ESCOLA
SUPERIOR
DE SAÚDE
POLITÉCNICO
DO PORTO

M

MESTRADO
Higiene e Segurança nas Organizações

Advancing the Understanding of Pupil Size Variation in Occupational Safety and Health: A Systematic Review and Evaluation of Open-Source Methodologies

Daniela Filipa Campos Ferreira

09/2023

ESCOLA
SUPERIOR
DE SAÚDE

Center for Translational Health
and Medical Biotechnology Research

Advancing the Understanding of Pupil Size Variation in Occupational Safety and Health: A Systematic Review and Evaluation of Open-Source Methodologies

Autor

Daniela Filipa Campos Ferreira

Orientador(es)

Prof. Doutora Matilde Alexandra Rodrigues da Escola Superior de Saúde do Instituto Politécnico do Porto

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Mestre em **Higiene e Segurança nas Organizações** – Segurança e Saúde Ocupacional pela Escola Superior de Saúde do Instituto Politécnico do Porto.

Agradecimentos

Gostaria de expressar a minha sincera gratidão a todas as pessoas que contribuíram para a conclusão desta dissertação.

Primeiramente, desejo agradecer à minha orientadora, Prof. Doutora Matilde Rodrigues, pela sua orientação valiosa e incansável. A sua orientação ajudou-me a aprofundar o meu conhecimento e a aprimorar a minha capacidade de pesquisa.

Gostaria também de agradecer ao Simão Ferreira, pela sua colaboração e suporte ao longo desta jornada.

A um nível mais pessoal, desejo expressar a minha profunda gratidão à minha mãe, Adelina Ferreira, ao meu irmão, Daniel Ferreira e à memória do meu amado pai, José Ferreira. A eles, dedico esta dissertação.

Por último, mas não menos importante, quero agradecer ao meu colega, amigo e companheiro Ricardo Sousa, por estar ao meu lado durante todo o percurso. O seu apoio emocional, encorajamento constante e compreensão foram essenciais para o meu bem-estar na realização deste projeto.

A todos aqueles que mencionei e a todos os outros que de alguma forma contribuíram para este trabalho, o meu mais profundo agradecimento. As suas palavras de incentivo, conselhos e suporte foram inestimáveis e ajudaram-me a superar os obstáculos ao longo deste caminho acadêmico.

À minha família, amigos e todos que me acompanharam nesta jornada, a minha gratidão eterna.

Resumo

O tamanho da pupila pode ser usado como um importante biomarcador de riscos ocupacionais. Nos últimos anos, tem havido um aumento no desenvolvimento de ferramentas de código aberto dedicadas à obtenção e medição do diâmetro da pupila. No entanto, a adequação dessas ferramentas para serem usadas em ambientes ocupacionais ainda não está clara. Este estudo explora a importância da variação do tamanho da pupila como biomarcador de riscos ocupacionais e revisa métodos de código aberto existentes para avaliar a sua potencial aplicabilidade em pesquisas e ambientes ocupacionais, com o objetivo de prevenir acidentes e melhorar a saúde e o desempenho dos trabalhadores. Para isso, foi realizada uma revisão sistemática da literatura em duas fases nas bases de dados Web of Science™, Science Direct® e Scopus®. Para a relevância de monitorizar a variação do tamanho da pupila em ambientes ocupacionais, foram incluídos 15 artigos. Os artigos foram divididos em três grupos: carga mental, stresse ocupacional e fadiga mental. Na maioria deles, a dilatação da pupila aumentou com o aumento da carga de trabalho e com os níveis mais altos de stresse. No que diz respeito à fadiga, à medida que sua intensificação era observada, notava-se uma diminuição no tamanho da pupila. Em relação às metodologias de código aberto, foram identificados 16 artigos, que foram separados em duas categorias: algoritmos e software. Para os algoritmos, as redes neurais convolucionais tiveram melhor desempenho. Considerando o grupo de software e os resultados anteriores, o MEYE, que pode usar uma webcam de computador, indica ser a melhor opção de sistema de código aberto a ser usado em ambientes ocupacionais. Pode ser implementado para trabalhadores com ambientes regulares, como desenvolvedores e administradores.

Palavras-chave: Rastreamento ocular; Avaliação do tamanho da pupila; Fatores psicológicos; Segurança e Saúde Ocupacional.

Abstract

Pupil size can be used as an important biomarker of occupational risks. In recent years, there has been an increase in the development of open-source tools dedicated to obtaining and measuring pupil diameter. However, the appropriateness of these tools to be used in occupational settings is yet not clear. This study explores the importance of pupil size variation as a biomarker for occupational risks and review existing open-source methods to assess their potential applicability in research and occupational settings, aiming to prevent accidents and improve worker's health and performance. To this end, a two-phase systematic literature review was carried out in the Web of Science™, Science Direct® and Scopus® databases. For the relevance of monitoring pupil size variation in occupational settings, 15 articles were included. The articles were divided into three groups: mental workload, occupational stress, and mental fatigue. In most of them, pupil dilation increased with workload enhancement, and with higher levels of stress. With regards to fatigue, as far it was observed its increase, a decrease in pupil size was noticed. In what regards to the open-source methodologies, 16 articles were identified, which were separated into two categories: algorithms and software. For algorithms, the convolutional neural networks had better performance. Considering the software group and the previous results, MEYE, which can use a computer webcam, should be the best open-source system option to be used in occupational settings. This feature positions MEYE as a particularly practical tool for workers in stable environments, like those of developers and administrators.

Keywords: Eyetracking; Pupil size evaluation; Psychological factors; Occupational Safety and Health.

Index

1. Introduction

According to statistics from the International Labour Organization (ILO), every 15 seconds, an employee dies as a result of a work-related accident, or due to a disease related to their professional activity. This translates to nearly 6,300 deaths per day out of a total of 2.3 million deaths per year (OIT, 2023). However, no one should suffer from work-related illnesses or accidents (European Commission, 2021). The European Strategic Framework on Health and Safety at Work 2021-2027 denotes that the prevention of occupational accidents and diseases is an important aspect of the sustainability and competitiveness of the economy of the European Union (EU) (European Commission, 2021). Consequently, risk management at organizations must prioritize the safety and well-being of workers in all working settings, particularly in the current context, characterized by ecological and digital transitions.

In today's world, the population is assisting to a significant technological transition in several occupational settings. The digitization and automation of workstations and processes are increasing (European Commission, 2021). At the same time, workers are spending more time behind screens, and the cognitive demands of their tasks are also increasing (European Commission, 2021). In fact, these developments come with new risks or increasing the existing ones. Mental and cognitive risk factors, that can result in occupational illnesses and work-related accidents are one of the problems raised by these changes (Bonsang & Caroli, 2021; Kalakoski et al., 2020). This was highlighted as one of the three objectives of the EU's Strategic Framework on Health and Safety at Work (European Commission, 2021). Therefore, ensuring the mental state and cognitive ability of workers is very important for health, safety, and economic reasons.

A way to address workers' mental and cognitive status might be through biometric measures, such as pupil size. The autonomic nervous system, which is involved in many bodily and behavioural operations, regulates pupil size variation (Wangwiwattana et al., 2018). The sympathetic system regulates pupil dilation, while the parasympathetic system governs pupil constriction. These actions are achieved through the control of eye muscles (Alberdi et al., 2016; Kret & Sjak-Shie, 2019; Nguyen et al., 2022; Nurçin et al., 2017).

Pupil size variation has been reported to be extensively linked to neural activity (Krol & Krol, 2017).This effect transcends various disciplines, making pupillary monitoring a valuable tool employed in numerous fields such as neuroscience, psychology, ophthalmology, and business applications, among others (Nguyen et al., 2022). Researchers have employed pupil size as an indicator of cognitive workload, attention load, fatigue, and emotional arousal (Nguyen et al., 2022). Specifically, during challenging tasks such as short-term memory tasks, pupils tend to dilate more (Nguyen et al., 2022). In addition, pupil size varies while people concentrate, focus, relax and generally think (Wangwiwattana et al., 2018). Thus, pupil size measurement could be used for Occupational Safety and Health (OSH) (Dalveren et al., 2018; Iqbal et al., 2018). Since mental state and cognitive ability of individuals can impact their performance, and consequently, the entire system, monitoring it through pupil size can yield valuable insights to prevent human errors, improve safety, and increase productivity (Dalveren & Cagiltay, 2018). Furthermore, stress is a growing problem in today's modern society and a part of humans' daily lives, including the workplace (Alberdi et al., 2016), and pupillometry is a procedure that allows to detect stress in real time (Graff et al., 2019). When integrated into a system, pupillometry can assist workers in recognizing when to pause or seek assistance, as well as alert co-workers to a mental workload threshold level or to a stressful situation.

Furthermore, Computer Vision Syndrome (CVS) is recognized as an occupational problem affecting employees who spend more than 3 to 4 hours a day in front of a computer during their work week (Coronel-Ocampos et al., 2022; Randolph, 2017; Zenbaba et al., 2021). It is also one of the leading occupational hazards of the twenty-first century (Zenbaba et al., 2021), affecting 75% to 90% of computer users (Boadi-Kusi et al., 2020; Coronel-Ocampos et al., 2022; Singh et al., 2022). Therefore, it is crucial to increase public awareness (Boadi-Kusi et al., 2020), find indicators to detect possible CVS symptoms, and improve workplace environments for those making intensive use of computers (Lapa et al., 2023). These measures are crucial not only to minimize causes that reduce occupational efficiency and impact workers' personal lives as well, but also to enhance productivity and workplace wellness (Gomes & Preto, 2015; Randolph, 2017; Seguí et al., 2015; Yan et al., 2008; Zayed et al., 2021). Since several of the CVS symptoms are eye-related (Blehm et al., 2005; Boadi-Kusi et al., 2020; Lapa et al., 2023; Munshi et al., 2017; Reddy et al., 2013) and pupil size changes also impact basic visual functions such as visual sensitivity and acuity (Binda et al., 2014), measuring pupil size and detecting its variations could be an indicator to help prevent or detect CVS.

Pupillometry is a non-invasive index that can be obtained without interfering with normal behavior, and that has been extensively studied in the literature (see, e.g., Binda et al., 2014; Binda & Murray, 2015; Cao et al., 2021). The pupil light response, a well-known reflex in which the pupil constricts when exposed to bright light and dilates when exposed to dim light, is an important consideration when acquiring pupil measurement data (Gao et al., 2020; Pan et al., 2022). The measurement of pupil diameter and its dynamic variations is subject to certain physiological limitations. It is essential for researchers to recognize and account for these limitations in their experiments, capturing data in high sensitivity ranges, within optical constrained operational limits. By controlling lighting conditions, researchers can regulate the level of ambient illumination and minimize potential confounding factors that may influence pupil size. Moreover, manipulating lighting conditions offers an opportunity to investigate the specific effects of different lighting levels or characteristics on pupil responses to other stimuli. This approach enables researchers to fully explore the range of sensitivity of the pupil as a dynamic indicator of other factors. Nowadays, pupil measurements are contactless, easily accessible, and objective, requiring only minor cooperation from the participant being tested (Kelbsch et al., 2019). Generally, most modern pupillary monitoring devices, whether for experiments or those already implemented, now use infrared-based eye trackers (Wangwiwattana et al., 2018). Simply put, pupil size can be measured using eye-tracking methodology (Bachurina et al., 2022; Kret & Sjak-Shie, 2019).

Moreover, pupillometry and eye-tracking methodologies offer the notable advantage of providing real-time data, enabling inference about an individual's current state or establishing a direct relationship between specific stimuli and their corresponding pupil responses. This real-time capability allows for immediate insights into cognitive and emotional processes as they unfold. Furthermore, the cost-effectiveness of this approach is noteworthy as it can utilize a simple web camera as an acquisition device for certain applications, reducing the need for specialized and expensive equipment.

Considering recent advancements in pupil diameter monitoring systems, it is crucial to gain a deeper understanding of their relevance for OSH and the applicable tools in occupational settings. To address this, two primary objectives were established for this study. The first objective is to conduct a systematic literature review to understand the importance of pupil size variation in OSH. The second objective is to evaluate existing open-source methodologies for acquiring pupil size variation data. The aim is to consolidate and discuss these findings in a single work to provide a comprehensive understanding.

1.1. Research Question

Understanding the impact of occupational risks on workers' health and safety is of paramount importance for organizations, regulatory bodies, and OSH practitioners. Traditional methods of risk assessment often rely on self-report measures or subjective evaluations, which may be prone to biases or limitations. The measurement of pupil size has gained attention as a potential objective and real-time measure of cognitive workload, mental stress, and attentional demands. Based on the PICO framework (Population, Intervention, Comparison, Outcome) the following research question was formulated: "*In working populations exposed to occupational risks, does the measurement of pupil size predict or correlate with adverse health outcomes, cognitive performance, or safety accidents?*" By critically analysing and synthesizing the available evidence, this review aims to provide insights into the potential role of pupil size measurement as an indicator of occupational risks and its implications for workers health and safety in occupational settings. Despite some drivers exerting professional activities, they are out of the scope of this research.

2. Material and Methods

The current systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021). The review was carried out in two phases, aligning with the defined primary objectives to this study.

2.1. Search strategy

The query was refined from December 10th of 2022 to the 15th of January 2023, and independent searches were conducted in selected electronic databases to find relevant articles published in the previous ten years (2012-2022): Web of Science™, Science Direct®, and Scopus®. The search was carried out using the keywords described in Table 1 that were selected through the Medical Subject Headings website. These keywords were searched in all fields for phase 1, as it is the main objective of the review. For phase 2, the keywords were searched in the article fields of title, keywords, and abstract, once it was the secondary goal of the study.

Due to the limitation of eight boolean connectors, in Science Direct® the keywords applied were fewer. For phase 1 the defined search equation was ("Pupil Size" OR "Pupillometry") AND ("Health" OR "Safety" OR "Accidents" OR "Human error") AND ("Occupational" OR "Job") NOT ("Drivers"). For phase 2 the search was made with ("Pupil Size" OR "Pupillometry" OR "Pupil Variation") AND ("Open source" OR "Open-source" OR "Web app").

Table 1 – Keywords for the review research.

Phase			
Phase	1 and 2	Pupil size AND	"Pupil Size" OR "Pupillometry" OR "Pupil size estimation" OR "Pupil dilation" OR "Pupil diameter" OR "Pupil detection" OR "Pupil Segmentation" OR "Pupil measurement" OR "Pupil constriction" OR "Pupillary response" OR "Pupil Variation" OR "Pupillography"
	1	Safety and Health AND	"Health" OR "Safety" OR "Mental Health" OR "Accidents" OR "Human error"
		Occupational Settings	"Occupational" OR "Job" OR "Work" OR "Industry" OR "Job site" OR "Workplace" OR "Work place" OR "Worker" OR "Employee" OR "Occupation" OR "Operators"
		NOT	"Drivers" OR "Traffic" OR "Children" OR "Infants"
	2	Open source	"Open source" OR "Open-source" OR "Web app"

2.2. Eligibility criteria

Exclusion criteria were established: 1) Date of publication: articles prior to 2013 were excluded; 2) Type of document: non-review articles were selected to reduce the likelihood of using repeated sources; 3) Language: articles not in English were removed. The fourth and final exclusion criterion differed between the phases, based on their respective objectives. For phase 1, the criterion was that articles should establish a relationship between pupil size variation and OSH, removing those that were inconsistent with the relationship. For phase 2, the criterion involved removing studies that did not provide a description of the methodology used to obtain pupil size variation values.

2.3. Screening criteria

After the initial search, in both phases, the first to third exclusion criteria were applied using database automatic tools. Then, search results were exported to EndNote® X9 software where duplicates were identified and removed. Subsequently, the articles underwent a screening process. The fourth exclusion criterion was manually done. Initially, the titles of articles were examined for relevance. Secondly, abstracts were screened, giving particular importance to the study aims and methodology. Finally, full text articles were retrieved for those studies appearing to meet the eligibility criteria, and for those where the information in the title and abstract was insufficient for exclusion (see Figure 1 and Figure 2).

2.4. Quality assessment

The quality of the studies included in phase 1 was evaluated using the Mixed Methods Appraisal Tool (MMAT) Version 2018 (Hong et al., 2018). The user guide provided instructions on the appropriate section to be applied in the articles and explained the questions to assess a correct answer. According to this guide, if one or both screening questions were answered with a "No" or "Can't tell", the article was not appraised using the MMAT. In this review, none of the included articles met these criteria (see Appendix 1), allowing for the use of the MMAT for all phase 1 articles. For each included study, the appropriate category was defined using the user guide's algorithm for selecting the study categories in the MMAT. Upon evaluation, the studies were classified under the third category "Quantitative non-randomized" specifically as either "Before-and-after study Time series" or "Non-randomized cross-over design". As a result, questions from the third category were consistently applied across all included studies. Although calculating an overall score from the ratings is discouraged, only one category was employed in this case. Therefore, the articles were categorized as low, medium, or high quality based on the number of "No" and "Can't tell" responses. No assessment of article quality was conducted in phase 2 as it was not the primary objective of the review but rather a compilation of existing methodologies.

3. Results

3.1. Relevance of pupil size for OSH

Initially, 1311 articles were selected after applying criteria one to three (Figure 1). Before proceeding to the fourth criterion analysis, 83 duplicate articles were excluded, resulting in 1228 titles. Based on the last exclusion criterion, 1142 titles were discarded. Thus, 86 abstracts were analyzed, and following the same premise, 40 articles were further excluded. The remaining 46 articles were examined in full. In this last evaluation, 19 articles did not specify the workgroup or type, leading to their exclusion. Additionally, five articles that involved random degree-seeking students as a sample and three articles that focused on a specific group of students without work experience were also discarded. One article, although having a clear workgroup as a sample, did not present a clear relationship between the factors and pupil size variation, and was therefore rejected. Furthermore, one article was removed as it represented a duplicate experiment with the same conclusions already included in a different journal. Moreover, two articles could not be accessed in full and were consequently excluded as well.

The articles included in this phase are described in Tables 2 to 4, providing details such as the sample size, evaluation methods, and their relationship to pupil size variation. For further analysis, the articles were categorized into three groups: workload, stress, and fatigue. The quality assessment of this phase is presented in detail in Appendix 1

Various brands of eye trackers were utilized among the studies included in this review. These brands include SensoMotoric Instruments (SMI) (Bertilsson et al., 2020; Cabrera-Mino et al., 2019; Couceiro et al., 2019; Y. Wu et al., 2019) (25%), Tobii Technology (Bhavsar et al., 2016; Gao et al., 2019; Naeeri et al., 2021; Naskrent et al., 2022; Othman & Romli, 2016; Sharma et al., 2021; Szulewski et al., 2017; C. Wu et al., 2020; Zhang et al., 2018) (56%), and VisionTrak (Mazur et al., 2013; Mosaly et al., 2013) (19%).

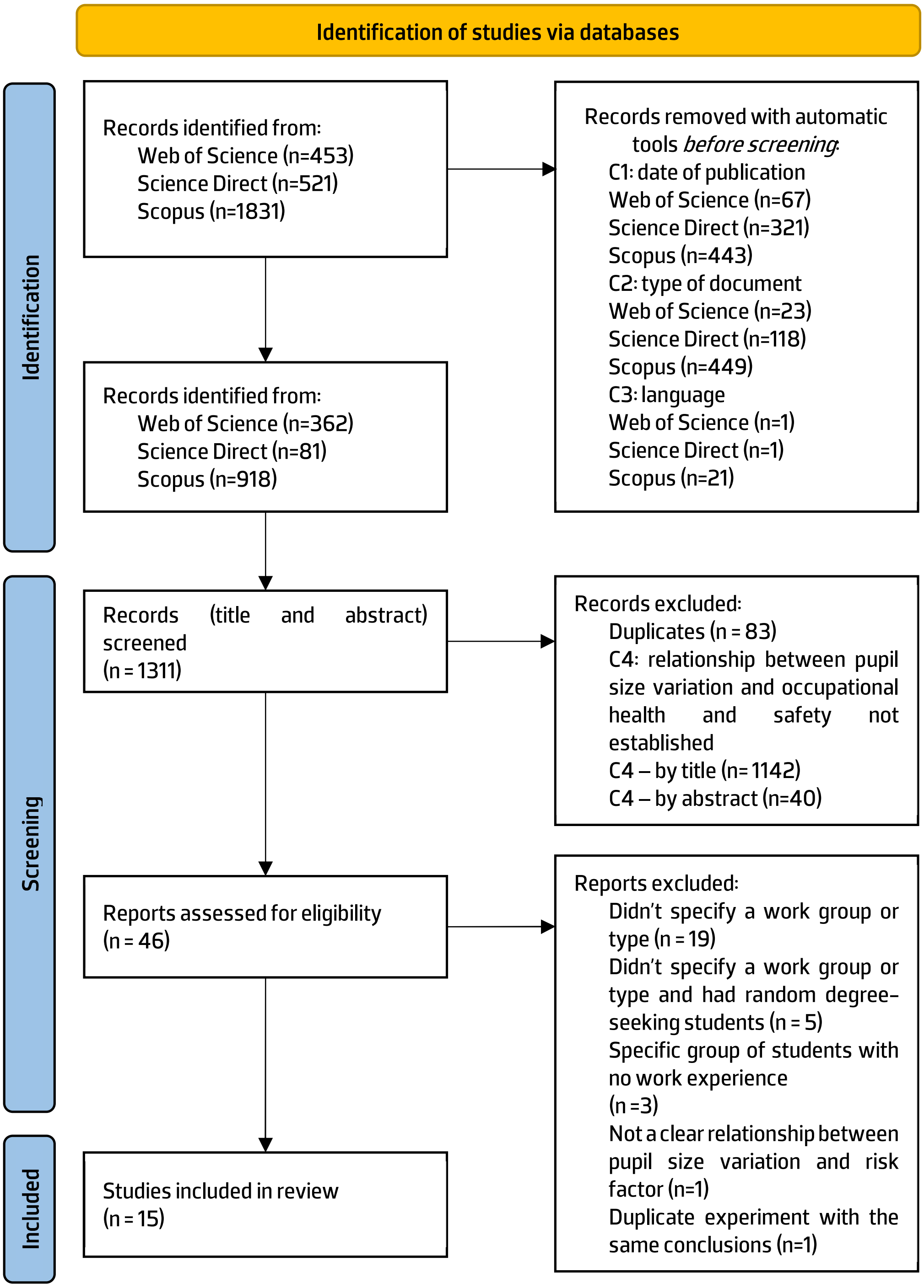


Figure 1: Flow diagram of study selection for systematic review about the relevance of pupil size for OSH.

3.1.1. Workload

The group of studies presented in Table 2 examines the relationship between workload and pupil size variations. The experiments were conducted with different professionals, including physicians, plant operators, pilots, nurses, programmers, engineers, and harvesting operators.

The use of pupillometry measurements to establish a relationship between workload and task performance has been explored in previous studies. Othman and Romli (2016) utilized pupillometry to assess mental workload in pilots and observed a correlation between an increase in workload and a decrease in task performance. They correlated NASA-TLX questionnaire results with pupil dilation, finding that pupil increases with workload enhancement.

Variation of workload measured through pupillometry across tasks with different difficulty levels was also analyzed in several of the selected studies. Couceiro et al. (2019) investigated workload in programmers using pupillography and NASA TLX questionnaire, they concluded that pupil dilation increases with task difficulty. Y Wu et al. (2019) and C Wu et al. (2020) studied workload in nuclear power plant operators and robotic surgical tasks, respectively, using NASA-TLX questionnaire and eye-tracking measurements to detect workload variations across different task difficulty levels. In these studies, pupil diameter and gaze entropy were used, in Y Wu et al (2019) both measures rose as task level difficulty increased, while in C Wu et al (2020) only gaze entropy had significant correlations with workload. Sharma et al. (2021) analyzed workload in ultrasound operators and observed changes in pupil diameter related to task phases. The pupil diameter grew larger with workload, after the first freeze task. The workload was related to the task but not to the experience. Naskrent et al. (2022) compared mental workload in different timber harvesting operations and found slightly higher workloads and pupil dilations during felling operations.

Pupillometry was also applied to assess cognitive load under stressful situations. Szulewski et al. (2017) explored cognitive load with Paas questionnaire (Paas, 1992) and pupillometry in physicians, observing higher cognitive load and larger pupil diameter in novices compared to experts. Also, Cabrera-Mino et al. (2019), compared novice and expert nurses' stress and workload levels, finding higher pupil dilation and cognitive load in novices. Bhavsar et al. (2016), by studying plant operators, observed an increase in pupil dilation when cognitive load increased during alarm situations.

Table 2 – Summary of studies that analysed the relationship of a workload and pupil size variation.

Reference	Sample	Evaluation method: Subjective	Evaluation method: Objective	Relation with pupil size
(Mazur et al., 2013)	9 radiology physicians (4 experient professors and 5 residents)	National Aeronautics and Space Administration Task Load Index (NASA-TLX) questionnaire	Samples at 60 Hz, were taken with a head-mounted VisionTrak eye-tracking device (Massachusetts, United States of America)	No correlation was found between the average pupil diameter and the NASA-TLX scores
(Mosaly et al., 2013)	8 radiology physicians (3 experienced professors and 5 residents)			
(Bhavsar et al., 2016)	44 plant operators	Not measured	Samples at 120 Hz were made with Tobii Technology (Danderyd, Sweden)	Pupil dilation increases with workload enhancement
(Othman & Romli, 2016)	13 pilots with commercial aircraft experience.	NASA-TLX questionnaire	Samples at 50 Hz were taken with Tobii Glasses 2.0 (Danderyd, Sweden)	
(Szulewski et al., 2017)	32 physicians with resuscitation medicine experience (13 novices; 9 senior residents; 10 experts)	The psychometric mental effort questionnaire developed by Paas (1992)	Samples at 30 Hz were obtained with the Tobii Glasses Eye Tracker (Danderyd, Sweden)	
(Zhang et al., 2018)	16 laparoscopic surgeons and medical students with laparoscopic experience.	NASA-TLX questionnaire	Eye tracking signals were captured with Tobii Glasses 2.0 (Danderyd, Sweden)	
(Couceiro et al., 2019)	30 experienced programmers in Java programming language (12 intermediate; 14 advanced; 4 experts).		Samples of 20 Hz were taken with a SensoMotoric Instruments (SMI) eye tracker (Berlin, Germany)	
(Gao et al., 2019)	24 postgraduate students with laparoscopic surgery training and clinical practice experience.		Samples at 50 Hz were taken with Tobii Glasses 2.0 (Danderyd, Sweden)	
(Y. Wu et al., 2019)	39 engineers (32 nonexperts and 7 experts)		Samples at 50 Hz were taken with the iView X head-mounted eye-tracking device (SMI, Germany)	
(C. Wu et al., 2020)	8 surgical trainees who participated in robotic skills training		Samples at 50 Hz were taken with Tobii Glasses 2.0 (Danderyd, Sweden)	Pupil diameter had no significant correlation with NASA-TLX scores
(Sharma et al., 2021)	12 obstetrics physicians (3 newly qualified and 9 experienced)	Not measured	Samples of 90 Hz were taken with a Tobii Eye Tracker (Danderyd, Sweden)	Pupil dilation increases with workload enhancement
(Naskrent et al., 2022)	24 experienced operators, 6 for each cutting variant		Samples were taken with Tobii Glasses 2.0 (Danderyd, Sweden)	

Other studies related both subjective measures (e.g. NASA TLX) and objective measures (pupillometry) to the task workload. In the field of radiology, Mazur et al. (2013) and Mosaly et al. (2013) investigated workload in physicians and found differences under different workloads between cases when subjective measures were applied, but not when objective measures were used. In opposite, studies on laparoscopic surgeons, conducted by Zhang et al. (2018) and Gao et al. (2019), identified increases in pupil diameter as indicators of workload, with NASA-TLX scores showing similar patterns.

3.1.2. Stress

Two studies focused on stress as a risk factor to work performance. In both cases authors used pupil diameter as a biomarker for stress. Cabrera-Mino et al. (2019) found that novices showed higher stress responses than experts in four out of six tasks, as indicated by larger pupil dilation.

Table 3 - Summary of studies that analysed the relationship of stress and pupil size variation.

Reference	Sample	Evaluation method		Relation with pupil size
		Subjective	Objective	
(Cabrera-Mino et al., 2019)	13 novice nurses (near graduation students) and 15 expert nurses (>5 years of clinical experience).	Not measured	Video capture was obtained with ETG's (SMI) software version 2.7 at a rate of 24 Hz audio and 1280 x 960 video resolution. The pupil diameter was measured at a rate of 60 Hz	The rise in stress and workload is significantly related to higher pupil dilation
(Bertilsson et al., 2020)	10 police officers with at least 5 *years* of experience with fieldwork	Two or three Go-Pro™ 3.0 cameras recorded participants' actions (including sound) and specialists rated participants' performances with self-made questionnaire of the study	Heart rate and body posture were recorded with Zephyr Bioharned® 2.0 (Annapolis, USA). Pupil diameter was obtained with an SMI eye tracker (Berlin, Germany) at a sampling rate of 30 Hz	The pupil diameter is larger for threats that appear immediately when compared to threats that appear with delay. The average size of the pupil was not significantly different during moderate threat tasks when compared with higher threat tasks. Repeating tasks decreases pupil diameter significantly

Bertilsson et al. (2020) examined police officers' responses to stress and concluded that subjective and objective evaluations of performance were concordant. The study found that task repetition led to improved performance and a significant decrease in pupil diameter. In terms of threat levels, high-threat situations significantly affected motor control performance, while biomarkers did not differ significantly between moderate and high stress levels. Immediate threats resulted in larger pupil diameter and lower perception skill scores compared to delayed threats. The findings suggest that pupil activity, along with heart rate, can serve as a reliable physiological marker for recording and assessing stress levels.

3.1.3. Fatigue

A single article that related pupil size with fatigue levels was identified. In that study involving pilots, Naeeri et al. (2021) correlated fatigue with eye-tracking measures. The authors found that expert pilots had faster reaction times and better performance compared to novices in the psychomotor vigilance test (PVT). Eye movement measures showed a decrease in eye fixation and pupil size over the flight, with novices exhibiting a higher rate of pupil size decrease. Eye fixation duration positively correlated with PVT measures, while pupil size and eye fixation number had a negative association. The study suggests that eye movement measures could be used as an alternative to the PVT for assessing pilot fatigue.

Table 4 – Summary of studies that analysed the relationship of fatigue and pupil size variation.

Reference	Sample	Evaluation method	Relation with pupil size
(Naeeri et al., 2021)	20 pilots (10 novices and 10 experts)	The PVT measures were assessed using the Psychology Experiment Building Language software, version 0.13. And eye tracking data was collected with the Tobii TX 300 eye tracker (having a 300 Hz data collection rate with 0.5 degrees of visual angle accuracy)	Pupil size decreases with fatigue augment

3.2. Open-source methodologies for pupil size analysis

Of the 51 articles held for further evaluation after the first three criteria, 20 were duplicates that were eliminated. Based on the fourth criterion, nine articles were excluded after reviewing their titles. Subsequently, 22 abstracts were screened, resulting in the exclusion of 10 articles. Among the remaining 12 articles that were assessed in full, one was excluded because it did not provide a description of the methodology used to obtain pupil size variation values (fourth exclusion criterion). From the final set of 11 articles, an additional five articles were retrieved and included in the review. The total of 16 included articles was further categorized into two groups: algorithms and software (Tables 5 and 6, respectively).

3.2.1. Algorithms

Eight of the articles selected for this phase concerned algorithms that enable pupil size detection (Figure 2). The methodology to do so and some detection rates with a 5-pixel error are described in Table 5, along with the reference and algorithm name. Regarding the mean detection rate percentage presented, the means were calculated as shown in Appendix 2.

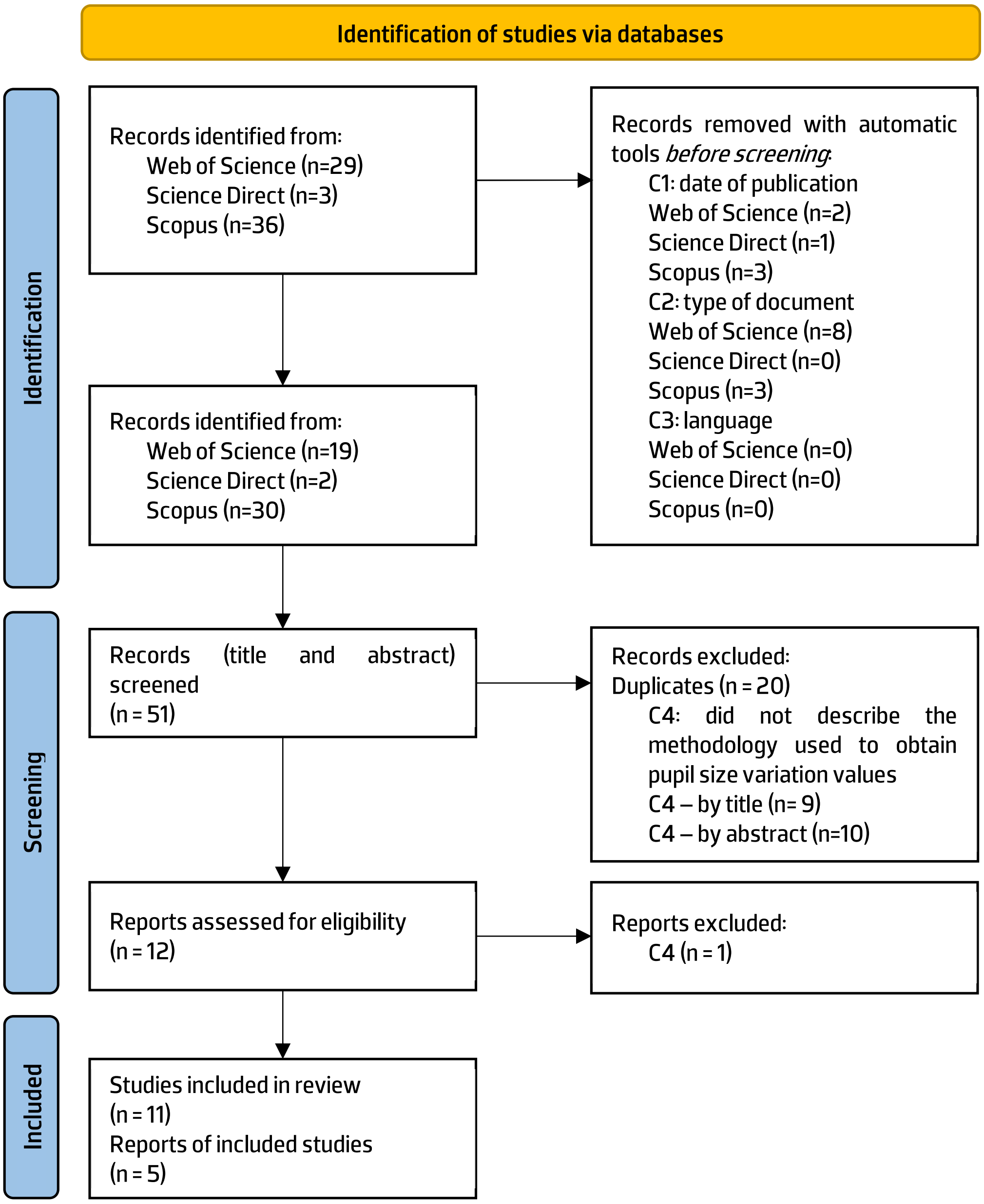


Figure 2: Flow diagram of study selection for systematic review about pen-source methodologies for pupil size analysis.

Table 5 – Summary of studies that address the algorithms to determine pupil size variation. ✓ = Applies; ⊗ = Does not apply; NM = Not measured.

Reference	System	Grayscale images [1]	Approach basis	Curve edge detection/filtering[3]	LSFE[5]	CHT[6]	S. DR[7]	E. DR[8]	EE. DR[9]
(Fuhl, Kübler, et al., 2015)	ExCuSe	✓	Edge and Thresholding	✓	✓	⊗	86,17%	79,20%	57,07%
(Javadi et al., 2015)	SET		Thresholding	✓	✓	⊗	63,0%	NM	25,87%
(Fuhl, Santini, et al., 2015)	ElSe		Edge	✓	✓	⊗	82,0%	79,39%	70%
(Santini, Fuhl, & Kasneci, 2017)	PuRe		Edge	✓	✓	⊗	80%	88%	76%
(Santini et al., 2018)	PuReST		Edge	✓	✓	⊗	86%	88%	79%
(Vera-Olmos et al., 2018)	DeepEye		DCNN	✓	⊗	⊗	54%	85%	87%
(Bonteanu et al., 2019)	PDA		Thresholding	✓	⊗	✓	NM	82,44%	NM
(Eivazi et al., 2019)	CNN-BPD		CNN	⊗	⊗	⊗	74%	88%	87%

[1] Input
[2] Utilization of values to refine image (binarization) or definition of dark and bright masks
[3] Curve edges detection, selection or/and filtering
[4] Least square fitting to ellipse
[5] Circular Hough Transformation
[6] Detection rate with 5 pixel error on Swirski (2012) dataset* – Highly off-axis images
[7] Detection rate with 5 pixel error on ExCuSe XII dataset (ExCuSe)* – Images with bad ilumination
[9] Mean detection rate with 5 pixel error of ExCuSe and ElSe datasets*
*This measures were taken from only two studies (with decimals from (Fuhl, Santini, et al., 2015), integers from (Eivazi et al., 2019)) to prevent significant differences regarding used CPUs.

The ExCuSe algorithm outperformed other methods, utilizing the least square fitting to an ellipse (LSFE) technique. The SET algorithm showed superior performance in outdoor and laboratory image settings. ElSe algorithm performed well by applying edge filtering and selecting curved lines for ellipse fitting. PuRe algorithm reconstructed the pupil using multiple processing stages, achieving detection rates over 60%. PuReST algorithm improved upon PuRe by incorporating tracking methods. DeepEye, a deep convolutional neural network (DCNN), outperformed other algorithms by generating a circular mask to represent the pupil circle. The PDA algorithm, based on the circular Hough transform (CHT), achieved high detection rates in all datasets.

The algorithms were tested in challenging datasets with many forms of noise, such as images that were highly off-axis, eyelashes, reflections, bad illumination, contact lenses, mascara, shifted contact lenses, recording errors, and a pupil at the image border. Additionally, Table 6 provides comparisons among the included algorithms in this review. It is worth noting that algorithms not reviewed here, but that were tested, are not included in the table.

In Table 6, it is evident that the ExCuSe algorithm was the most frequently compared, as four of the seven algorithms (excluding itself) included a comparison with ExCuSe. Therefore, ExCuSe (Fuhl, Kübler, et al., 2015) serves as a reference for pupil detection.

Table 6 – Compared algorithms in the performance evaluation tests. Gray shading = no comparison done. White shading with an "x" = comparison done.

	Algorithms	ExCuSe	SET	ElSe	PuRe	PuReST	DeepEye	PDA	CNN-BPD	Total**
Reference System	ExCuSe									0
	SET									0
	ElSe	x	x							2
	PuRe	x								1
	PuReST									0
	DeepEye	x		x						2
	PDA	x								1
	CNN-BPD			x	x	x	x			4
	Total*	4	1	2	1	1	1	0	0	

*Number of comparisons with the algorithm; **Number of presented algorithms compared with the system in the study.

3.2.2. Software

The software category includes articles in which a software was defined for pupil size measurement as one of the extracted first order metrics, regardless of whether any hardware is used or indicated to be used with the described software. In this group, eight articles were included. Table 7 presents some characteristics of the software, such as the system name, first order metrics, pupil size methodology, kinds of images that can be used (videos or real-time images), and types of cameras suitable for real-time image methodology (external devices, head-mounted, or computer webcams).

Within this group the experimental set-up or the evaluation tests were performed on human participants (Bianchetti et al., 2013; Kassner et al., 2014; Santini, Fuhl, Geisler, et al., 2017), on human datasets (Kumar et al., 2019; Yiu et al., 2019; Zandi et al., 2021), on human and mouse subjects (Arvin et al., 2021) or on human and mouse datasets (Mazziotti et al., 2021).

Table 7 – Summary of studies that address systems to determine pupil size. ✔ = Applies; ⊗ = Does not apply; NR – Not relevant.

Reference	System	1st order metrics	Pupil size methodology	Images		For real-time images: cameras		
				Videos	Real-tine	External	Head-mounted	Webcam
(Bianchetti et al., 2013)	Blik	Gaze; Pupil size	1. RGB image converted to grayscale 2. Spatial filtering with Gaussian masks – noise reduction 3. Thresholding and segmentation 4. Area filtering and abnormal pupils' exclusion 5. Circularity filter is defined	⊗	✔	✔	⊗	⊗
(Kassner et al., 2014)	Pupil	Gaze; Pupil size	1. Conversion the image to grayscale. 2. Curve edge detection/filtering 3. Ellipse fitting	✔	✔	⊗	✔	⊗
(Santini, Fuhl, Geisler, et al., 2017)	EyeRecToo	Gaze; Pupil size	Else, the default, ExCuSe, Swirski, and Starburst can be used.	⊗	✔	⊗	✔	✔
(Yiu et al., 2019)	DeepVOG*	Gaze; Pupil size Blink	1. Probabilistic output prediction 2. Small gaps filling 3. Biggest perimeter extraction 4. Ellipse fitting	✔	⊗	NR	NR	NR
(Kumar et al., 2019)	Name not defined	Pupil size	1. Curve edge detection/filtering 2. Thresholding – binary image 3. Small gaps filling of the pupil	✔	⊗	NR	NR	NR
(Arvin et al., 2021)	EyeLoop	Pupil size	1. Engine: detects the pupil and corneal reflexes 2. Modules: import or extract data from the engine	✔	✔	✔	⊗	⊗
(Zandi et al., 2021)	PupilEXT	Pupil size	Starburst, Swirski, ExCuSe, Else, PuRe, and PuReST can be used.	✔	✔	✔	⊗	⊗
(Mazziotti et al., 2021)	MEYE*	Pupil size; Blink	1. Map of characteristics construction 2. Spatial resolution reducing 3. Probability map of pixels for pupil size determination	✔	✔	⊗	⊗	✔

*Convolutional Neural Network methods

Blick (2013) is a tabletop and low-cost pupillometer-eye tracker. It achieved a phase rate of up to 30 frames per second (fps) and a maximum error in pupil diameter determination of 0.05 mm. The algorithm involved image processing steps, including grayscale conversion, histogram equalization, noise reduction with Gaussian masks, thresholding, blob labeling, area filtering, polygon pupil approximation, and circularity filtering.

Pupil (Kassner et al., 2014) demonstrated an accuracy of 0.6 degrees under ideal conditions, with a precision of 0.08 degrees and a phase rate of up to 30 fps with 0.124 seconds of latency. It employed infrared light illumination and utilized the Canny edge filter, dark-pupil detection, reflection removal, ellipse fitting, and combinatorial search to obtain the pupil size.

EyeRecToo (Santini, Fuhl, Geisler, et al., 2017) is an open-source software for a real-time head-mounted eye tracking system. It integrated supervised and unsupervised calibration approaches and employed algorithms such as Else, ExCuSe, Swirski, and Starburst for pupil detection and ellipse fitting. With mean angular errors of 0.82 and 0.59 degrees, respectively, the higher latency for time-consuming operations observed was under 0.010 seconds. Its performance was obtained in an experiment with five human subjects.

DeepVOG (Yiu et al., 2019) utilized a fully connected convolutional neural network (FCNN) to generate a probabilistic output prediction. The algorithm involved steps such as binarization, morphological operations, contour extraction, and ellipse fitting to obtain the pupil contour and size. It got a mean dice coefficient of 0.96, a median Hausdorff distance of 2.8 pixels and a latency of 0.017 seconds. The Hausdorf distance (unit: [pixel]) calculates the greatest pupil contour distance between prediction and manual labelling, while the dice coefficient computes the overlap between predicted and manually marked regions of the pupil (range: 0 to 1, with 1.0 showing flawless overlap).

Kumar et al. (2019) presented a technique for evaluating pupillary light reflex using infrared and white light illumination. The algorithm involved pupil edge detection, binarization, infrared LED reflection removal, and pupil margin identification.

EyeLoop (Arvin et al., 2021) comprised an engine and modules for pupil and corneal reflex detection. It involved phases such as corneal reflection selection, gaussian kernel-based processing, contour detection, point matrix filtering, and fitting models to parameterize the shape.

PupilEXT (Zandi et al., 2021) offered a flexible platform for high-resolution pupillometry. It incorporated six different algorithms, including Starburst, Swirski, ExCuSe, Else, PuRe, and

PuReST, for pupil detection and ellipse fitting to obtain pupil size. This platform exhibited a measurement accuracy of 0.0059 mm on a 30 fps rate phase.

MEYE (Mazziotti et al., 2021) is a web application for real-time pupillometry that employed three CNN models: mini-UNet, DeepLabv3+/ResNet-50, and DeepLabv3+/Lite-MobileNet-V3-Small. In their comparison mini-UNet displayed the best dice coefficient, of 0.84. MEYE has a latency of 1 second. The algorithm used filters and selection layers to extract features from input data, followed by high-level reasoning layers for pupil size determination and blink detection.

Three of the softwares (Arvin et al., 2021; Bianchetti et al., 2013; Zandi et al., 2021) can be used with external cameras (including eye trackers), and one of them (Kassner et al., 2014) was designed for head-mounted cameras. Similarly, another one (Santini, Fuhl, Geisler, et al., 2017) can either use head-mounted cameras or the computer webcam. Lastly, MEYE was designed to operate with the webcam.

To measure pupil size, EyeRecToo and PupilEXT use some of the algorithms included in this review, such as ExCuse and ElSe. Additionally, Blik, the Kumar et al (2019) system, and Pupil use self-made algorithms. On the contrary, DeepVOG and MEYE use CNN methodologies.

4. Discussion

4.1. Relevence of pupil size for workload measurment

Results of this study showed that pupillometry can be used to determine mental workload in occupational settings, by replacing in some cases subjective measurements. In fact, self-rated mental effort has been the most widespread method for evaluating mental workload and continues to be widely used (Leppink & Pérez-Fuster, 2019). Questionnaires, such as NASA-TLX, are commonly employed subjective measures in the literature to assess mental workload (Derouin & Salway, 2018; Matthews et al., 2015). However, evaluating mental workload in real occupational settings can be challenging due to the arbitrary and often invasive nature of conventional measures (Matthews et al., 2015; Zheng et al., 2022). In this regard, eye tracking provides a non-intrusive method for monitoring mental activity and collecting data (Zheng et al., 2022). Pupil diameter has been frequently associated with measuring mental workload (Pauszek, 2023). In this review six out of nine studies have shown statistically significant correlations between subjective and objective measures of mental workload, suggesting that eye tracking indices can either replicate or substitute subjective methods, which are widely accepted for workload assessment.

However, it is important to note that while pupillometry has shown applicability in assessing workload, there have been discrepancies in the findings of some studies regarding the correlation between pupil dilation and questionnaires scores (Mazur et al., 2013; Mosaly et al., 2013; C. Wu et al., 2020). These discrepancies were related to various factors, which could be considered as limitations when using pupil size as a measure of cognitive workload in occupational settings. The lighting environment may be one of the reasons, in particular poor illumination conditions and rapid changes in illumination, since it affects pupil size obtention (Holmqvist et al., 2022). Another explanation for the lack of correlation is related to individual differences or the misconception of considering workload as a single latent concept, as proposed by Matthews et al. (2015). Other characteristics that need to be taken into consideration when interpreting the results is the sample size. In this regard, the studies that found no correlation between objective and subjective measures (Mazur et al., 2013; Mosaly et al., 2013; C. Wu et al., 2020) had the smallest samples, all with fewer than or equal to eight participants, which could also explain why no correlation was found.

To address the illumination limitation and overcome the challenges encountered in real-time images (Bozomitu et al., 2019), the detection rates of the reviewed algorithms were compared using both the ElSe and ExCuSe datasets, that include images with a wide range of challenges. Regarding bad illumination conditions, the reviewed algorithms PuRe, PuReST, and CNN-BPD perform equally, showing an 88 % detection rate. When comparing the performance of the algorithms in highly off-axis images, ExCuSe and PuReST algorithms have the best detection rate percentage. Thus, PuReST outperformed the others considering these two challenges. However, the ExCuSe and ElSe datasets comprehend more images and more challenges, and both CNN algorithms, DeepEye and CNN-BPD, had the highest detection rate of 87 %, outperforming the other algorithms, including PuReST, that presented 79 %. Thus, CNN-based algorithms seem to be the best methodology to detect and measure the pupil. It is important to note that despite the fact that both studies, from which detection rates were taken, were performed at different CPU powers, they both measured ElSe performance. It presented a deviation of 1,11 % when comparing both means, which was considered nonsignificant for this review. Therefore, all the detection rates were directly compared.

The algorithms presented used various methods to define and locate the pupil size, which is a first order data as reviewed elsewhere (Sharafi et al., 2020). LSFE is the most commonly used technique to fit an ellipse to the pupil. According to Bozomitu et al. (2019) LSFE is numerically stable, fast, and exact. The key disadvantage is that the identified ellipse may not be the best geometrical model of the eye pupil in the case of outlier points. This can happen due to noisy frames and nonuniform or variable illumination conditions. These scenarios can lead to large mistakes in pupil centre coordinate identification.

Furthermore, in the study by Wu et al. (2020), gaze entropy demonstrated a significant correlation, suggesting that increases in gaze entropy corresponded to higher perceived workload. This finding is crucial as it demonstrates that not only pupil size, but also other eye metrics should be utilized to assess workload more accurately (Pauszek, 2023).

Ergo, objective, non-invasive measures, such as pupil size with eye tracking technologies, can be used to assess workload in substitution for subjective, more invasive measures, such as NASA-TLX and the Paas scale. It is important to mention that they can also be used together. However, subjective measures are dependent on workers' ability to provide accurate descriptions, which can be lower or higher than reality, and objective measures are dependent on physiological behaviors, which are harder to forge (Matthews et al., 2015).

Regarding the four articles that did not assess subjective workload, Bhavsar et al. (2016) and Cabrera-Mino et al. (2019) found a statistical correlation between pupil size and workload. These emphasizes the usage of pupil size to measure workload. On the other hand, Sharma et al. (2021) and Naskrent et al. (2022) directly employed eye tracking metrics, including pupil size, to directly measure workload. In all of these studies, as well as those where a correlation between subjective and objective measures was to some extent described, pupil dilation consistently increased with workload enhancement. Accordingly, the same relationship was verified before in other studies such as Ahmad et al (Ahmad et al., 2020) and Liao et al (2021).

In view of the above, the majority of workload studies indicate an increase in pupil dilation with higher workload, accompanied by a decrease in pupil size variation (the pupil remained dilated) during periods of increased workload. In addition, some of the studies (Mazur et al., 2013; Mosaly et al., 2013; Othman & Romli, 2016), the performance of the workers was also evaluated and compared with workload. Majorly it was seen that performance declined with workload enhancement. Likewise, literature states that workload negatively and significantly affects performance (Aliyyah et al., 2021) . In some of the studies, not only performance but also experience was evaluated (Cabrera-Mino et al., 2019; Sharma et al., 2021; Szulewski et al., 2017). Mainly inexperienced or less experienced workers showed higher workloads. However, in Mazur et al (2013), the residents showed higher workloads on task-wise pooled data than faculty physicians, but no other significant differences were found regarding experience. In contrast, in Mosaly et al (2013), experience did not affect workload, and in Gao et al 2019), no significant differences were seen. These findings indicate that workload can be higher for experts or newcomers depending on the task or work type. So, the level of experience matters but differs in impact depending on the job or task evaluated. This goes in line with Zheng's findings that experienced, and novice workers show different patterns to achieve the same goal in manufacturing and logistics (Zheng et al., 2022).

The evaluation method for objective measurements varied a lot. Three studies used Tobii Glasses 2.0 at a sampling rate of 50 Hz. Five studies used other Tobii Technologies eye trackers with different sampling rates, from 30 Hz to 120 Hz. As stated above, two studies used VisionTrak Technologies, both with 60 Hz samples. Lastly, three other studies used SMI eye trackers with sampling rates of 20 Hz, 50 Hz, and 60 Hz. This wide range of sampling rates does not seem to affect the results. Additionally, two of the studies didn't provide sampling rate acquisition, and therefore no interpretation was made when comparing them to the other studies.

4.2. Relevance of pupil size for stress evaluation

In terms of stress, two studies (Bertilsson et al., 2020; Cabrera-Mino et al., 2019) evaluated it objectively, both consistently showing that pupil size increases with higher stress levels, which aligns with the literature (Pauszek, 2023). They both used SMI technologies with different sampling rates, which once again presents the possibility of sampling rate alterations not affecting the results. However, more analysis on sampling rates needs to be done for this to be certain. For instance, the same work group or type should be evaluated with the same eye tracker in the same experiment setup with different sampling rates for a more conclusive interpretation and comparison.

Studies about stress were too much less to the ones about workload. Well, there is no accepted gold standard for workload measurement, and it has been brought into question whether other variables, like stress, might be the ones being measured using subjective techniques (Szulewski et al., 2017). Therefore, to address this issue, objective measures may be more suitable (Matthews et al., 2015).

4.3. Relevance of pupil size for fatigue assessment

In respect to fatigue, only one study was included in this review. Nonetheless, in this study (Naeeri et al., 2021) pupil size decreased with fatigue augment. In fact, fatigue augmentation is accompanied by a decrease in pupil size, as demonstrated in previous studies (see, e.g., Hopstaken et al., 2015; Bertlisson et al., 2020). This shows that pupil size is an important metric to be applied in occupational safety, since it is one of the most important sources of human errors (Bhavsar et al. , 2016). Addicionally, it is important to note that fatigue is related to the time and in short time experiences the results can be not significative. In Zhang et al. (2018), the authors mentioned fatigue and concluded that because of their short experience, they could not assess the influence of fatigue. This remark is coherent with Naeeri et al. (2021), because fatigue increases with time, so short-term evaluations should not be appropriate to measure this risk factor. Naskrent et al. (2022) stated that long working shifts combined with high monotony (task repetition) can increase fatigue and also stated that stress affects the rate at which fatigue is developed. Likewise, in Wu et al (2020), the authors measured the percentage of eyelid closure (blink) and discussed, based on the literature, that it was linked to fatigue. This also shows other eye metrics can be used to assess fatigue.

Moreover, this review revealed that while many experiments used eye tracking technologies in laboratory settings, only a few studies evaluate cognitive workload, stress and fatigue in real-world field settings. These findings are consistent with another recent review conducted by Zheng et al. (2022). Furthermore, attention is another commonly analyzed risk factor related to occupational accidents; however, several studies that initially explored the relationship between attention and pupil size were excluded from this review due to a lack of defined workgroups, as they involved random participants in laboratory environments.

4.4. Quality analysis of phase 1 studies

Regarding the quality of the studies analysed for relevance of pupil size for OSH, significant limitations were observed in the included studies, particularly in terms of population representation, complete outcome data, confounder consideration, and intended exposure. On the overall score, all studies passed the screening questions and were included in the evaluation. In terms of appropriate measures, 100% of the studies received a positive score, indicating that they utilized appropriate measurement techniques. However, when considering the representation of the population, the studies received a negative score (No – N – or Can't Tell – CT) of 73.3%. This suggests that the included studies did not adequately capture the diverse characteristics of the target population. Only a small proportion of the studies (20%) provided comprehensive data on the outcomes being assessed, resulting in a low score for complete outcome data. Additionally, confounding factors, which could potentially influence the observed outcomes, were only partially considered, as indicated by a score of 60% for confounder consideration. Moreover, the intended exposure received a score of 6.7%, indicating that only a limited number of studies provided a clear and precise description of the exposure under investigation. Consequently, based on these scores, one study was evaluated as high quality, another as low quality, and the remaining studies were considered medium quality. The studies are mostly medium quality, with noticeable negative scores on the representation of the population. This is due to participants being volunteers, the lack of explanation for inclusion/exclusion criteria, or significant differences in group sizes. Another aspect that received unfavourable responses was the consideration of confounding circumstances. Several researchers included participants with varying degrees of experience/education, which is a potential confounder that was not adequately addressed in the study's quality assessment. The experiments were conducted in the same manner for all individuals, but control procedures such as stratification, regression, matching, standardization, or inverse probability weighting were not applied at the statistical level. These findings highlight the need to address these limitations in future research to ensure more robust and reliable findings.

4.5. Open source softwares to measure pupil size variations

The best interest of this review is to point out some software that can be used in occupational settings to assess the analysed risk factors in real time. DeepVOG and Kumar et al (2019) system cannot be used in real-time images but were included for contrast. The systems whose pupil size detection and measurement methodologies that were not CNN-based presented different evaluation data. Kumar et al (2019) system and EyeLoop do not provide maximum error in diameter, accuracy, or latency. Therefore, they were not included in the following considerations. The other systems were discussed depending on the comparable measures presented.

Blik has a maximum error in the diameter of the pupil of 0,05 mm, which is higher than PupilEXT's error of 0,0059 mm. This difference could be due to the fact that Blik was tested on a small sample of subjects in real time, while PupilEXT was tested on a dataset that had more subjects and was offline. Nevertheless, PupilEXT offers more than one algorithm for pupil detection and measurement, in contrast with Blik, which has only one algorithm. Thereby, PupilEXT could be used as a system for external camera inputs.

Pupil had 0,6 degrees of accuracy and 0,124 seconds of latency, while EyeRecToo had 0,82 degrees of accuracy and 0,010 seconds of latency. Thus, EyeRecToo outperformed Pupil. Since they both were evaluated in real-time images, the difference in the results probably arises from the fact that EyeRecToo has four different algorithms to detect and measure the pupil whereas Pupil has only one methodology. Also, the sample size is larger in Pupil (8 subjects) than in EyeRecToo (5 subjects), which could implicate less variations in EyeRecToo results. This way, EyeRecToo could be used as a system for head-mounted camera inputs.

Regarding the CNN methods, while DeepVOG had a 0,96 mean dice coefficient and a latency of 0,017 seconds, MEYE had 0,84 dice coefficient and a latency of 1 second. Comparing these results, DeepVOG showed the best performance, but it can only be used in offline images, which explains the best scores. MEYE had more challenging images because they used human and mouse datasets to test it.

In conclusion, since CNN methods have better performance and DeepVOG cannot be used in real-time images, MEYE should be the best open-source system to be used in occupational settings. Once MEYE uses the webcam to acquire the inputs, it should be very useful when applied to office workers and every worker who spends more than three hours in front of the computer. For other jobs that do not require a computer but have a designated work environment, such as production line operators, forklift truckers, pilots, and any kind of driver, a system with external cameras (PupilEXT) should be the best choice. Not discarding a system with head-mounted cameras (EyeRecToo) for this type of worker. Finally, for workers that do not have a designated spot to work, such as high-rise building window cleaners, construction workers, and many more, the system with head-mounted cameras (EyeRecToo) seems more appropriate. It is better to use head-mounted cameras that are integrated into glasses because they are more comfortable, less intrusive, and can be worn with other technologies (Pauszek, 2023).

Other systems for detection and measuring pupil size could exist and outperform the ones here presented, since the review methodology used did not search the keywords in all fields for phase 2. This method was chosen because it was not the main goal of this review to extensively assess all the methodologies existent. Instead, this review aimed to provide some options for user-friendly software that could be further implemented in the worker environment to improve OSH.

5. Limitations

Regarding the relevance of pupil size for OSH, only one study was reviewed for fatigue. However, several articles on fatigue and another mental and cognitive conditions, such as attention, were sorted and discarded. This highlights the scale of studies that were excluded because they were conducted in laboratory settings rather than in real working settings.

When considering open-source methodologies for pupil size analysis, it is worth noting that runtime comparisons could have provided valuable insights into algorithm performance. However, conducting such comparisons would have required evaluating all algorithms using a consistent CPU and programming language across studies. Unfortunately, none of the reviewed articles compared all the algorithms or employed the same versions of algorithms in a uniform programming language. Consequently, the discussion of runtime as a measure was not feasible within the scope of this analysis.

6. Conclusion

This systematic review aimed to understand whether pupil size could be used as an indicator for occupational risks and which technologies could measure it and be implemented in the worker's environment. The findings from the reviewed studies demonstrate that eye tracking metrics, including pupil size, can be utilized to assess workload, stress, and fatigue. It became clear that pupil size increases with workload and stress enhancement and that the trend is for pupil size to decrease upon fatigue augmentation.

Despite the many methodologies that exist, CNN-based algorithms appear to be the most effective technique for identifying and measuring the pupil. Thus, MEYE, which is CNN-based, was the best-performing system considering real-time image acquisition and therefore the one recommended.

Notwithstanding, this review pointed out that pupil size is not the only eye metric that could be used as indicators of occupational risks. Therefore, eye tracking systems with many metrics are useful to maintain or improve OSH. Consequently, performance and production should improve as well. Additionally, as the analysed studies encompassed diverse workgroups, the insights gained can be broadly applied in different occupational settings. Therefore, implementing these measures in the workplace, such as in forklifts, trucks, computers, or safety equipment, can be useful to effectively prevent accidents and enhance workers' health, safety and performance. Through timely alerts, both workers and responsible parties can be made aware of their mental and cognitive state, facilitating proactive intervention.

References

Ahmad, M. I., Keller, I., Robb, D. A., & Lohan, K. S. (2020). A framework to estimate cognitive load using physiological data. *Personal and Ubiquitous Computing*. https://doi.org/10.1007/s00779-020-01455-7

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

Aliyyah, N., Prasetyo, I., Rusdiyanto, R., Endarti, E. W., Mardiana, F., Winarko, R., Chamariyah, C., Mulyani, S., Grahani, F. O., Rochman, A. S. U., Kalbuana, N., Hidayat, W., & Tjaraka, H. (2021). What Affects Employee Performance Through Work Motivation? *Journal of Management Information and Decision Science*, *24*, 1–14.

Arvin, S., Rasmussen, R. N., & Yonehara, K. (2021). EyeLoop: An Open-Source System for High-Speed, Closed-Loop Eye-Tracking. *Frontiers in Cellular Neuroscience*, *15*. https://doi.org/10.3389/fncel.2021.779628

Bachurina, V., Sushchinskaya, S., Sharaev, M., Burnaev, E., & Arsalidou, M. (2022). A machine learning investigation of factors that contribute to predicting cognitive performance: Difficulty level, reaction time and eye-movements. *Decision Support Systems*, *155*, 113713. https://doi.org/10.1016/j.dss.2021.113713

Bertilsson, J., Niehorster, D., Fredriksson, P., Dahl, M., Granér, S., Fredriksson, O., Mårtensson, J., Magnusson, M., Fransson, P., & Nyström, M. (2020). Towards systematic and objective evaluation of police officer performance in stressful situations. *Police Practice and Research*, *21*(6), 655–669. https://doi.org/10.1080/15614263.2019.1666006

Bhavsar, P., Srinivasan, B., & Srinivasan, R. (2016). Pupillometry Based Real-Time Monitoring of Operator’s Cognitive Workload To Prevent Human Error during Abnormal Situations. *Industrial & Engineering Chemistry Research*, *55*(12), 3372–3382. https://doi.org/10.1021/acs.iecr.5b03685

Bianchetti, A., Perez, L. I., & Comastri, S. A. (2013). *Development of a low cost pupillometer-eyetracker and applications* (M. F. P. C. Martins Costa, Ed.; p. 8785DA). https://doi.org/10.1117/12.2021054

Binda, P., & Murray, S. O. (2015). Spatial attention increases the pupillary response to light changes. *Journal of Vision*, *15*(2), 1–1. https://doi.org/10.1167/15.2.1

Binda, P., Pereverzeva, M., & Murray, S. O. (2014). Pupil size reflects the focus of feature-based attention. *Journal of Neurophysiology*, *112*(12), 3046–3052. https://doi.org/10.1152/jn.00502.2014

Blehm, C., Vishnu, S., Khattak, A., Mitra, S., & Yee, R. W. (2005). Computer Vision Syndrome: A Review. *Survey of Ophthalmology*, *50*(3), 253–262. https://doi.org/10.1016/j.survophthal.2005.02.008

Boadi-Kusi, S. B., Abu, S. L., Acheampong, G. O., Adueming, P. O.-W., & Abu, E. K. (2020). Association between Poor Ergophthalmologic Practices and Computer Vision Syndrome among University Administrative Staff in Ghana. *Journal of Environmental and Public Health*, *2020*, 1–8. https://doi.org/10.1155/2020/7516357

Bonsang, E., & Caroli, E. (2021). Cognitive Load and Occupational Injuries. *Industrial Relations: A Journal of Economy and Society*, *60*(2), 219–242. https://doi.org/10.1111/irel.12277

Bonteanu, P., Bozomitu, R. G., Cracan, A., & Bonteanu, G. (2019). A New Pupil Detection Algorithm Based on Circular Hough Transform Approaches. *2019 IEEE 25th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, 260–263. https://doi.org/10.1109/SIITME47687.2019.8990887

Bozomitu, R. G., Păsărică, A., Tărniceriu, D., & Rotariu, C. (2019). Development of an Eye Tracking-Based Human-Computer Interface for Real-Time Applications. *Sensors*, *19*(16), 3630. https://doi.org/10.3390/s19163630

Cabrera-Mino, C., Shinnick, M. A., & Moye, S. (2019). Task-Evoked Pupillary Responses in Nursing Simulation as an Indicator of Stress and Cognitive Load. *Clinical Simulation in Nursing*, *31*, 21–27. https://doi.org/10.1016/j.ecns.2019.03.009

Cao, Y., Ding, Y., Proctor, R. W., Duffy, V. G., Liu, Y., & Zhang, X. (2021). Detecting users' usage intentions for websites employing deep learning on eye-tracking data. *Information Technology and Management*, *22*(4), 281–292. https://doi.org/10.1007/s10799-021-00336-6

Coronel-Ocampos, J., Gómez, J., Gómez, A., Quiroga-Castañeda, P. P., & Valladares-Garrido, M. J. (2022). Computer Visual Syndrome in Medical Students From a Private University in Paraguay: A Survey Study. *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.935405

Couceiro, R., Duarte, G., Duraes, J., Castelhano, J., Duarte, C., Teixeira, C., Castelo Branco, M., Carvalho, P., & Madeira, H. (2019). Pupillography as Indicator of Programmers' Mental Effort and Cognitive Overload. *2019 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 638–644. https://doi.org/10.1109/DSN.2019.00069

Dalveren, G. G., & Cagiltay, N. E. (2018). Using Eye-Movement Events to Determine the Mental Workload of Surgical Residents. *Journal of Eye Movement Research*, *11*(4). https://doi.org/10.16910/jemr.11.4.3

Dalveren, G. G., Cagiltay, N. E., Ozcelik, E., & Maras, H. (2018). Insights From Pupil Size to Mental Workload of Surgical Residents: Feasibility of an Educational Computer-Based Surgical Simulation Environment (ECE) Considering the Hand Condition. *Surgical Innovation*, *25*(6), 616–624. https://doi.org/10.1177/1553350618800078

Derouin, A., & Salway, A. (2018). Enhancing Workload Assessments for Validation Activities Associated with DBA and BDBA Scenarios. *Nuclear Technology*, *201*(2), 165–173. https://doi.org/10.1080/00295450.2017.1413922

Eivazi, S., Santini, T., Keshavarzi, A., Kübler, T., & Mazzei, A. (2019). Improving real-time CNN-based pupil detection through domain-specific data augmentation. *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 1–6. https://doi.org/10.1145/3314111.3319914

European Commission. (2021). *Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions, Eu Strategic Framework On Health And Safety At Work 2021-2027, Occupational Safety And Health In A Changing World Of Work*. https://ec.europa.eu/info/strategy/priorities-2019-2024/economy-works-people/jobs-growth-and-

Fuhl, W., Kübler, T., Sippel, K., Rosenstiel, W., & Kasneci, E. (2015). ExCuSe: Robust Pupil Detection in Real-World Scenarios. *Computer Analysis of Images and Patterns*, 39–51. https://doi.org/10.1007/978-3-319-23192-1_4

Fuhl, W., Santini, T. C., Kuebler, T., & Kasneci, E. (2015). *ElSe: Ellipse Selection for Robust Pupil Detection in Real-World Environments*.

Gao, J., Ko, A., Yabe, Y., Goodale, M. A., & Chen, J. (2020). Pupil size is modulated by the size of equal-luminance gratings. *Journal of Vision*, *20*(8), 4. https://doi.org/10.1167/jov.20.8.4

Gao, J., Liu, S., Feng, Q., Zhang, X., Jiang, M., Wang, L., Zhang, J., & Zhang, Q. (2019). Subjective and Objective Quantification of the Effect of Distraction on Physician's Workload and

Performance During Simulated Laparoscopic Surgery. *Medical Science Monitor*, *25*, 3127–3132. https://doi.org/10.12659/MSM.914635

Gomes, C., & Preto, S. (2015). Should the light be static or dynamic? *6th International Conference on Applied Human Factors and Ergonomics*.

Graff, T. C., Luke, S. G., & Birmingham, W. C. (2019). Supportive hand-holding attenuates pupillary responses to stress in adult couples. *PLOS ONE*, *14*(2), e0212703. https://doi.org/10.1371/journal.pone.0212703

Holmqvist, K., Örbom, S. L., Hooge, I. T. C., Niehorster, D. C., Alexander, R. G., Andersson, R., Benjamins, J. S., Blignaut, P., Brouwer, A.-M., Chuang, L. L., Dalrymple, K. A., Drieghe, D., Dunn, M. J., Ettinger, U., Fiedler, S., Foulsham, T., van der Geest, J. N., Hansen, D. W., Hutton, S. B., … Hessels, R. S. (2022). Eye tracking: empirical foundations for a minimal reporting guideline. *Behavior Research Methods*, *55*(1), 364–416. https://doi.org/10.3758/s13428-021-01762-8

Hong, Q. N., Gonzalez-Reyes, A., & Pluye, P. (2018). Improving the usefulness of a tool for appraising the quality of qualitative, quantitative and mixed methods studies, the Mixed Methods Appraisal Tool (MMAT). *Journal of Evaluation in Clinical Practice*, *24*(3), 459–467. https://doi.org/10.1111/jep.12884

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015). A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology*, *52*(3), 305–315. https://doi.org/10.1111/psyp.12339

Iqbal, M. U., Srinivasan, B., & Srinivasan, R. (2018). *Towards Obviating Human Errors in Real-time through Eye Tracking* (pp. 1189–1194). https://doi.org/10.1016/B978-0-444-64235-6.50207-2

Javadi, A.-H., Hakimi, Z., Barati, M., Walsh, V., & Tcheang, L. (2015). SET: a pupil detection method using sinusoidal approximation. *Frontiers in Neuroengineering*, *8*. https://doi.org/10.3389/fneng.2015.00004

Kalakoski, V., Selinheimo, S., Valtonen, T., Turunen, J., Käpykangas, S., Ylisassi, H., Toivio, P., Järnefelt, H., Hannonen, H., & Paajanen, T. (2020). Effects of a cognitive ergonomics workplace intervention (CogErg) on cognitive strain and well-being: a cluster-randomized controlled trial. A study protocol. *BMC Psychology*, *8*(1), 1. https://doi.org/10.1186/s40359-019-0349-1

Kassner, M., Patera, W., & Bulling, A. (2014). *Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction*.

Kelbsch, C., Strasser, T., Chen, Y., Feigl, B., Gamlin, P. D., Kardon, R., Peters, T., Roecklein, K. A., Steinhauer, S. R., Szabadi, E., Zele, A. J., Wilhelm, H., & Wilhelm, B. J. (2019). Standards in Pupillography. *Frontiers in Neurology*, *10*. https://doi.org/10.3389/fneur.2019.00129

Kret, M. E., & Sjak-Shie, E. E. (2019). Preprocessing pupil size data: Guidelines and code. *Behavior Research Methods*, *51*(3), 1336–1342. https://doi.org/10.3758/s13428-018-1075-y

Krol, M., & Krol, M. (2017). A novel approach to studying strategic decisions with eye-tracking and machine learning. *Judgment and Decision Making*, *12*(6), 596–609. https://doi.org/10.1017/S1930297500006720

Kumar, A. S., Padmavathi, R., Maruthy, K., Sowjanya, B., & Kumar, K. M. (2019). An Innovative Technique to Evaluate Quantitative Pupillary Light Reflex by Dynamic Pupillometry using Infrared Videography. *JOURNAL OF CLINICAL AND DIAGNOSTIC RESEARCH*. https://doi.org/10.7860/JCDR/2019/41051.12805

Lapa, I., Ferreira, S., Mateus, C., Rocha, N., & Rodrigues, M. A. (2023). Real-Time Blink Detection as an Indicator of Computer Vision Syndrome in Real-Life Settings: An Exploratory Study. *International Journal of Environmental Research and Public Health*, *20*(5), 4569. https://doi.org/10.3390/ijerph20054569

Leppink, J., & Pérez-Fuster, P. (2019). Mental Effort, Workload, Time on Task, and Certainty: Beyond Linear Models. *Educational Psychology Review*, *31*(2), 421–438. https://doi.org/10.1007/s10648-018-09460-2

Liao, P.-C., Sun, X., & Zhang, D. (2021). A multimodal study to measure the cognitive demands of hazard recognition in construction workplaces. *Safety Science*, *133*, 105010. https://doi.org/10.1016/j.ssci.2020.105010

Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2015). The Psychometrics of Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(1), 125–143. https://doi.org/10.1177/0018720814539505

Mazur, L. M., Mosaly, P. R., Hoyle, L. M., Jones, E. L., & Marks, L. B. (2013). Subjective and objective quantification of physician's workload and performance during radiation therapy planning tasks. *Practical Radiation Oncology*, *3*(4), e171–e177. https://doi.org/10.1016/j.prro.2013.01.001

Mazziotti, R., Carrara, F., Viglione, A., Lupori, L., Lo Verde, L., Benedetto, A., Ricci, G., Sagona, G., Amato, G., & Pizzorusso, T. (2021). MEYE: Web App for Translational and Real-Time Pupillometry. *Eneuro*, *8*(5), ENEURO.0122-21.2021. https://doi.org/10.1523/ENEURO.0122-21.2021

Mosaly, P. R., Mazur, L. M., Jones, E. L., Hoyle, L., Zagar, T., Chera, B. S., & Marks, L. B. (2013). Quantifying the impact of cross coverage on physician's workload and performance in radiation oncology. *Practical Radiation Oncology*, *3*(4), e179–e186. https://doi.org/10.1016/j.prro.2013.02.007

Munshi, S., Varghese, A., & Dhar-Munshi, S. (2017). Computer vision syndrome-A common cause of unexplained visual symptoms in the modern era. *International Journal of Clinical Practice*, *71*(7), e12962. https://doi.org/10.1111/ijcp.12962

Naeeri, S., Kang, Z., Mandal, S., & Kim, K. (2021). Multimodal Analysis of Eye Movements and Fatigue in a Simulated Glass Cockpit Environment. *Aerospace*, *8*(10), 283. https://doi.org/10.3390/aerospace8100283

Naskrent, B., Grzywiński, W., Polowy, K., Tomczak, A., & Jelonek, T. (2022). Eye-Tracking in Assessment of the Mental Workload of Harvester Operators. *International Journal of Environmental Research and Public Health*, *19*(9), 5241. https://doi.org/10.3390/ijerph19095241

Nguyen, K. T., Liang, W.-K., Juan, C.-H., & Wang, C.-A. (2022). Time-frequency analysis of pupil size modulated by global luminance, arousal, and saccade preparation signals using Hilbert-Huang transform. *International Journal of Psychophysiology*, *176*, 89–99. https://doi.org/10.1016/j.ijpsycho.2022.03.011

Nurçin, F. V., Imanov, E., Işın, A., & Ozsahin, D. U. (2017). Lie detection on pupil size by back propagation neural network. *Procedia Computer Science*, *120*, 417–421. https://doi.org/10.1016/j.procs.2017.11.258

Othman, N., & Romli, F. I. (2016). Mental Workload Evaluation of Pilots Using Pupil Dilation. *International Review of Aerospace Engineering (IREASE)*, *9*(3), 80. https://doi.org/10.15866/irease.v9i3.9541

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*, *372*. https://doi.org/10.1136/BMJ.N71

Pan, J., Klímová, M., McGuire, J. T., & Ling, S. (2022). Arousal-based pupil modulation is dictated by luminance. *Scientific Reports*, *12*(1), 1390. https://doi.org/10.1038/s41598-022-05280-1

Pauszek, J. R. (2023). An introduction to eye tracking in human factors healthcare research and medical device testing. *Human Factors in Healthcare*, *3*, 100031. https://doi.org/10.1016/j.hfh.2022.100031

Randolph, S. A. (2017). Computer Vision Syndrome. *Workplace Health & Safety*, *65*(7), 328–328. https://doi.org/10.1177/2165079917712727

Reddy, S. C., Low, C., Lim, Y., Low, L., Mardina, F., & Nursaleha, M. (2013). Computer vision syndrome: a study of knowledge and practices in university students. *Nepalese Journal of Ophthalmology*, *5*(2), 161–168. https://doi.org/10.3126/nepjoph.v5i2.8707

Santini, T., Fuhl, W., Geisler, D., & Kasneci, E. (2017). EyeRecToo: Open-source Software for Real-time Pervasive Head-mounted Eye Tracking. *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 96–101. https://doi.org/10.5220/0006224700960101

Santini, T., Fuhl, W., & Kasneci, E. (2017). *PuRe: Robust pupil detection for real-time pervasive eye tracking*. https://doi.org/10.1016/j.cviu.2018.02.002

Santini, T., Fuhl, W., & Kasneci, E. (2018). PuReST. *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, 1–5. https://doi.org/10.1145/3204493.3204578

Seguí, M. del M., Cabrero-García, J., Crespo, A., Verdú, J., & Ronda, E. (2015). A reliable and valid questionnaire was developed to measure computer vision syndrome at the workplace. *Journal of Clinical Epidemiology*, *68*(6), 662–673. https://doi.org/10.1016/j.jclinepi.2015.01.015

Sharafi, Z., Sharif, B., Guéhéneuc, Y.-G., Begel, A., Bednarik, R., & Crosby, M. (2020). A practical guide on conducting eye tracking studies in software engineering. *Empirical Software Engineering*, *25*(5), 3128–3174. https://doi.org/10.1007/s10664-020-09829-4

Sharma, H., Drukker, L., Papageorghiou, A. T., & Noble, J. A. (2021). Machine learning-based analysis of operator pupillary response to assess cognitive workload in clinical ultrasound imaging. *Computers in Biology and Medicine*, *135*, 104589. https://doi.org/10.1016/j.compbiomed.2021.104589

Singh, S., McGuinness, M. B., Anderson, A. J., & Downie, L. E. (2022). Interventions for the Management of Computer Vision Syndrome. *Ophthalmology*, *129*(10), 1192–1215. https://doi.org/10.1016/j.ophtha.2022.05.009

Szulewski, A., Gegenfurtner, A., Howes, D. W., Sivilotti, M. L. A., & van Merriënboer, J. J. G. (2017). Measuring physician cognitive load: validity evidence for a physiologic and a psychometric tool. *Advances in Health Sciences Education*, *22*(4), 951–968. https://doi.org/10.1007/s10459-016-9725-2

Vera-Olmos, F. J., Pardo, E., Melero, H., & Malpica, N. (2018). DeepEye: Deep convolutional network for pupil detection in real environments. *Integrated Computer-Aided Engineering*, *26*(1), 85–95. https://doi.org/10.3233/ICA-180584

Wangwiwattana, C., Ding, X., & Larson, E. C. (2018). PupilNet, Measuring Task Evoked Pupillary Response using Commodity RGB Tablet Cameras. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(4), 1–26. https://doi.org/10.1145/3161164

Wu, C., Cha, J., Sulek, J., Zhou, T., Sundaram, C. P., Wachs, J., & Yu, D. (2020). Eye-Tracking Metrics Predict Perceived Workload in Robotic Surgical Skills Training. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *62*(8), 1365–1386. https://doi.org/10.1177/0018720819874544

Wu, Y., Liu, Z., Jia, M., Tran, C. C., & Yan, S. (2019). Using Artificial Neural Networks for Predicting Mental Workload in Nuclear Power Plants Based on Eye Tracking. *Nuclear Technology*, *206*(1), 94–106. https://doi.org/10.1080/00295450.2019.1620055

Yan, Z., Hu, L., Chen, H., & Lu, F. (2008). Computer Vision Syndrome: A widely spreading but largely unknown epidemic among computer users. *Computers in Human Behavior*, *24*(5), 2026–2042. https://doi.org/10.1016/j.chb.2007.09.004

Yiu, Y.-H., Aboulatta, M., Raiser, T., Ophey, L., Flanagin, V. L., zu Eulenburg, P., & Ahmadi, S.-A. (2019). DeepVOG: Open-source pupil segmentation and gaze estimation in neuroscience using deep learning. *Journal of Neuroscience Methods*, *324*, 108307. https://doi.org/10.1016/j.jneumeth.2019.05.016

Zandi, B., Lode, M., Herzog, A., Sakas, G., & Khanh, T. Q. (2021). PupilEXT: Flexible Open-Source Platform for High-Resolution Pupillometry in Vision Research. *Frontiers in Neuroscience*, *15*. https://doi.org/10.3389/fnins.2021.676220

Zayed, H. A. M., Saied, S. M., Younis, E. A., & Atlam, S. A. (2021). Digital eye strain: prevalence and associated factors among information technology professionals, Egypt. *Environmental Science and Pollution Research*, *28*(20), 25187–25195. https://doi.org/10.1007/s11356-021-12454-3

Zenbaba, D., Sahiledengle, B., Bonsa, M., Tekalegn, Y., Azanaw, J., & Kumar Chattu, V. (2021). Prevalence of Computer Vision Syndrome and Associated Factors among Instructors in Ethiopian Universities: A Web-Based Cross-Sectional Study. *The Scientific World Journal*, *2021*, 1–8. https://doi.org/10.1155/2021/3384332

Zhang, J., Liu, S., Feng, Q., Gao, J., Cheng, J., Jiang, M., Lan, Y., & Zhang, Q. (2018). Ergonomic Assessment of the Mental Workload Confronted by Surgeons during Laparoscopic Surgery. *The American Surgeon*, *84*(9), 1538–1543. https://doi.org/10.1177/000313481808400964

Zheng, T., Glock, C. H., & Grosse, E. H. (2022). Opportunities for using eye tracking technology in manufacturing and logistics: Systematic literature review and research agenda. *Computers & Industrial Engineering*, *171*, 108444. https://doi.org/10.1016/j.cie.2022.108444

Appendixes

Appendix 1 – Quality assessment result for phase 1 with quantitative non-randomized studies sector from mixed methods quality assessment tool.

	References	S1	S2	3.1	3.2	3.3	3.4	3.5	Quality
Workload	(Mazur et al., 2013)	Y	Y	N	Y	Y	N	Y	Medium
	(Mosaly et al., 2013)	Y	Y	N	Y	Y	N	Y	Medium
	(Bhavsar et al., 2016)	Y	Y	CT	Y	Y	NA	Y	Medium
	(Othman & Romli, 2016)	Y	Y	N	Y	Y	NA	Y	Medium
	(Szulewski et al., 2017)	Y	Y	CT	Y	Y	N	Y	Medium
	(Zhang et al., 2018)	Y	Y	CT	Y	Y	N	Y	Medium
	(Couceiro et al., 2019)	Y	Y	N	Y	Y	N	Y	Medium
	(Gao et al., 2019)	Y	Y	N	Y	Y	NA	Y	Medium
	(Y. Wu et al., 2019)	Y	Y	N	Y	Y	N	Y	Medium
	(C. Wu et al., 2020)	Y	Y	N	Y	N	NA	N	Low
	(Sharma et al., 2021)	Y	Y	N	Y	Y	N	Y	Medium
	(Naskrent et al., 2022)	Y	Y	Y	Y	Y	NA	Y	High
Stress	(Cabrera-Mino et al., 2019)	Y	Y	Y	Y	N	N	Y	Medium
	(Bertilsson et al., 2020)	Y	Y	Y	Y	N	NA	Y	Medium
Fatigue	(Naeeri et al., 2021)	Y	Y	Y	Y	Y	N	Y	Medium
Total of negative (N and CT) answers		0	0	11	0	3	9	1	

S1. Are there clear research questions?; S2. Do the collected data allow to address the research questions?; 3.1. Are the participants representative of the target population?; 3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?; 3.3. Are there complete outcome data?; 3.4. Are the confounders accounted for in the design and analysis?; 3.5. During the study period, is the intervention administered (or exposure occurred) as intended?
Y – Yes; N– No; CT – Can't Tell; NA – Not applicable.
Quality rating (Considering 3.1 to 3.5): High: 0 N or CT; Medium: 1 or 2 N or CT; Low: 3 or more N or CT

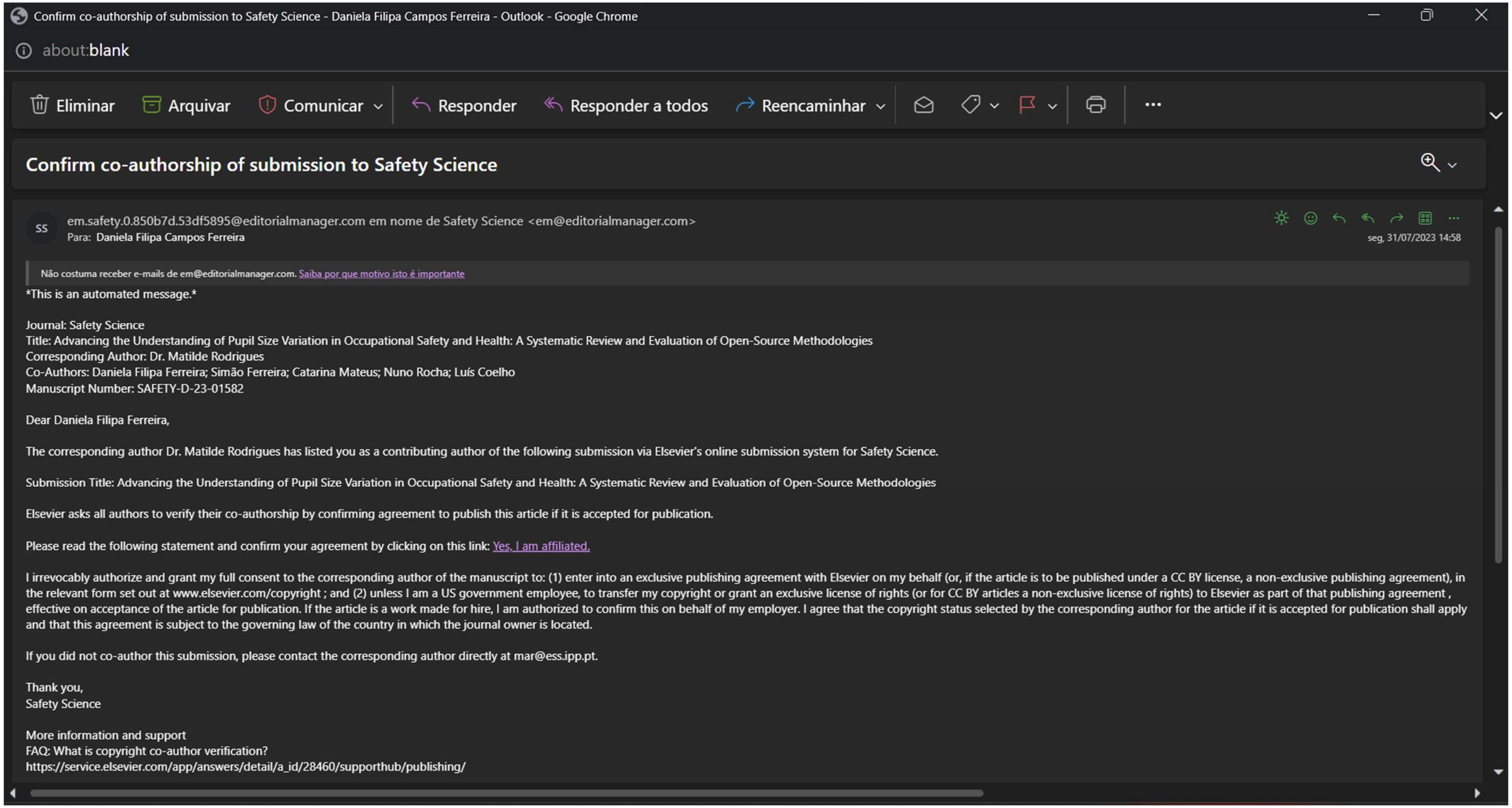
Appendix 2 – Detection rates with 5-pixel errors measured for ExCuSe (I-XVII) and ElSe (XVIII-XXIV) datasets.

	Algorithms							
Dataset	**SET (%) [1]**	**ExCuSe (%) [1]**	**ElSe (%) [1]**	**ElSe (%) [2]**	**PuRe (%) [2]**	**PuReST (%) [2]**	**DeepEye (%) [2]**	**CNN-BPD (%) [2]**
I	10,27	70,95	85,52	86	87	89	86	**90**
II	43,76	34,26	65,35	65	29	48	83	**88**
III	12,23	39,44	63,6	64	73	77	**93**	88
IV	4,03	81,58	83,24	83	89	90	**93**	**93**
V	18,08	77,28	84,87	85	87	87	**97**	**97**
VI	10,3	53,18	77,52	78	89	91	93	**94**
VII	2,19	46,91	59,51	60	68	73	**84**	81
VIII	36,67	56,83	68,41	68	54	60	87	**89**
IX	10,2	74,6	86,72	87	91	91	**92**	91
X	57,62	79,76	78,93	79	90	90	92	**94**
XI	23,51	56,49	75,27	75	88	87	94	**97**
XII	56,11	79,2	79,39	79	**88**	**88**	85	**88**
XIII	33,4	70,26	73,52	74	**85**	80	79	83
XIV	46,27	57,57	84,22	84	88	89	96	**97**
XV	38,29	52,34	57,3	57	62	71	**89**	77
XVI	57,17	49,49	59,95	60	79	66	82	**87**
XVII	91,04	77,99	89,55	90	95	95	95	**96**
XVIII	1,32	22,24	50,86	57	68	69	74	75
XIX	4,75	26,45	33,04	33	48	53	**78**	46
XX	3,2	52,37	67,9	78	83	86	**92**	89
XXI	2,29	43,54	41,47	47	70	81	88	**89**
XXII	1,91	27,93	48,98	53	62	72	**80**	76
XXIII	55,43	93,86	94,34	94	97	93	96	**100**
XXIV	0,94	45,21	52,97	53	60	65	55	**73**
Mean	25,87	57,07	69,27	70,38	76,25	78,79	**86,79**	86,58

[1] Results from (Fuhl, Santini, et al., 2015)

[2] Results from (Eivazi et al., 2019)

Appendix 3 – Proof of submission.



Confirm co-authorship of submission to Safety Science - Daniela Filipa Campos Ferreira - Outlook - Google Chrome

about:blank

Eliminar | Arquivar | Comunicar | Responder | Responder a todos | Reencaminhar

Confirm co-authorship of submission to Safety Science

SS

em.safety.0.850b7d.53df5895@editorialmanager.com em nome de Safety Science <em@editorialmanager.com>

Para: Daniela Filipa Campos Ferreira

seg, 31/07/2023 14:58

Não costuma receber e-mails de em@editorialmanager.com. Saiba por que motivo isto é importante

This is an automated message.

Journal: Safety Science
Title: Advancing the Understanding of Pupil Size Variation in Occupational Safety and Health: A Systematic Review and Evaluation of Open-Source Methodologies
Corresponding Author: Dr. Matilde Rodrigues
Co-Authors: Daniela Filipa Ferreira; Simão Ferreira; Catarina Mateus; Nuno Rocha; Luís Coelho
Manuscript Number: SAFETY-D-23-01582

Dear Daniela Filipa Ferreira,

The corresponding author Dr. Matilde Rodrigues has listed you as a contributing author of the following submission via Elsevier's online submission system for Safety Science.

Submission Title: Advancing the Understanding of Pupil Size Variation in Occupational Safety and Health: A Systematic Review and Evaluation of Open-Source Methodologies

Elsevier asks all authors to verify their co-authorship by confirming agreement to publish this article if it is accepted for publication.

Please read the following statement and confirm your agreement by clicking on this link: Yes, I am affiliated.

I irrevocably authorize and grant my full consent to the corresponding author of the manuscript to: (1) enter into an exclusive publishing agreement with Elsevier on my behalf (or, if the article is to be published under a CC BY license, a non-exclusive publishing agreement), in the relevant form set out at www.elsevier.com/copyright ; and (2) unless I am a US government employee, to transfer my copyright or grant an exclusive license of rights (or for CC BY articles a non-exclusive license of rights) to Elsevier as part of that publishing agreement , effective on acceptance of the article for publication. If the article is a work made for hire, I am authorized to confirm this on behalf of my employer. I agree that the copyright status selected by the corresponding author for the article if it is accepted for publication shall apply and that this agreement is subject to the governing law of the country in which the journal owner is located.

If you did not co-author this submission, please contact the corresponding author directly at mar@ess.ipp.pt.

Thank you,
Safety Science

More information and support
FAQ: What is copyright co-author verification?
https://service.elsevier.com/app/answers/detail/a_id/28460/supporthub/publishing/