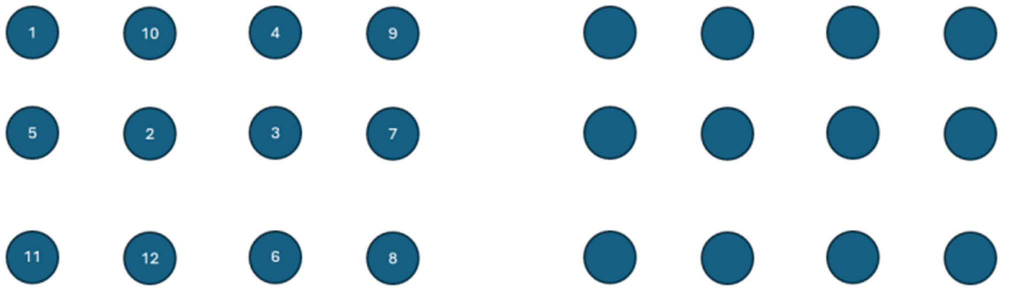

Figure 2. Easy condition of the point-and-click task. Participants had to click on the empty circles on the right following the order presented on the left.

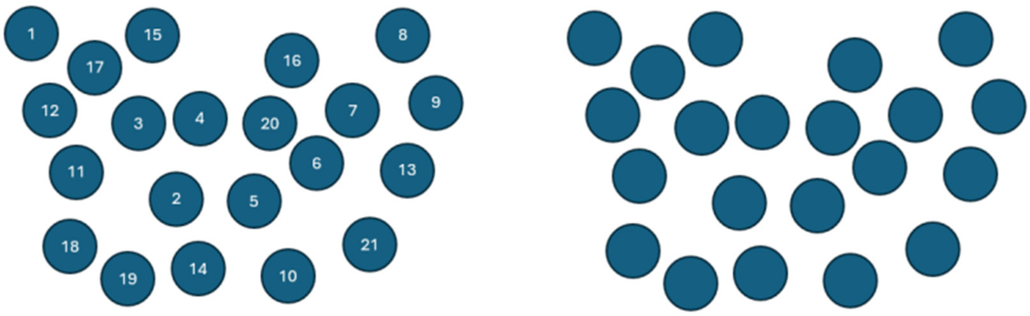

Figure 3. Hard condition of the point-and-click task. Participants had to click on the empty circles on the right following the order presented on the left.

2.3.4. Self-Reported Cognitive Load

The NASA-TLX questionnaire [8] was used to assess self-reported load. Participants were asked to rate on a scale from 1 (very low) to 21 (very high) the level of cognitive load experience during the experimental condition. The questionnaire had six items involving mental demand, physical demand, temporal demand, performance, effort, and frustration.

2.4. Procedure

Before starting the experiment, participants were asked to complete a consent form and provide demographic information. After receiving a general overview of the study, they were instructed on how to perform the calibration process of the eye tracker: participants were required to fixate on a red circle that moved to occupy nine distinct positions on the screen. Next, participants received instructions on how to complete the experimental phase (see Figure 4 for an example of the experimental setup).

During the experimental phase, participants completed a total of four trials for each task: two of easy difficulty and two of hard difficulty. Each trial lasted 2–3 min in the block-building task and 20–120 s in the point-and-click task. Video footage of the participant's face was recorded throughout the experimental phase via a generic webcam and later analyzed for eye blink detection (see Section 2.5). The number of eye blinks was also recorded using the Gazepoint eye tracker, which was later visually inspected by a research assistant to ensure exact accuracy. The presentation of pairs of trials with the same difficulty and the order of the tasks was counterbalanced using a Latin square design. At the end of each pair (i.e., either easy or hard), participants provided their subjective cognitive load ratings via the NASA-TLX questionnaire. The next trial began when participants indicated they were ready.

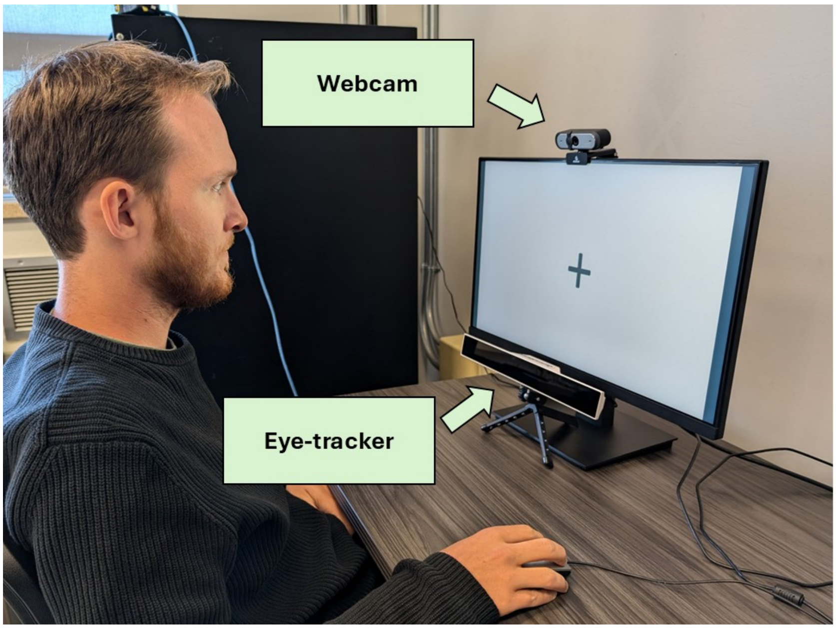

Figure 4. Example of experimental setup.

2.5. Data Processing

2.5.1. Generic Camera-Based Eye Blink Calculation

The generic camera-based eye blink calculation was the same as [19]. First, a Histogram of Oriented Gradients (HOG) based face detection algorithm was applied to a particular frame. The result of this algorithm is the coordinates of a rectangle encapsulating the human face in the frame. From this region of interest covering the human face, shape prediction algorithms that are capable of localizing key points were applied. After the coordinates of the rectangle were localized, the dlib facial landmark detector was applied. This detector algorithm provides 68 landmarks (X and Y coordinates) distinguishing different features of the participants' faces. Of these landmarks, the ones of interest are coordinates corresponding to the right eye and to the left eye. These eye landmark positions on the frame are used to calculate the scalar value called the eye aspect ratio (EAR). The formula below shows how EAR was calculated for the left eye; a similar formula was used for the right eye.

$$\text{EAR}_{left\ eye} = \frac{\parallel p_{38} - p_{42} \parallel + \parallel p_{39} - p_{41} \parallel}{2 \parallel p_{37} - p_{40} \parallel}$$

where $p38$ and $p39$ are the landmarks corresponding to the left eye's top eyelid, $p42$ and $p41$ are the landmarks corresponding to the left eye's bottom eyelid, $p37$ is the landmark corresponding to the left eye's medial canthus, and $p40$ is the landmark corresponding to the left eye's lateral cantus.

EAR was calculated individually for each eye and then averaged across the two eyes. Recorded EAR values were then compared against a threshold to determine the occurrence of eye blinks. Following [19], we adopted an adaptive threshold. The average EAR was calculated for a video that was recorded at 30 Hz. Due to this high sampling, the resulting EAR was very noisy. Hence, a moving average filter was employed to remove noise. The number of frames chosen for the data smoothing process was contingent upon an initial visual assessment of the data, where the noise level and the video duration were taken into consideration. Consequently, longer videos utilized larger frame windows, while shorter videos employed smaller ones, with frame ranges spanning from 5 to 30 frames for smoothing. Each video was then split into 5 s time windows. Considering that a normal blink rate is approximately 12 blinks per minute, the duration of each 5 s window was determined so at least one blink would fall within each window. The mean and standard deviation (SD) of EAR were calculated for each window. Within each window, a threshold of 2 SD below the mean was set (as in [19]). The duration of individual eye blinks approximates 250 milliseconds. With this in mind, considering the webcam's sampling rate of 30 Hz, a threshold of 3 consecutive frames was set for blink detection. This means that, to be classified as a blink, the recorded EAR must fall below the adaptive threshold (means −2 SD) for at least 3 consecutive frames; otherwise, it would be classified as a non-blink.

2.5.2. Eye Tracker Eye-Blink Calculation

The Gazepoint eye tracker was initially used to record the number of eye blinks for each participant in each condition. Additionally, a research assistant visually inspected the data to ensure they were free of artifacts and manually coded any missing blinks. The output from the Gazepoint eye tracker was then compared with the results from the generic camera-based eye blink detection.

2.6. Analytical Approach

In this study, our primary objective was to test the absence of a difference in eye blink measurements obtained via two different methods (see Objective 1). When testing for the absence of an effect and measuring evidence for the null hypothesis, traditional null hypothesis significance testing (NHST) is insufficient. In NHST, the *p*-value calculated by a statistical test measures the discrepancy from a null sampling distribution, with *p*-values smaller than a significance criterion α (usually set to 0.05) leading to the rejection of the null hypothesis [23,24]. However, NHST does not account for evidence supporting the null hypothesis, as *p*-values only describe the likelihood of committing a Type I error when rejecting the null hypothesis (i.e., the likelihood of finding a false positive). Thus, NHST is unsuitable for testing Objective 1.

Instead, Bayes factor analysis can directly quantify whether the data increase or decrease the likelihood that the null hypothesis is true [25]. Drawing from the Bayesian approach, it is possible to set up two competing models, one in favor of the null hypothesis and the other in favor of the alternative hypothesis, and estimate the conditional probability of each model in explaining the observed data [26]. The Bayes factor (BF) is calculated as the ratio of the posterior odds of the null hypothesis to those of the alternative hypothesis, with a BF > 1 indicating evidence for the alternative hypothesis, a BF < 1 indicating evidence for the null hypothesis, and a BF = 1 indicating no support for either model [23,24]. Notably, the BF varies from 0 to infinite and is directly interpretable. For instance, a BF = 10 suggests that the alternative hypothesis is 10 times more likely than the null hypothesis based on the data, while a BF = 0.10 suggests that the null hypothesis is 10 times more likely than the alternative [26]. In other words, the larger the BF, the stronger the evidence for the alternative hypothesis; the closer the BF is to zero, the stronger the evidence for the null hypothesis. To facilitate a clearer interpretation of BFs, we chose to use [25,27] categorization into different levels of evidence (see Table 1).

Table 1. Categorization of Bayes Factors according to [25,27].

Evidence for the Alternative Hypothesis		Evidence for the Null Hypothesis	
Bayes Factor	**Strength of Evidence**	**Bayes Factor**	**Strength of Evidence**
1 to 3	Weak	1 to 1/3	Weak
3 to 10	Moderate	1/3 to 1/10	Moderate
10 to 30	Substantial	1/10 to 1/30	Substantial
30 to 100	Strong	1/30 to 1/100	Strong
>100	Very Strong	<1/100	Very Strong

In this study, a Bayesian approach was preferred over traditional NHST due to its ability to provide evidence for both the alternative and null hypotheses. However, traditional NHST metrics (i.e., p-values and F-statistics) are also reported for comparison. Analyses were conducted using R (version 4.4.1) and RStudio (version 2022.02.0). The BayesFactor (version 0.9.12-4.7) library was adopted for Bayesian analyses.

3. Results

The results are presented following the study's objectives. The blink rate (blinks per minute) was calculated by dividing the total number of blinks by the time of completion of each trial (in seconds) and then multiplied by 60. Data from one participant were excluded from the analyses involving eye blinks due to the eye tracker malfunctioning in one trial.

3.1. Objective 1. Assess the Accuracy of the Camera-Based System in Detecting Eye Blinks During Realistic Computer Tasks

To test the accuracy of the camera-based system, three repeated-measure Bayesian analysis of variance models were conducted with blink rate as a dependent variable and participant as a random factor. Comparable NHST metrics are reported in round parentheses. Marginal means and standard errors (SE) are displayed in Table 2, while Figure 5 visually shows the results.

Table 2. Marginal means and standard errors (SE) for blink rate, sorted by type of task, difficulty, and method.

Type of Task	Difficulty	Method	Blink Rate (Blinks/Min)	
			Mean	SE
Block-building	Easy	Camera-based	7.56	0.679
	Easy	Eye tracker	7.40	0.823
	Hard	Camera-based	7.66	0.733
	Hard	Eye tracker	6.98	0.826
Point-and-click	Easy	Camera-based	4.03	0.589
	Easy	Eye tracker	5.21	0.935
	Hard	Camera-based	3.31	0.580
	Hard	Eye tracker	4.92	0.852

A model with method (camera-based detection vs. eye tracker detection) as the independent factor was set up to investigate overall differences in blink rate across the two methods. A BF of 0.281 was found (F = 3.92, $p = 0.062$), indicating moderate evidence for the null hypothesis and suggesting that, overall, blink rates obtained using the camera-based system were not different from those obtained with the eye tracker (camera-based detection: mean = 5.64, SE = 0.521; eye tracker detection: mean = 6.13, SE = 0.704). A second model was set up to test the interaction between method and difficulty (easy vs. hard), finding a BF of 0.214 (F = 0.02, $p = 0.887$). This indicated moderate evidence for the null hypothesis, suggesting that no interaction was present. Finally, a third model was set up to test the interaction between method and type of task (block-building vs. point-and-click). Here, a BF of 1.725 was found (F = 19.04, $p < 0.001$), indicating weak evidence in support of the alternative hypothesis. To further test this interaction, two additional repeated-measure Bayesian analysis of variance were conducted, contrasting differences in the method within each task. When considering the block-building task, a BF of 0.524 was found (t = 1.78, $p = 0.087$; the p-value was not corrected for multiple comparisons, as such a correction would increase the likelihood of finding a null result), indicating weak evidence for the null hypothesis and suggesting no differences between methods. When considering the point-and-click task, a BF of 8.762 was found (t = 3.64, $p = 0.001$; the *p*-value was not corrected for multiple comparisons), indicating moderate evidence for the alternative hypothesis and suggesting a difference in blink rates when using different methods in this task (camera-based: mean = 3.67, SE = 0.520; eye tracker: mean = 5.07, SE = 0.781). Intuitively, this result might be explained by noting that participants were faster in completing the point-and-click task compared to the block-building task (55.6 s and 183.6 s on average, respectively), and the short amount of time might have caused the camera-based detection errors to impact more on the blink rate measure. To further explore this, we tested the same repeated-measure Bayesian analysis of the variance model using the total number of blinks (instead of blink rate) during the point-and-click task as the dependent variable, finding weak evidence for the null hypothesis (BF = 0.863) (see the discussion for more details).

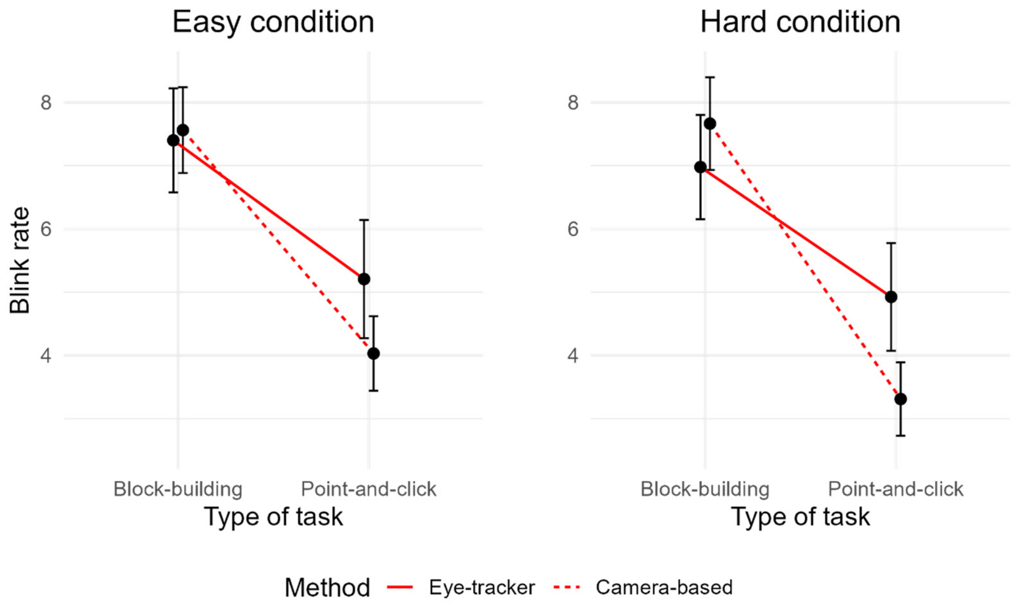

Figure 5. Mean blink rate, sorted by method and type of task. The left panel shows the eye blink rate in easy conditions; the right panel shows the blink rate in hard conditions. Error bars represent standard errors.

Overall, these results show evidence that the camera-based approach does not differ from the eye tracker approach, indicating that the camera-based system is accurate in detecting eye blinks during realistic computer tasks.

3.2. Objective 2. Assess the Accuracy of the Camera-Based System in Detecting Differences in Cognitive Load

NASA-TLX. First, we aimed to assess whether self-reported load differed between conditions. To test this, a raw TLX score was calculated by averaging the scores of each scale of the NASA-TLX questionnaire for each participant into a single variable. This was performed as past works have shown raw TLX scores to be as reliable as weighted TLX scores (see [28]). Two repeated-measures Bayesian analyses of variance were conducted with raw TLX as the dependent variable and participant as the random factor. A model with difficulty as the independent factor was set up to investigate differences in cognitive load between conditions. Here, a BF of 1.86×10^4 was found ($F = 29.49$, $p < 0.001$), indicating very strong evidence in support of the alternative hypothesis and suggesting that participants perceived the easy condition as less demanding than the hard condition (easy: mean = 6.16, SE = 0.729; hard: mean = 8.68, SE = 0.821). A second model was set up with the type of task as the independent factor to test the differences in self-reported cognitive load between tasks. Here, a BF of 2.029 was found ($F = 5.16$, $p = 0.032$), indicating weak evidence for the alternative hypothesis. In detail, the point-and-click task was perceived as less demanding than the block-building task (point-and-click: mean = 6.81, SE = 0.730; block-building: mean = 8.03, SE = 0.843). Finally, to explore the impact of each scale, six repeated-measures Bayesian analyses of variance models were conducted with each single scale as the dependent variable. Table 3 shows the result of each model, while Figure 6 visually displays the results.

Table 3. Results of the Bayesian analyses on single NASA-TLX scales. Each cell shows the Byes factor (BF) calculated for a specific effect (either type of task, difficulty, or the interaction). BFs > 1 indicate evidence for the alternative hypothesis, while BFs < 1 indicate evidence for the null hypothesis. Refer to Table 1 for a complete overview of BF categorizations.

Dependent Variable (Scale)	**Independent Variable**		
	Type of Task (BF)	**Difficulty (BF)**	**Type of Task × Difficulty (BF)**
Mental demand	0.251	1.356×10^6	1.508
Physical demand	33.959	197.504	0.274
Temporal demand	1.273	52.266	4.380
Performance	0.586	1.571	0.660
Effort	1.375	1.990×10^4	3.558
Frustration	0.706	17.050	1.928

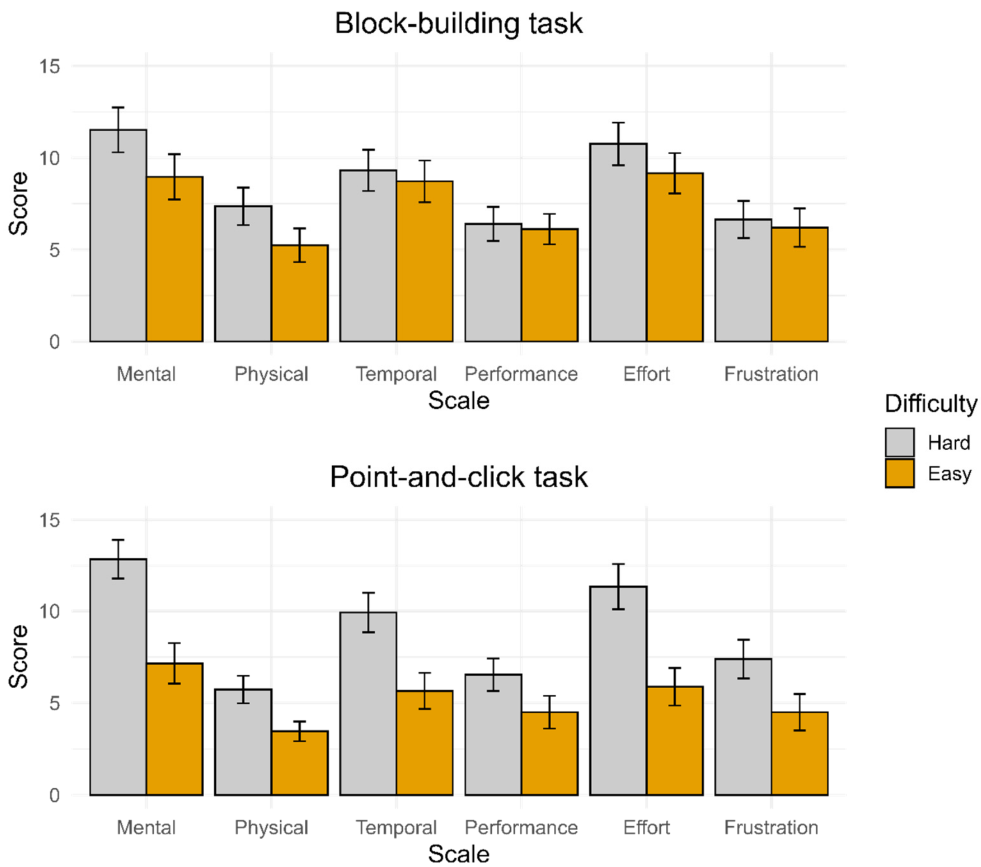

Figure 6. NASA-TLX scores sorted by scale and difficulty. The top panel displays scores referring to the block-building task; the lower panel displays scores referring to the point-and-click task. Error bars represent standard errors.

Overall, our analyses indicated that participants perceived the easy condition as less demanding. Moreover, weak evidence was found in support of the point-and-click task to be perceived as less demanding.

Blink rate. To assess the accuracy of the camera-based system in detecting differences in cognitive load, repeated-measure Bayesian analyses of variance models were conducted with camera-based blink rate as the dependent variable and participant as a random factor (marginal means are displayed in Table 2). Our analysis revealed a BF of 1.76×10^{10} (F = 41.73, $p < 0.001$) when the type of task was included as an independent factor, indicating very strong evidence for the alternative hypothesis and suggesting that blink rate differed between tasks. In detail, participants showed a lower blink rate during the point-and-click task compared to the block-building task (point-and-click: mean = 3.67, SE = 0.520; block-building: mean = 7.61, SE = 0.677), indicating that the block-building task imposed a higher cognitive load on participants. In contrast, both the model testing the main effect of difficulty and the model testing the interaction between difficulty and type of task revealed BFs of 0.233 and 0.335, respectively (F = 0.97, p = 0.334; F = 1.36, p = 0.255), indicating evidence for the null hypothesis. Similar results were found when using the blink rate coded with the eye tracker as the dependent variable, with a BF of 71.35 (F = 9.41, p = 0.005) when the type of task was included as an independent factor, a BF of 0.243 (F = 0.45, p = 0.507) when difficulty was included as an independent factor, and a BF of 0.276 (F = 0.02, p = 0.883) when the interaction was tested (see Figure 5 for reference).

Overall, blink rate results seem to suggest that the imposed mental load was higher during the block-building task but not during hard conditions. When difficulty is tested, results on blink rates are different from those found via self-reported measures. These findings are further discussed in Section 4.

4. Discussion

The first objective of this study was to assess whether the eye blink detection algorithm using a generic camera is as accurate as a scientific eye tracker visually inspected by an experimenter in detecting eye blinks during computer tasks that resemble stationary manufacturing and office tasks. Consistent with previous research using the same method (i.e., [19]), Bayesian statistics found no difference between the two methods. However, when analyzing the two tasks individually, we observed a difference between the two methods during the point-and-click task. Intuitively, this discrepancy was likely due to the short duration of this task (each trial lasted an average of ~25 s), during which participants blinked infrequently (5.07 blinks per minute when coded with the eye tracker). As a result, even a single eye blink detection error could significantly impact the blinks-per-minute calculation. Consistent with this result, previous research failed to validate our same camera-based method when using short video recordings (e.g., [15,29]). In fact, when the total number of blinks was used as the dependent variable instead of the blink rate, Bayesian statistics did not support the alternative hypothesis. In summary, our results indicate that camera-based systems using generic webcams can serve as a feasible solution for tracking blink rates in stationary working contexts, highlighting their potential as an alternative to scientific-grade eye trackers. Notably, however, researchers interested in using these tools should avoid employing a short measurement period (i.e., lower than 30 s), as this can impair the accuracy of the blink rate calculation.

The second objective was to test whether the camera-based system could accurately reflect differences in cognitive load between computer tasks that resemble stationary manufacturing and office tasks. As expected, both camera-based and eye tracker blink rate measures indicated differences in cognitive load between tasks, with the point-and-click task imposing lower cognitive demand on participants compared to the block-building task. This difference was also reflected in the self-reported measures from the NASA-TLX questionnaire, where the block-building task was perceived as more demanding in terms of overall cognitive load.

However, the difficulty of the task appeared to have affected only self-reported measures and not the blink rate. Consistent with previous research, this observation highlights the already mentioned divergence between subjective and physiological measures, suggesting that they might assess different constructs (cf., [11]). Indeed, while subjective measures could be biased by individuals' personal beliefs about their own efficacy or by beliefs on the self-reported scales themselves, physiological measures are not susceptible to people's subjective biases. In our view, to understand these results, it is important to note that the hard conditions were consistently longer than the easy ones (147 s for hard and 92.2 s for easy, on average). In line with these time differences, Bayesian statistics provided strong evidence supporting the alternative hypothesis when examining the effect of task difficulty on the temporal demand scale (BF = 52.3) but offered no evidence for either hypothesis when testing the type of task (BF = 1.3; see also Table 3 and Figure 6). This suggests that NASA-TLX scores may reflect differences in task duration rather than differences in the actual cognitive load imposed by the task. As already mentioned in the Introduction, several researchers have posed concerns about the uncritical use of the NASA-TLX questionnaire, indicating that in some cases, this instrument may lack sufficient construct validity (e.g., [9,30]). Additionally, it is important to note that while differences in completion times were observed both between tasks and conditions, a discrepancy between physiological and self-reported measures was found only in difficult conditions. This, once again, suggests a stronger reliance on self-reported measures on duration rather than blink rate. In summary, although NASA-TLX has traditionally been considered the gold standard in assessing cognitive load in manufacturing disciplines (cf., [9]), we believe that TLX scores should be compared with other measures, such as physiological ones, to ensure a comprehensive overview. Indeed, a camera-based eye blink rate can provide a cost-effective and efficient solution to complement self-reported measures.

In conclusion, our camera-based system appears to be a more affordable yet equally accurate alternative to scientific-grade eye trackers for measuring variations in cognitive load during stationary manufacturing and office tasks. This is an important finding, as generic webcams are commonly available in most workplaces and do not require calibration or expensive equipment, allowing for quick and efficient application. This method could be particularly useful in those working environments that require workers to perform stationary work. For instance, workers who perform various office-based tasks or stationary manufacturing tasks (such as activities involving quality control inspections) can easily monitor their cognitive load on a day-to-day basis and proactively detect overload before its onset. Given that perceived load does not always align with actual load (cf. [11]), a more objective physiological measure can be crucial in accurately preventing cognitive overload and, consequently, reducing the risk of psychological or physical complications [5,6].

Finally, despite its promising applicability, this study has some limitations. First, this study was conducted within a laboratory setting with a limited number of participants, which, while allowing for better control of confounding variables, lacks the ecological validity of in situ experiments. Future research should aim to replicate these results in real working environments with larger sample sizes. Second, this camera-based approach is limited to stationary positions where workers face the camera continuously. For other typologies of work, workers would need to use wearable devices, such as wearable eye trackers. Future studies could build on our findings to refine the camera-based blink detection algorithm for more dynamic conditions (e.g., see [31] for a novel approach to collecting eye metrics during dynamic tasks). Third, this study would have benefited from having tasks and conditions of equal duration. Future studies are encouraged to ensure consistent completion times across conditions by, for instance, asking participants to complete as many trials as possible within a specific time window.

Author Contributions: Conceptualization, F.G. and F.N.B.; Methodology, N.J. and L.Z.; Formal Analysis, N.V.; Data Curation, N.V. and N.J.; Investigation, F.N.B.; Writing—original draft, N.V.; Supervision, F.N.B.; Project administration, F.N.B.; Funding acquisition, F.N.B. All authors have read and agreed to the published version of the manuscript.

Funding: The authors received financial support for the research from Atlas Copco Inc., Mitacs, Social Sciences and Humanities Research Council (SSHRC), Natural Sciences and Engineering Research Council of Canada (NSERC), and WE-SPARK Health Institute.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

Conflicts of Interest: Author Frida Graf was employed by Atlas Copco Inc. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. And the authors declare that this study received funding from Atlas Copco Inc. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

References

1. Biondi, F.N. Adopting Stimulus Detection Tasks for Cognitive Workload Assessment: Some Considerations. *Hum. Factors* **2024**, *66*, 2561–2568. [CrossRef] [PubMed]
2. Engström, J.; Tech, V.; Kingdom, U.; Victor, T. Effects of Cognitive Load on Driving Performance: The Cognitive Control Hypothesis. *Hum. Factors* **2017**, *59*, 734–764. [CrossRef] [PubMed]
3. Cohen, J.; LaRue, C.; Cohen, H.H. Attention Interrupted: Cognitive Distraction & Workplace Safety. *Prof. Saf.* **2017**, *62*, 28–34.
4. Biondi, F.N.; Saberi, B.; Graf, F.; Cort, J.; Pillai, P.; Balasingam, B. Distracted worker: Using pupil size and blink rate to detect cognitive load during manufacturing tasks. *Appl. Ergon.* **2023**, *106*, 103867. [CrossRef]
5. Bonsang, E.; Caroli, E. Cognitive Load and Occupational Injuries. *Ind. Relat.* **2021**, *60*, 219–242. [CrossRef]
6. Vanneste, P.; Raes, A.; Morton, J.; Bombeke, K.; Van Acker, B.B.; Larmuseau, C.; Depaepe, F.; Van den Noortgate, W. Towards measuring cognitive load through multimodal physiological data. *Cogn. Technol. Work* **2021**, *23*, 567–585. [CrossRef]
7. Iskander, M. Burnout, Cognitive Overload, and Metacognition in Medicine. *Med. Sci. Educ.* **2019**, *29*, 325–328. [CrossRef]
8. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Adv. Psychol.* **1988**, *52*, 139–183.
9. de Winter, J.C.F. Controversy in human factors constructs and the explosive use of the NASA-TLX: A measurement perspective. *Cogn. Technol. Work* **2014**, *16*, 289–297. [CrossRef]
10. Rubio, S.; Díaz, E.; Martín, J.; Puente, J.M. Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Appl. Psychol.* **2004**, *53*, 61–86. [CrossRef]
11. Matthews, G.; De Winter, J.; Hancock, P.A. What do subjective workload scales really measure? Operational and representational solutions to divergence of workload measures. *Theor. Issues Ergon. Sci.* **2020**, *21*, 369–396. [CrossRef]
12. Braarud, P.Ø. Investigating the validity of subjective workload rating (NASA TLX) and subjective situation awareness rating (SART) for cognitively complex human–machine work. *Int. J. Ind. Ergon.* **2021**, *86*, 103233. [CrossRef]
13. Magliacano, A.; Fiorenza, S.; Estraneo, A.; Trojano, L. Eye blink rate increases as a function of cognitive load during an auditory oddball paradigm. *Neurosci. Lett.* **2020**, *736*, 135293. [CrossRef] [PubMed]
14. Al-Gawwam, S.; Benaissa, M. Robust eye blink detection based on eye landmarks and Savitzky-Golay filtering. *Information* **2018**, *9*, 93. [CrossRef]
15. Dewi, C.; Chen, R.C.; Chang, C.W.; Wu, S.H.; Jiang, X.; Yu, H. Eye Aspect Ratio for Real-Time Drowsiness Detection to Improve Driver Safety. *Electronics* **2022**, *11*, 3183. [CrossRef]
16. King, D.E. Dlib-ml: A machine learning toolkit. *J. Mach. Learn. Res.* **2009**, *10*, 1755–1758.
17. Kumar, A.; Patra, R. Driver drowsiness monitoring system using visual behaviour and machine learning. In Proceedings of the 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 28–29 April 2018; pp. 339–344. [CrossRef]
18. Kuwahara, A.; Nishikawa, K.; Hirakawa, R.; Kawano, H.; Nakatoh, Y. Eye fatigue estimation using blink detection based on Eye Aspect Ratio Mapping(EARM). *Cogn. Robot.* **2022**, *2*, 50–59. [CrossRef]
19. Biondi, F.N.; Graf, F.; Pillai, P.; Balasingam, B. On validating a generic camera-based blink detection system for cognitive load assessment. *Cogn. Comput. Syst.* **2023**, *5*, 255–264. [CrossRef]
20. Ellegast, R.P.; Kraft, K.; Groenesteijn, L.; Krause, F.; Berger, H.; Vink, P. Comparison of four specific dynamic office chairs with a conventional office chair: Impact upon muscle activation, physical activity and posture. *Appl. Ergon.* **2012**, *43*, 296–307. [CrossRef]

21. Benedetto, S.; Pedrotti, M.; Minin, L.; Baccino, T.; Re, A.; Montanari, R. Driver workload and eye blink duration. *Transp. Res. Part F Traffic Psychol. Behav.* **2011**, *14*, 199–208. [CrossRef]
22. Coyne, J.; Sibley, C. Investigating the use of two low cost Eye tracking systems for detecting pupillary response to changes in mental workload. *Proc. Hum. Factors Ergon. Soc.* **2016**, *60*, 37–41. [CrossRef]
23. Greenland, S.; Senn, S.J.; Rothman, K.J.; Carlin, J.B.; Poole, C.; Goodman, S.N.; Altman, D.G. Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *Eur. J. Epidemiol.* **2016**, *31*, 337–350. [CrossRef] [PubMed]
24. Quintana, D.S.; Williams, D.R. Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry* **2018**, *18*, 178. [CrossRef] [PubMed]
25. Held, L.; Ott, M. On *p*-Values and Bayes Factors. *Annu. Rev. Stat. Its Appl.* **2018**, *5*, 393–419. [CrossRef]
26. Rouder, J.N.; Speckman, P.L.; Sun, D.; Morey, R.D.; Iverson, G. Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* **2009**, *16*, 225–237. [CrossRef]
27. Held, L.; Ott, M. How the Maximal Evidence of P-Values Against Point Null Hypotheses Depends on Sample Size. *Am. Stat.* **2016**, *70*, 335–341. [CrossRef]
28. Hart, S.G. NASA-task load index (NASA-TLX); 20 years later. *Proc. Hum. Factors Ergon. Soc.* **2006**, *50*, 904–908. [CrossRef]
29. Pandey, N.N.; Muppalaneni, N.B. Real-Time Drowsiness Identification based on Eye State Analysis. In Proceedings of the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 25–27 March 2021; pp. 1182–1187. [CrossRef]
30. McKendricka, R.D.; Cherry, E. A deeper look at the NASA TLX and where it falls short. *Proc. Hum. Factors Ergon. Soc.* **2018**, *1*, 44–48. [CrossRef]
31. Chen, H.; Zendehdel, N.; Leu, M.C.; Yin, Z. Real-time human-computer interaction using eye gazes. *Manuf. Lett.* **2023**, *35*, 883–894. [CrossRef]