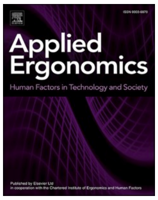
Applied Ergonomics 92 (2021) 103342

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: http://www.elsevier.com/locate/apergo



The Index of Cognitive Activity - Eligibility for task-evoked informational strain and robustness towards visual influences

Julia N. Czerniak [a,*], Nikolas Schierhorst [a], Valeria Villani [b], Lorenzo Sabattini [b], Christopher Brandl [a], Alexander Mertens [a], Maximilian Schwalm [c], Verena Nitsch [a]

[a] *Institute of Industrial Engineering and Ergonomics (IAW), RWTH Aachen University, Aachen, Germany*
[b] *Department of Sciences and Methods for Engineering (DISMI), University of Modena and Reggio Emilia, Reggio Emilia, Italy*
[c] *Institute of Highway Engineering (ISAC), RWTH Aachen University, Aachen, Germany*

ARTICLE INFO

Keywords:
Index of Cognitive Activity (ICA)
Visual stimuli
Workload

ABSTRACT

Various researchers have proposed pupillometric indicators to assess a person's cognitive strain. However, to distinguish the variation of pupil light response from psychosensory pupil response in experimental field conditions is a challenge. The Index of Cognitive Activity (ICA) addresses this problem by wavelet separation. This research investigates the ICA's sensitivity for multiple level task-evoked cognitive activity and visual influences concerning informational work tasks. Objective and subjective measures assessed cognitive strain of participants ($N = 22$) during various tasks. In a first experiment, mental arithmetic tasks were used to induce different levels of cognitive activity. In a second experiment, influences of screen polarity and presentation of information were investigated ($N = 18$). The results indicate that eye metrics are rarely sensitive to slight variations in task difficulty. Moreover, the ICA is likely to be sensitive towards constant screen illumination and shows tendencies regarding changes in displayed information. Possible ramifications for the objective assessment of cognitive strain are discussed.

1. Introduction

Fulfilling informational work tasks requires effort based on a certain amount of available cognitive capacity. These cognitive resources are limited and specifically have been implicated in showing arousal, in literature (Kahneman, 1973). Cognitive effort in combination with limited resources can result in cognitive strain, which may negatively affect work performance and well-being at the workplace (e.g. see Kirschner (2002); Young et al. (2015). The assessment of cognitive strain is therefore an essential part of industrial engineering in the design of ergonomic workplaces. In this regard, pupil dilation has been used as an indicator to capture cognitive strain across a wide range of empirical eye tracking studies (see for example Bækgaard et al. (2019); Fletcher et al. (2017); Orlandi and Brooks (2018)). However, a pupil does not solely dilate with cognitive activity, but reacts to changing light conditions, as this is the pupil's main function besides accommodation adaptation (Beatty and Lucero-Wagoner, 2000). Concerning empirical investigations, this fact requires standardization of illumination throughout the entire experiment to allow valid interpretation of the indicator, which mostly limits eye tracking methods to be used in laboratory conditions. Therefore, this research investigates pupillometric indicators with regard to their sensitivity concerning informational work tasks and visual influences, such as light or information manipulation.

The interaction of pupil size and cognitive processes is based on changes in the central nervous system influencing pupil dilation (Beatty and Lucero-Wagoner, 2000). This is actuated by the locus coeruleus that is altering the activity of prefrontal cortex (Mathôt, 2018). In contrast to light reflex, which is characterized by large scale pupillary movement, cognitive activity is reflected in phasic small scale rapid fluctuations of less than 0.5 mm (Beatty and Lucero-Wagoner, 2000). More over, research has consistently found a relationship between task difficulty and pupil diameter (e.g. Ahern and Beatty, 1979; Beatty, 1982; Beatty and Kahneman, 1966; Hess and Polt, 1964; Payne et al., 1968 or Peavler, 1974). According to the current state of knowledge, the adaptive gain theory found the role of the locus coeruleus system to be even more

---

* Corresponding author.

E-mail addresses: j.czerniak@iaw.rwth-aachen.de (J.N. Czerniak), n.schierhorst@iaw.rwth-aachen.de (N. Schierhorst), valeria.villani@unimore.it (V. Villani), lorenzo.sabattini@unimore.it (L. Sabattini), c.brandl@iaw.rwth-aachen.de (C. Brandl), a.mertens@iaw.rwth-aachen.de (A. Mertens), schwalm@dornieden-gruppe.com (M. Schwalm), v.nitsch@iaw.rwth-aachen.de (V. Nitsch).

https://doi.org/10.1016/j.apergo.2020.103342
Received 8 April 2020; Received in revised form 24 November 2020; Accepted 10 December 2020

complex and specific, reflecting exploitation and exploration behaviour (Aston-Jones and Cohen, 2005).

However, in an experiment assessing cognitive workload, attention has to be paid to remove the variation of pupil light response from psychosensory pupil response, making it difficult to implement pupillometry in realistic settings. In order to address this problem, the Index of Cognitive Activity (ICA) has been introduced as a measure for cognitive activity (Marshall, 2000). The index is calculated by means of wavelet analysis that separates large scale pupillary movements and small scale rapid fluctuations to obtain different signal frequencies (Demberg et al., 2013). However, there is little empirical evidence to assess its validity, since only a small number of studies have addressed illumination effects and demonstrated that the index is robust against ambient light (Marshall, 2002) or quickly changing screen illumination (Rerhaye et al., 2018).

The ICA has been investigated in different task settings such as dual-task driving simulations, linguistic processing, mental arithmetics, learning tasks, or visual search tasks. This far, results are contradictory. While its relation to cognitive load has been shown in some studies (Ankener et al., 2018; Demberg et al., 2013; Demberg and Sayeed, 2016; Marshall, 2007; Schwalm et al., 2008), others state no significant effects (Debue and van de Leemput, 2014; Korbach et al., 2018). In a third group of studies, effects were found for only a few of the tested factors (Kahya et al., 2018; Matthews et al., 2015; Platten, 2013; Vogels et al., 2018). In addition, only a little number of studies have evaluated the ICA concerning task-evoked responses of more than two tasks (Bartels and Marshall, 2012; Korbach et al., 2018; Platten, 2013). Further, the relation between the ICA and pupil diameter has seldomly been investigated, since only a few studies focus on multiple measures (Debue and van de Leemput, 2014; Demberg et al., 2013).

In summary, the ICA's sensitivity to differentiate between multiple levels of task-evoked response and illumination requires further analysis. For this reason, two experimental studies were conducted. The objective of the first study was to analyse the ICA as a measure for increasing task difficulty amongst multiple mental arithmetics tasks in comparison to pupil dilation and NASA-TLX (Hart and Staveland, 1988). In a second study, it was investigated if there are visual effects when manipulating screen polarity or information density displayed. In this regard, for instance Backs and Walrath (1992) found evidence that display design, e.g. color or symbol coding influence pupil response. Similarly, Bernhardt et al. (2019) suggest that pupil diameter is likely to be sensitive towards increasing visual load. Influences of polarity due to screen illumination on pupil dilation were found by Dobres et al. (2017).

2. Experiment 1: mental arithmetics

2.1. Method

The first experiment investigated the question if the ICA is a sensitive measure for task evoked pupillary response. We hypothesized that the ICA increases with increasing task difficulty of mental arithmetics and shows significant differences between the conditions in line with pupil diameter (H01). In order to point to increasing task difficulty, and, thus, increasing mental strain and sensitivity to different levels of task difficulty, additionally subjective strain was evaluated with the NASA Task Load Index (Hart and Staveland, 1988) and performance was assessed by the number of solved problems. We hypothesized that NASA-TLX values increase with increasing task difficulty of mental arithmetics and show significant differences between the levels of difficulty (H02).

2.1.1. Participants

An opportunity sample of $N = 22$ participants (8 male, 14 female) aged between 20 and 56 ($AM = 28$, $SD = 9$) took part in the experiment. Participants were recruited in Aachen, Germany. An exclusion criterion for the participants was a limited visual acuity which could not be compensated with the help of an appropriate visual aid (min 0.8 diopters). The subjective assessment of mathematical knowledge ($AM = 3.5$, $SD = 0.9$) and the ability of mental arithmetics ($AM = 3.1$, $SD = 0.8$) were assessed using a 5-point Likert scale (1 = very low, 5 = very high).

2.1.2. Experimental design

The experiment followed a single factor within-subject design with the factor "difficulty" manipulated on 5 levels. The difficulty levels D1 to D5 of the arithmetic problems per stage were determined according to the values from Table 1. The lowest difficulty level (D1) contained multiplication problems such as "6 × 9" or "7 × 13". The gradation of the individual stages was chosen based on previous studies in which a change of the pupillary diameter was measured depending on the difficulty of the given problem (Ahern and Beatty, 1979; Beatty, 1982; Payne et al., 1968) with the extension of two additional stages D4 and D5.

The squared products (e.g. 6 × 6, 16 × 16) were removed, leaving a total of 22 possible problems per stage. The study examined the following dependent variables: the ICA of both eyes (ICA_L/ICA_R), the relative pupil dilation of both eyes (PD_L/PD_R) in millimeters with regard to the baseline, and the NASA-TLX value. Since conditions were not randomized and in order to address within-subjects effects, baseline measurements were taken for each single participant before each difficulty condition. Participants were asked to solve the problems for each stage consecutively to generate a constant strain level for each difficulty stage according to the stress-strain concept (Rohmert, 1986). To reduce the effects of the more difficult problems on those of a lesser difficulty, the order of the conditions was not randomized, but presented in ascending order, starting with the lowest difficulty level. For instance, Truschzinski et al. (2018) found a significant effect between participants mood in relation to their performance.

2.1.3. Apparatus

The study was carried out under laboratory conditions using Seeing Machines' stationary eye tracker FOVIO running the Eyeworks™ software, including the Cognitive Workload Module (developed by Eyetracking Inc.). The eye tracker was placed centrally in front of a monitor screen (24" 16/9 LED screen, resolution 1,920 × 1,200), on which the math problems were displayed. Both the eye-tracker and the participants' chair were adjusted to standardize individual physical properties of each participant such as distance of the eyes and height. The lighting conditions in the test room were standardized using a preinstalled ceiling light and identical for all participants. The vision screening instrument used for the eye test was a Vistec Rhoda device.

2.1.4. Procedure

The study was performed in five stages. As part of the introduction (10 min), the participants were fully briefed on the purpose and study procedure as well as the use of the collected data and their right to ask questions and to withdraw from the study at any time without adverse

Table 1
Difficulty levels with arithmetic problems.

Difficulty Condition		Integer Multiplication Problems			
		First Multiplier		Second Multiplier	
Name	Description	Lower Bound	Upper Bound	Lower Bound	Upper Bound
D1	Very low	6	9	6	9
		6	9	11	14
D2	Low	6	9	16	19
		11	14	11	14
D3	Intermediate	11	14	16	19
		16	19	16	19
D4	High	21	24	26	29
		26	29	26	29
D5	Very high	31	34	36	39
		36	39	36	39

consequences. Upon this briefing, participants signed an Informed Consent Declaration. Afterwards, they filled out the preliminary survey assessing demographic data (5 min). After the introduction, visual acuity was investigated (5 min). Before the start of the experiment, participants were able to familiarize themselves with the procedure of the experiment during a trial run. During the baseline measurements, they were instructed to relax and look at a white area displayed on the monitor for 30 seconds. This white area held the same hue, saturation and size as the area in which the arithmetic problems were presented. In each difficulty stage D1 - D5, the participants were given 2 min to solve as many problems as possible. While solving the arithmetic problems for each stage, the participants were asked to maintain their gaze in the direction of the eye tracker. After solving each problem, the answer was spoken aloud by the participant to check for mistakes. The participants were not told if their answer was correct. The subjective assessment of the cognitive strain was measured immediately after the completion of each difficulty condition followed by a break. The break duration started at 2 minutes and increased for 1 minute after each ascending stage. The experiment lasted approximately 1 hour for each participant.

2.1.5. *Statistical analysis and data preparation*

The statistical analysis was carried out using IBM SPSS Statistics 25, calculating multiple repeated measures ANOVAs. The accepted α-level was $p = .05$. If the assumption of sphericity was violated, F-ratios were adjusted using a Greenhouse-Geisser correction. For post-hoc pairwise comparisons, the accepted level was adjusted using a Bonferroni correction to adjust for an inflated familywise error rate. Data were prepared by using a Matlab tool and Microsoft Excel. All values except from “-99”, indicating a recording failure or blink, were included in the analysis. The further analysis uses the mean values of the ICA without using any normalization. Baseline pupil diameter values were subtracted from task mean pupil diameter (Laeng and Alnaes, 2019).

2.2. *Results*

All data were included in the experimental analysis ($N = 22$). Descriptive statistics, including baseline values, means, and standard deviations are shown in Table 2. ICA and PD baseline means and standard deviations can be derived from Fig. 1.

Further, descriptive statistics for ICA_L and ICA_R show similar mean values with no increase in task difficulty (Fig. 1). However, baseline means for the ICA of the left and right eye were higher than condition means. Mean changes of PD_L and PD_R have shown slightly different developments with respect to the difficulty levels. The largest mean change for both PD_L and PD_R was observed for the lowest difficulty level (D1). Smaller mean changes were found for D2 - D4, with the smallest for D5. Mean NASA-TLX values tended to increase with increasing task difficulty.

Table 2

Experiment 1 (mental arithmetic): change to baseline measurements: arithmetic mean *(AM)* and standard deviation *(SD)*. ICAL and ICAR denote the ICA of the left and right eye, while PDL and PDR are pupil diameter of the left and right eye, respectively. Bold values indicate the highest value for each dependent variable.

Change to Baseline AM (SD)	D1	D2	D3	D4	D5
ICA_L	−0.011 (0.035)	−0.017 (0.041)	−0.017 (0.065)	−0.002 (0.065)	**−0.020** **(0.074)**
ICA_R	−0.010 (0.067)	−0.013 (0.050)	**−0.033** **(0.066)**	−0.028 (0.064)	−0.031 (0.060)
PD_L [mm]	**0.141** **(0.105)**	0.048 (0.124)	0.047 (0.116)	0.053 (0.139)	0.019 (0.113)
PD_R [mm]	**0.153** **(0.100)**	0.063 (0.130)	0.060 (0.118)	0.062 (0.143)	0.027 (0.112)

2.2.1. *Subjective and performance measures*

The main effect of NASA-TLX was significant, $F\,(4, 84) = 11.07$, $p < .001$, partial $\eta^2 = 0.345$. The results of the Bonferroni-corrected post-hoc tests revealed that there is a significant difference between factor level 1 in comparison to factor levels 3, 4 and 5 and between factor level 2 and factor level 5, $p < .023$. Results reveal a significant main effect of performance, $F(2.462, 51.705) = 137.786$, $p < .001$, partial $\eta^2 = 0.868$. Bonferroni corrected pairwise comparisons show significant differences between all conditions ($p < .001$), except condition 4 and 5 ($p = .055$).

2.2.2. *Scaled ICA and pupil diameter*

The Greenhouse-Geisser corrected main effect of the ICA_L for univariate analysis was not significant for the different conditions, $F(2.8, 58.81) = 1.191$, $p = .32$, partial $\eta^2 = 0.054$. ICA_R was not significantly affected by increasing task difficulty $F(2.77, 58.14) = 0.187$, $p = .892$, partial $\eta^2 = 0.009$, either. For PD_L, the univariate analysis determined that mean difficulty levels showed a statistically significant difference between measurements, $F(4, 84) = 7.892$, $p < .001$, partial $\eta^2 = 0.273$. Pairwise comparisons with Bonferroni-correction showed statistically different results between conditions 1 and the other conditions, $p < .009$. The PD_R analysis determined that the main effect of the univariate factor difficulty showed a statistically significant difference between measurements of the factor levels, $F(4,84) = 8.348$, $p = .001$, partial $\eta^2 = 0.284$. Bonferroni-corrected pairwise comparisons revealed a significant difference between the first condition compared to the other conditions, $p < 0.018$. Correlation analysis of left and right mean values reveal strong positive correlations for all difficulty conditions D1 - D5, ICA: $r > 0.809$, $p < .001$ (D3), PD: $r > 0.827$, $p < .001$. Results reveal a significant main effect for performance, $F(2.462, 51.705) = 137.786$, $p < .001$, partial $\eta^2 = 0.868$. Bonferroni corrected pairwise comparisons show significant differences between all conditions ($p < .001$), except condition 4 and 5 ($p = .055$).

2.3. *Discussion*

The goal of this research was to determine whether the ICA could capture within-task difficulty effects. Pupil dilation was tested, additionally. The NASA-TLX was used to check if task-evoked difficulty manipulations represent different stress levels. It was hypothesized that the dependent variables would increase with task difficulty, and that difficulty levels would show significant differences as well as that the repeated measures ANOVA would reveal significant results, accordingly. However, univariate results showed inconsistent tendencies.

2.3.1. *Subjective and performance measures*

NASA-TLX mean values showed significant differences for two difficulty levels. However, the null hypothesis of H02 cannot be rejected. This measure shows that difficulty was increasing between task levels. These findings are supported by performance measures that showed to significantly decrease with increasing difficulty and to distinguish between difficulty levels up to level 4. Thus, the experimental setting was shown to elicit at least two levels of task difficulty, as expected. Results reveal a dissociation between subjective and physiological measures. This can often be observed in empirical studies, since the subjective experience of workload does not necessarily comply to the physiological reaction of the body (see for instance Orlandi and Brooks (2018)), since they require the participants ability to introspect to their cognitive processes (Boekaerts, 2017), and further they do not capture variabilities over a certain period of time (Chen et al., 2016). Apparently, the physiological reaction induced in the present study was not strong enough to distinguish between the workload levels, whereas subjectively increasing workload could be perceived between the conditions. It is being assumed that psychological factors are likely to have influenced these findings.

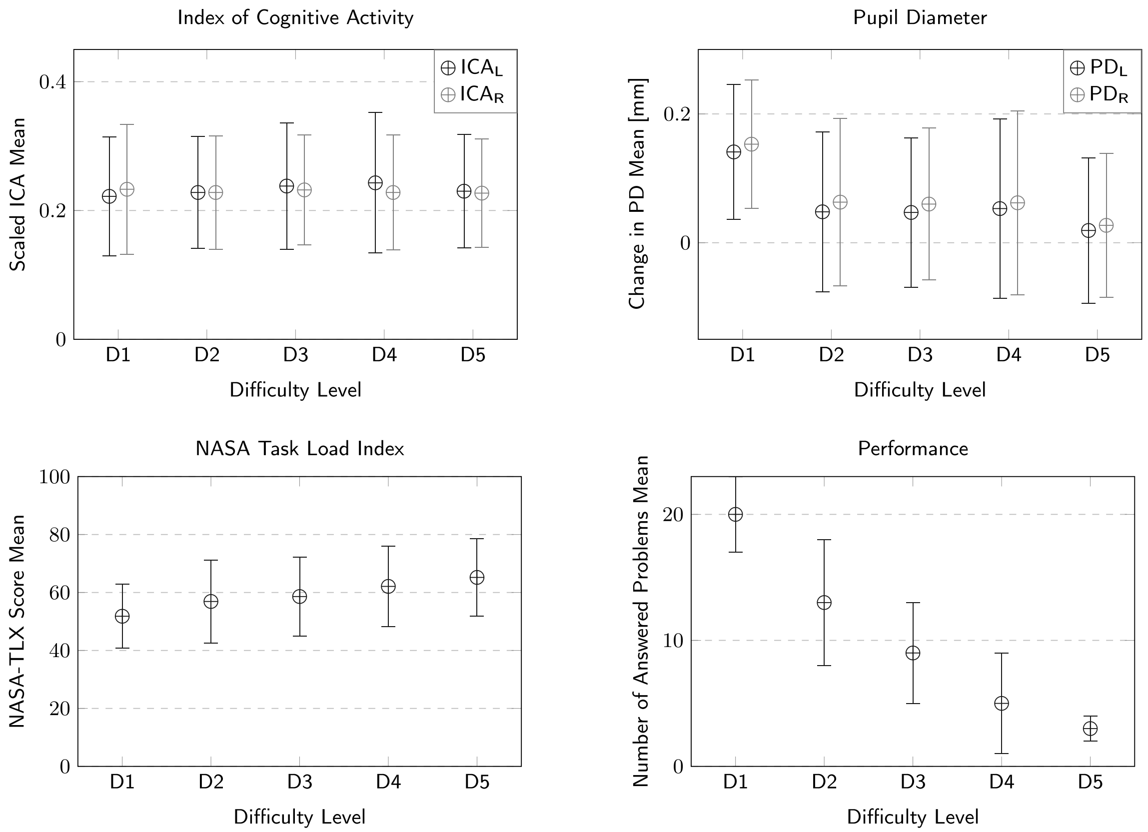


Fig. 1. Mean and standard deviations (circles and error bars) of dependent variables Index of Cognitive Activity, Pupil Diameter, NASA-TLX, and Performance across all difficulty levels D1 - D5.

2.3.2. *Scaled ICA and pupil diameter*

However, data show that within-task difficulty did not relate with changes in ICA values. Unlike subjective measures, ICA and pupil diameter did not show to be sensitive for the evoked cognitive stress in the present study. Post-hoc test concerning pupil dilation reveal that PD might reflect participants' stress due to nervousness at the beginning of the experiment, since the difficulty conditions were not randomized. Therefore, the null hypothesis of H01 cannot be rejected. A substantial difference to prior research is the number of factor levels that were tested. Korbach et al. (2018), Marshall (2002), and Schwalm et al. (2008) for instance tested a baseline condition against a workload condition instead of consecutive steps of difficulty. Hence, the scaled ICA might be able to indicate differences between a relaxed state and an engaged state, but, according to the results, does not seem to be a sensitive indicator for different levels of increasing task difficulty. Moreover, the index is likely to be covered by other effects, such as arousal, response selection or a ceiling effect of the scale.

Unlike the PD values, the measured baseline values of the scaled ICA were higher in the baseline conditions than in the task conditions. This effect was inexplicable at this point. Since baseline conditions were taken in front of a white screen, it is conceivable that there might have been visual effects regarding the screen polarity or the information presented. Based on this finding, further hypotheses were formulated and tested in Experiment 2.

Since positive correlations were found between mean of both eyes, it is assumed that there is no difference between both eyes.

3. Experiment 2: visual influences

3.1. *Method*

In the previous study, the effect of high ICA baseline values could not be explained by means of the experimental variation. We assume that either tension or visual effects of the information shown on the screen are the reason even if this is contradicting with PD results. In this regard, for instance Backs and Walrath (1992) found evidence that display design, e.g. color or symbol coding, influences pupil response. Influences of polarity due to screen illumination on pupil dilation were found by Dobres et al. (2017). For this reason, we conducted a second experiment focusing on screen polarity and presentation of information in two trials. For instance, Bernhardt et al. (2019) suggest that pupil diameter is likely to be sensitive towards increasing visual load. We hypothesize that screen polarity influences pupil diameter but not the scaled ICA (H03). Further, we hypothesize that pupil diameter and the scaled ICA is sensitive to visual strain by information manipulation (H04).

3.1.1. *Participants*

Twenty participants were recruited in Aachen, Germany (11 male, 9 female). Their ages ranged from 19 to 34 years ($AM = 26$, $SD = 4$). The exclusion criterion was similar to the first experiment concerning irreparable visual acuity with visual aid lower than 0.8 dioptres. No participant was excluded due to this criterion.

3.1.2. *Experimental design*

The experiment followed a within-subject design for three independent variables: information, polarity, and trial. Factor levels of information were: none (1), fixation cross (2), low density (3), high density (4), centred circle (5), and moving circle (6). Density was tested with a small number of nine centred dots (3×3 rows) versus a high number of dots (9×13 rows) spread over the entire screen, according to findings of Bernhardt et al. (2019). Regarding movement, a stationary filled circle was compared to a circle moving with constant speed. Factor levels of polarity were: white (1) and black (2) (Dobres et al., 2017). Fig. 2 shows an overview of factors levels of polarity and illumination.

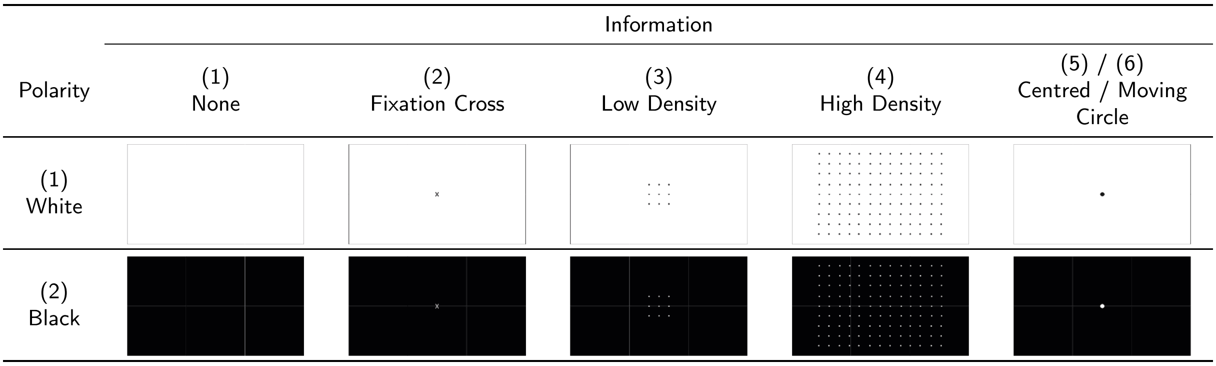


Fig. 2. Information displayed for the factors polarity and information.

Each person did the experiment twice in order to exclude habituation effects. The Latin square method was used to randomize conditions. Dependent variables were PD and ICA. Mean values were calculated by arithmetic means between left and right eyes, since correlations were found in the previous experiment, and since the procedure is useful to reduce artefacts (Laeng and Alnaes, 2019). Additionally, the galvanic skin response (GSR) was used as control variable, since it is sensitive for cognitive strain.

3.1.3. *Apparatus*

The same eye tracker and software and a similar arrangement was used as in the first experiment. GSR was analysed using finger electrodes and CAPTIVE software. A headrest was positioned 45 cm in front of the monitor. Illumination in the room was kept constant and identical for all participants. Additionally, direct illumination was avoided in the direction of the eye tracker by using a single programmable uplight. There was no other light source in the room.

3.1.4. *Procedure*

Participants were introduced to the study as they were in the first experiment. Visual acuity was investigated before the introduction. Participants completed a questionnaire concerning general questions regarding demographic data. Afterwards, the GSR sensor was placed at the ring finger and the little finger of the left hand. The eye tracker was calibrated using a 5-spot calibration. A baseline measurement of 30 seconds was taken at the beginning of each trial. During this measurement, participants gazed into the direction of the eye tracker at the switched-off screen. Each factor level was displayed for 30 seconds. Before the experiment started, participants were instructed to look at the stimulus on the monitor if one was present, or straight at the screen if no stimulus was shown.

3.1.5. *Statistical analysis and data preparation*

Statistical analysis was carried out similar to the first experiment. With regards to GSR, data were prepared as follows. Phasic SCR values of galvanic skin response was analysed using Ledalab 3.4.9 performing CDA analysis in order to avoid biases of classic peak detection methods (Benedek and Kaernbach, 2010).

3.2. *Results*

Data of two participants were excluded due to artefacts and missing records. Therefore, data of eighteen participants were included in the analysis and a total of 36 data sets were analysed. Fig. 3 shows the means of all ICA and PD values of the first trial. The following baseline mean values can be derived from the results (trial 1), ICA: 0.38 (±0.18), PD: 4.66 (±0.18). Results reveal that mean values were higher when the screen was white for the dependent variables PD and ICA amongst all factor levels of information in both trials. GSR values tended to be similar in both polarity conditions.

Significant univariate main effects of information were shown for PD, $F(5, 85) = 8.667, p < .001$, partial $\eta^2 = 0.338$, but not for the ICA, F

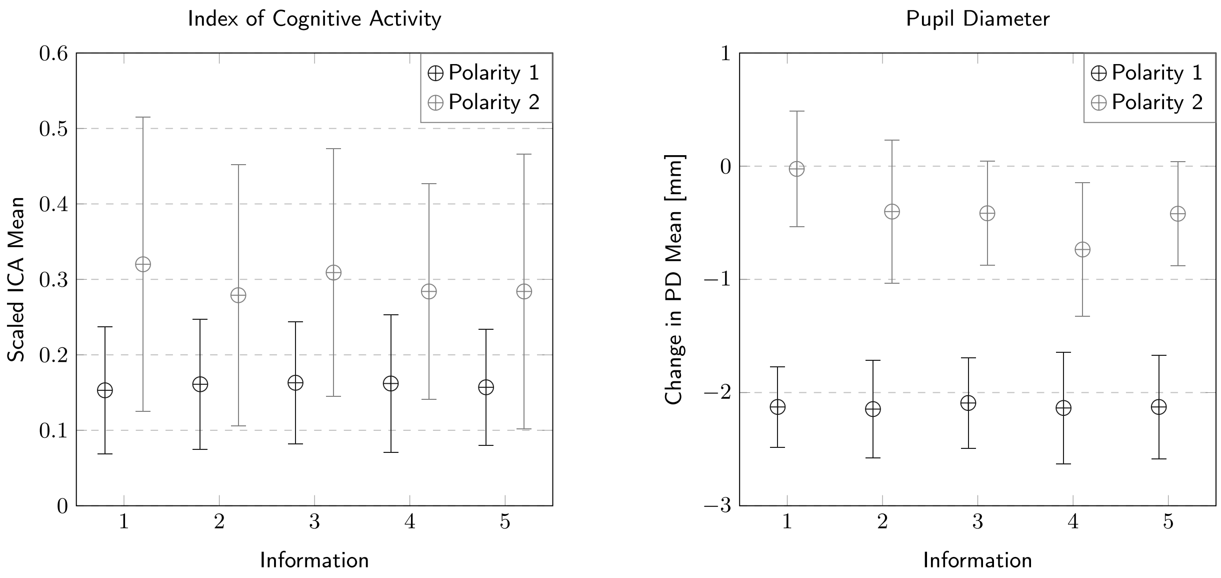


Fig. 3. Mean and Standard Deviations (circles and error bars) of Dependent Variables: Mean ICA, Mean Pupil Diameter Change PD across all Levels of Information, Trial 1. Polarity 1 - white, polarity 2 - black.

$(5, 85) = 1.555$, $p = .182$, partial $\eta^2 = 0.048$. The factor polarity revealed a significant univariate effect of both, ICA, $F(1, 17) = 27.104$, $p < .001$, partial $\eta^2 = 0.615$, and PD, $F(1, 17) = 225.278$, $p < .001$, partial $\eta^2 = 0.93$. The interaction between information and polarity showed significant univariate effects for PD, $F(5, 85) = 10.749$, $p < .001$, partial $\eta^2 = 0.387$, the main interaction effect of the ICA, on the other hand, was not shown to be significant, $F(5, 85) = 1.602$, $p = .168$, partial $\eta^2 = 0.086$. GSR did not reveal a significant univariate main effect on any factor, and moreover, the factor trial was not significant for any dependent variable ($p > .05$). Bonferroni-corrected post-hoc tests of information revealed significant differences for PD, $p < .005$, between factor levels 1 and 4, 1 and 5, as well as 2 and 4. Respective pairwise comparisons of the factor polarity were significant for ICA and PD, $p < .001$.

3.3. Discussion

The aim of this study was to investigate the ICA with regards to visual influences, such as light and information manipulation.

Firstly, the ICA was hypothesized not to be affected by changing light conditions in contrast to PD. This hypothesis was supported by the findings of Marshall (2002) and Rerhaye et al. (2018), who did not find significant effects, when participants were sitting in a dark versus illuminated room or when showing a flickering monitor. In contrast, results of the present experiment show significant differences between screen polarity, increasing with negative polarity like PD change. Since no stress besides visual manipulation was induced, we assume that changes in PD and ICA rely on screen polarity rather than on psychosensory pupil response, as also reported by Brunyé et al. (2013) or Dobres et al. (2017). GSR results support this assumption, as they indicate no significant differences for the factor polarity. We assume that the ICA may not be affected by room illumination or flickering light, but is likely to be sensitive towards screen polarity when exposition duration is kept constant for a couple of seconds. However, the null hypothesis of H03 cannot be rejected.

Secondly, it was assumed that the ICA would be sensitive towards visual strain by means of information manipulation. Since the results do not indicate significant effects for the ICA on information, the null hypothesis of H04 cannot be rejected. This finding implies that the ICA is not likely to indicate visual load besides cognitive strain and, thus might not be dependent on the kind of information presentation. However, based on the results it cannot be ruled out that visual stimuli might have an effect with longer exposure times.

However, ICA means reveal a higher variance between negative in comparison to white polarity. Therefore, it is unlikely that the white baseline screen would cause higher baselines than task means. PD on the other hand, showed significant differences between factor levels with increasing information density, and thus, the null hypothesis of H04b can be rejected. Since the ICA showed similar tendencies as the PD for polarity, it can be assumed that the ICA might be likely to vary for information when exposed to less light. Further empirical research shall cope with this question using longer exposure times of single stimuli.

4. Conclusion and outlook

Two studies were conducted that investigated the ICA and PD regarding their sensitivity to detect task-evoked response and visual effects. According to the results of the first experiment, both of these dependent variables were found to be non-sensitive towards cognitive strain induced by the used math problems. In contrast to these findings, subjective measures showed to be sensitive to distinguish for two difficulty levels. It was concluded that the indicators may be sensitive to cognitive strain compared to non-cognitive states. However, it was concluded that the ICA reflects other effects besides cognitive workload. Subsequently, a second experiment was conducted to examine visual influences of screen polarity or information presentation and found significant differences for polarity concerning the ICA. Hence, it was concluded that the ICA is likely to be sensitive for constant screen polarity. However, despite no significant differences for the type of information presentation, the ICA and the changes in pupil diameter tended to show a higher variance for negative screen polarity. This might indicate that effects of information presentation on the ICA and changes in pupil diameter may be detectable when inducing a longer exposure time to the stimuli.

In summary, the sensitivity of pupillometric measures is likely to be limited to distinguish cognitive states from relaxed states, since results did not reflect multiple levels of task-evoked response. Further, the suggestion of the ICA as a promising cognitive indicator for environments with changing light conditions could not be supported in this research. However, the ICA is likely to indicate perceptive strain induced by visual influences of negative polarity. Concerning practical implementation in field studies, these results could not prove that the Index of Cognitive Activity has advantages over pupillary measures with regard to its robustness towards light influences. However, pupillary measures have the potential for field investigations for instance in production, since they represent a continuous real-time measure for task demand (Fletcher et al., 2017). In order to do so, further research is required to analyse visual effects in detail in order to utilize the ICA or similar pupillometric measures as indicators of cognitive strain.

Declaration of competing interest

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

Acknowledgements

The research has been carried out within the “Smart and Adaptive Interfaces for INCLUSIVE Work Environment” project, funded by the European Union’s Horizon 2020 Research and Innovation Program under Grant Agreement N723373. The research is continued within the “Smart Working Environments for all Ages” project WorkingAge, funded by the European Union’s Horizon 2020 Research and Innovation Program under Grant Agreement N826232. The authors would like to express their gratitude for the given support.

References

Ahern, S., Beatty, J., 1979. Pupillary responses during information processing vary with scholastic aptitude test scores. Science 205, 1289–1292. https://doi.org/10.1126/science.472746.

Ankener, C.S., Sekicki, M., Staudte, M., 2018. The influence of visual uncertainty on word surprisal and processing effort. Front. Psychol. 9, 2387. https://doi.org/10.3389/fpsyg.2018.02387.

Aston-Jones, G., Cohen, J.D., 2005. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu. Rev. Neurosci. 28, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709.

Backs, R.W., Walrath, L.C., 1992. Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. Appl. Ergon. 23, 243–254. https://doi.org/10.1016/0003-6870(92)90152-L.

Bækgaard, P., Jalaliniya, S., Hansen, J.P., 2019. Pupillary measurement during an assembly task. Appl. Ergon. 75, 99–107. https://doi.org/10.1016/j.apergo.2018.09.004.

Bartels, M., Marshall, S.P., 2012. Measuring cognitive workload across different eye tracking hardware platforms. In: Morimoto, C.H. (Ed.), Proceedings of the Symposium on Eye Tracking Research and Applications. ACM, New York, NY, pp. 161–164. https://doi.org/10.1145/2168556.2168582.

Beatty, J., 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychol. Bull. 91, 276–292. https://doi.org/10.1037/0033-2909.91.2.276.

Beatty, J., Kahneman, D., 1966. Pupillary changes in two memory tasks. Psychonomic Sci. 5, 371–372. https://doi.org/10.3758/BF03328444.

Beatty, J., Lucero-Wagoner, B., 2000. The pupillary system. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G. (Eds.), Handbook of Psychophysiology. Cambridge University Press, New York, NY, USA, pp. 142–162.

Benedek, M., Kaernbach, C., 2010. Decomposition of skin conductance data by means of nonnegative deconvolution. Psychophysiology 47, 647–658. https://doi.org/10.1111/j.1469-8986.2009.00972.x.

Bernhardt, K.A., Poltavski, D., Petros, T., Ferraro, F.R., Jorgenson, T., Carlson, C., Drechsel, P., Iseminger, C., 2019. The effects of dynamic workload and experience on commercially available eeg cognitive state metrics in a high-fidelity air traffic control environment. Appl. Ergon. 77, 83–91. https://doi.org/10.1016/j.apergo.2019.01.008.

Boekaerts, M., 2017. Cognitive load and self-regulation: attempts to build a bridge. Learn. InStruct. 51, 90–97. https://doi.org/10.1016/j.learninstruc.2017.07.001.

Brunyé, T.T., Howe, J.L., Kimball, B.R., Eddy, M.D., Mahoney, C.R., 2013. Variable transmission lens influences on the dynamics of pupillary light reflexes. Ergonomics 56, 1745–1753. https://doi.org/10.1080/00140139.2013.832806.

Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S.Z., Khawaji, A., Conway, D., 2016. Robust Multimodal Cognitive Load Measurement. Human–Computer Interaction Series. Springer International Publishing.

Debue, N., van de Leemput, C., 2014. What does germane load mean? an empirical contribution to the cognitive load theory. Front. Psychol. 5, 1099. https://doi.org/10.3389/fpsyg.2014.01099.

Demberg, V., Kiagia, E., Sayeed, A., 2013. The index of cognitive activity as a measure of linguistic processing. In: Proceedings of the 35th annual meeting of the cognitive science society (cogsci-13).

Demberg, V., Sayeed, A., 2016. The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. PloS One 11, e0146194. https://doi.org/10.1371/journal.pone.0146194.

Dobres, J., Chahine, N., Reimer, B., 2017. Effects of ambient illumination, contrast polarity, and letter size on text legibility under glance-like reading. Appl. Ergon. 60, 68–73. https://doi.org/10.1016/j.apergo.2016.11.001.

Fletcher, K., Neal, A., Yeo, G., 2017. The effect of motor task precision on pupil diameter. Appl. Ergon. 65, 309–315. https://doi.org/10.1016/j.apergo.2017.07.010.

Hart, S.G., Staveland, L.E., 1988. Development of nasa-tlx (task load index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (Eds.), Advances in Psychology, vol. 52. Elsevier, North-Holland, pp. 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9.

Hess, E.H., Polt, J.M., 1964. Pupil size in relation to mental activity during simple problem-solving. Science 143, 1190–1192. https://doi.org/10.1126/science.143.3611.1190.

Kahneman, D., 1973. Attention and Effort. In: Prentice Hall Series in Experimental Psychology. Prentice Hall, Englewood Cliffs.

Kahya, M., Wood, T.A., Sosnoff, J.J., Devos, H., 2018. Increased postural demand is associated with greater cognitive workload in healthy young adults: a pupillometry study. Front. Hum. Neurosci. 12, 288. https://doi.org/10.3389/fnhum.2018.00288.

Kirschner, P.A., 2002. Cognitive load theory: implications of cognitive load theory on the design of learning. Learn. InStruct. 12, 1–10. https://doi.org/10.1016/S0959-4752(01)00014-7.

Korbach, A., Brünken, R., Park, B., 2018. Differentiating different types of cognitive load: a comparison of different measures. Educ. Psychol. Rev. 30, 503–529. https://doi.org/10.1007/s10648-017-9404-8.

Laeng, B., Alnaes, D., 2019. Pupillometry. In: Klein, C., Ettinger, U. (Eds.), Eye Movement Research, Studies in Neuroscience, Psychology and Behavioral Economics. Springer, Cham, pp. 449–502. https://doi.org/10.1007/978-3-030-20085-5_11.

Marshall, S.P., 2000. Method and Apparatus for Eye Tracking and Monitoring Pupil Dilation to Evaluate Cognitive Activity.

Marshall, S.P., 2002. The index of cognitive activity: measuring cognitive workload. In: Persensky, J.J., Hallbert, B., Blackman, H. (Eds.), Proceedings of the IEEE 7th Conference on Human Factors and Power Plants. IEEE, pp. 5–9. https://doi.org/10.1109/HFPP.2002.1042860.

Marshall, S.P., 2007. Identifying cognitive state from eye metrics. Aviat Space Environ. Med. 78, B165–B175.

Mathôt, S., 2018. Pupillometry: psychology, physiology, and function. J. cogn. 1, 16. https://doi.org/10.5334/joc.18.

Matthews, G., Reinerman-Jones, L.E., Barber, D.J., Abich IV, J., 2015. The psychometrics of mental workload: multiple measures are sensitive but divergent. Hum. Factors 57, 125–143. https://doi.org/10.1177/0018720814539505.

Orlandi, L., Brooks, B., 2018. Measuring mental workload and physiological reactions in marine pilots: building bridges towards redlines of performance. Appl. Ergon. 69, 74–92. https://doi.org/10.1016/j.apergo.2018.01.005.

Payne, D.T., Parry, M.E., Harasymiw, S.J., 1968. Percentage of pupillary dilation as a measure of item difficulty. Percept. Psychophys. 4, 139–143. https://doi.org/10.3758/BF03210453.

Peavler, W.S., 1974. Pupil size, information overload, and performance differences. Psychophysiology 11, 559–566. https://doi.org/10.1111/j.1469-8986.1974.tb01114.x.

Platten, F., 2013. Analysis of Mental Workload and Operating Behavior in Secondary Tasks while Driving. Dissertation. Technische Universität Chemnitz. Chemnitz, Deutschland.

Rerhaye, L., Blaser, T., Alexander, T., 2018. Evaluation of the index of cognitive activity (ica) as an instrument to measure cognitive workload under differing light conditions. In: Bagnara, S., Tartaglia, R., Albolino, S., Alexander, T., Fujita, Y. (Eds.), Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018). Springer, Cham, pp. 350–359. https://doi.org/10.1007/978-3-319-96059-3_38.

Rohmert, W., 1986. Ergonomics: concept of work, stress and strain. Appl. Psychol. 35, 159–181. https://doi.org/10.1111/j.1464-0597.1986.tb00911.x.

Schwalm, M., Keinath, A., Zimmer, H.D., 2008. Pupillometry as a method for measuring mental workload within a simulated driving task. Human Factors for Assistance and Automation 1–13.

Truschzinski, M., Betella, A., Brunnett, G., Verschure, P.F.M.J., 2018. Emotional and cognitive influences in air traffic controller tasks: an investigation using a virtual environment? Appl. Ergon. 69, 1–9. https://doi.org/10.1016/j.apergo.2017.12.019.

Vogels, J., Demberg, V., Kray, J., 2018. The index of cognitive activity as a measure of cognitive processing load in dual task settings. Front. Psychol. 9 https://doi.org/10.3389/fpsyg.2018.02276.

Young, M.S., Brookhuis, K.A., Wickens, C.D., Hancock, P.A., 2015. State of science: mental workload in ergonomics. Ergonomics 58, 1–17.