Acta Psychologica 220 (2021) 103402

Contents lists available at ScienceDirect

Acta Psychologica

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

A meta-analytic review of the reliability of the Index of Cognitive Activity concerning task-evoked cognitive workload and light influences

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

[a] *Institute of Industrial Engineering and Ergonomics (IAW), RWTH Aachen University, Eilfschornsteinstraße 18, 52062 Aachen, Germany*
[b] *Institute of Highway Engineering (ISAC), RWTH Aachen University, Mies-van-der-Rohe-Straße 1, 52074 Aachen, Germany*

ARTICLE INFO

Keywords:
Index of Cognitive Activity
Light influences
meta-analysis
Task-evoked cognitive workload

ABSTRACT

The Index of Cognitive Activity (ICA) was introduced as a promising pupillary workload measure for field investigations since, unlike pupil dilation, it is not affected by illumination. Recent studies have investigated the ICA for task-evoked cognitive workload with contradictory findings. However, few studies investigated the influence of illumination on the ICA. Therefore, to examine inconsistencies regarding the reliability for workload measurement and the effects of light, a meta-analysis was conducted based on a structured literature review. The meta-analysis considered k = 14 studies with a total sample size of $N = 751$ participants. Results showed significant effects for workload ($r = 0.61$) and light ($r = 0.45$) on the ICA. Since moderating effects were found for several between-study differences, it seems likely that different cognitive processes and settings affect the indicator and should be considered in empirical investigations. According to the findings, the ICA is a reliable indicator for task-evoked workload. However, light influences were found which indicates that evidence-based conclusions regarding the ICA's practical applicability require further research.

1. Introduction

The Index of Cognitive Activity (ICA) has been subject to increasing attention in ergonomic and psychological research within the last decades. Introduced by Marshall (2000) it is associated with cognitive pupillary response separated from light influences. The ICA is a pupillary-based algorithm, which calculates a Daubechies wavelet analysis separating workload dilation from light dilation. The ICA asserts that workload dilation is reflected by phasic small rapid pupillary signals extending the tonic threshold value of light dilation. The calculation is finalized by adjusting for outliers employing a hyperbolic tangent transformation (Marshall, 2000). According to literature, it provides an alternative to task-evoked pupillary workload measurements in field investigations.

However, the ICA has been facing some criticism amongst literature, since it is a proprietary measure due to intransparency of the algorithm. Addressing this fact, Duchowski et al. (2018) for instance have developed a similar pupillary wavelet-based indicator, the Index of Pupillary Activity (IPA). In contrast to the ICA, the IPA algorithm is completely accessible. The IPA differs from the ICA in certain key aspects, such as the use of symlet-16 wavelets instead of Daubechies wavelets and a different threshold approach by using periodic DWT (Duchowski et al., 2018, p. 282). Though, little effort was done to investigate the influence of illumination on the ICA in literature, yet.

In general, the cognitive pupillary response physiology is related to the locus coeruleus due to the activity of the central nervous system when cognitive workload is induced (Mathôt, 2018). The resulting dilation is characterized by rapid small phasic pupil signals below 0.5 mm amplitude, peaking about 0.1 mm above the tonic dilation level (Beatty & Lucero-Wagoner, 2000). Therefore, pupillary response is associated with information processing effort in current empirical research fields (Backs & Walrath, 1992). Empirical evidence on this "task-evoked pupillary response" (TEPR) was found in numerous research work since the 1960s (Ahern & Beatty, 1979; Beatty, 1982; Beatty & Kahneman, 1966; Chen et al., 2016; Matton et al., 2020; Oliva, 2019; Orlandi & Brooks, 2018; Peavler, 1974; Tao et al., 2019; van der Wel & van Steenbergen, 2018; Wu et al., 2019).

However, standardization of light conditions is yet an unsolved

* Corresponding author.
E-mail addresses: j.czerniak@iaw.rwth-aachen.de (J.N. Czerniak), n.schierhorst@iaw.rwth-aachen.de (N. Schierhorst), c.brandl@iaw.rwth-aachen.de (C. Brandl), a.mertens@iaw.rwth-aachen.de (A. Mertens), schwalm@dorniedengruppe.com (M. Schwalm), v.nitsch@iaw.rwth-aachen.de (V. Nitsch).

https://doi.org/10.1016/j.actpsy.2021.103402
Received 26 January 2021; Received in revised form 4 August 2021; Accepted 19 August 2021
Available online 8 September 2021

problem in research, limiting this kind of physiological workload evaluation to standardized laboratory settings. Realizing reliable pupillary measurement in uncontrolled work environments dealing with changing light conditions, e.g. production or office environments, would be a promising measure concerning physiological workload assessment. Thus, the ICA could contribute to further advances in ergonomic research improving working conditions.

1.1. State of the research

1.1.1. Single tasks

The Index of Cognitive Activity has been investigated in numerous settings with regards to cognitive activity (Fairclough et al., 2009; Korbach et al., 2017, 2018; Rerhaye et al., 2018; Richstone et al., 2010; Schwalm, 2009). Initially, scientific evidence for the relationship to cognitive workload has been found concerning the difficulty of mental arithmetic tasks (Marshall, 2000, 2002; Marshall, 2007; Marshall et al., 2004) and was related to arousal (Marshall, 2002) reflecting cognitive states during information processing (Marshall, 2007). Supporting these findings, Schwalm (2009) reported significant effects on the ICA, of both task difficulty and perception modality, whereby auditory information input led to significantly higher ICA means than visual provided information. Further, statistical evidence for the ICA's sensitivity to increasing task difficulty was found by Tourtouri et al. (2019) in a 6-back task setting. The findings of Rerhaye et al. (2018) showed differences in significance between task types. However, there are also contradicting findings that did not confirm an effect of task-evoked cognitive activity on the ICA (Czerniak et al., 2021; Korbach et al., 2017, 2018).

1.1.2. Multiple tasks

The ICA has also been investigated in more complex informational tasks (Bartels & Marshall, 2012; Demberg, 2013; Demberg et al., 2013; Dlugosch et al., 2013; Marshall et al., 2003; Matthews et al., 2015; Platten, 2012; Reinerman-Jones et al., 2014; Schwalm, 2009; Schwalm et al., 2008; Vogels et al., 2018). In this context, some authors increased complexity involving multiple parallel tasks to provide a more realistic setup. Effects of task difficulty on the ICA in these settings were found in several studies. Bartels and Marshall (2012) investigated a combination of smooth pursuit, mental arithmetic, and attention tasks reporting significant differences between the conditions. In particular, dual-task designs are employed frequently in driving simulations, where a secondary cognitive task is applied to the visual-motor reactive main task of driving lane changes and traffic attention. For instance, Dlugosch et al. (2013) investigated driving tasks with additional cognitive tasks and concluded that the ICA significantly increased in these conditions. Schwalm et al. (2008) argued that participants focused more on the secondary task, indicating that the ICA is likely to reflect strategy and attention shifts of the driver during dual tasks, which supports the findings of Marshall et al. (2003) and Demberg et al. (2013). Consistently, Vogels et al. (2018) found a significant decrease in the ICA in dual task settings as compared to the single task and even bigger effects for more difficult secondary tasks. In contrast, the results of Matthews et al. (2015) could not support this assumption.

1.1.3. Linguistic processing

As linguistic processing can be related to phasic small pupillary responses, a few studies investigated the ICA in that context (Ankener et al., 2018; Demberg et al., 2013; Demberg & Sayeed, 2016; Sekicki & Staudte, 2018; Tourtouri et al., 2019). According to Demberg et al. (2013), the ICA reflects linguistic complexity, which was supported by the findings of Demberg and Sayeed (2016). Connected to processing effort, information-theoretic concepts of entropy (Shannon & Weaver, 1949) and surprisal (Levy, 2008) have been subject to recent studies. In this context, findings of Ankener et al. (2018) reveal the ICA was affected by surprisal, which caused processing effort of the given linguistic stimuli. In line with that, Tourtouri et al. (2019) found significant differences in comprehension of different words combined with visual cues of manipulated entropy. The ICA further showed that visual attention shifts towards the cued object lowers the effort required for processing the linguistic reference (Sekicki & Staudte, 2018).

1.1.4. Visual influences

Concerning the indicator's sensitivity for visual influences, such as light, only a few studies exist (Czerniak et al., 2021; Debue & van de Leemput, 2014; Marshall et al., 2004; Rerhaye et al., 2018). Results of Marshall et al. (2004) showed statistical evidence for the indicator's robustness towards light, which was supported by results of Rerhaye et al. (2018) during a mental rotation task. Interestingly, the ICA seemed to be significantly affected by light in their Stroop task, although the light manipulation was similar for both tasks. The findings of Czerniak et al. (2021), on the other hand, investigated visual effects without any cognitive task, revealing significant effects of screen polarity, which is a crucial finding concerning the indicator's definition. However, no significant differences were found for visual load by information presentation. These results were supported by findings of Debue and van de Leemput (2014), who investigated visual influences by manipulating the type of information displayed.

1.2. Motivation and research questions

In summary, literature does not reveal a uniform picture of the reliability of Index of Cognitive Activity concerning task-evoked response and light. Although all studies followed similar research questions, results reveal inconsistent results concerning significance and effect sizes, thus prohibiting a valid conclusion concerning the indicator's evidence. It seems likely that this inconsistency can be related to between-study variability, and thus moderating effects have to be considered as influencing factors. For one, it can be assumed that different cognitive tasks require different cognitive resources and affect arousal differently, for instance, when time pressure is induced by task pace or multiple tasks have to be conducted in parallel. Further, the type or number of modalities addressed by the task is likely to have an influence on the ICA. Moreover, studies differed in sample sizes, repeated measures, and normalization, which are all features likely to affect results and effect sizes. It seems further likely that differences in results may be related to the question of how sensitive is the indicator to multiple workload levels. Although the used hardware seems unlikely to have an effect on the results according to Marshall et al. (2004), it still seems important to consider this as an influence due to quality and measurement differences. Furthermore, the number of studies investigating the actual effect of light towards the ICA is rare and likewise generated contradictory results (Czerniak et al., 2021; Marshall et al., 2004; Rerhaye et al., 2018). Between-study differences with a potential moderating effect can be found with regards to the type of light, the speed of change, the type of cognitive activity, modality influences and the quality of the studies. In order to give deeper insight into this topic, the following research questions were evaluated utilizing a meta-analysis:

- is the ICA a reliable indicator for task-evoked cognitive workload,
- is the ICA sensitive to light, and
- which influencing factors moderate the relationship between workload and the ICA?

2. Method

The present research provides a meta-analytical approach based on a structured literature research investigating the Index of Cognitive Activity as a workload measure following the PRISMA statement (Moher et al., 2009). Meta-analysis is a technique consolidating the results of multiple studies on a similar topic to a single estimate of the magnitude of an effect concerning a given hypothesis. In comparison to commonly practiced narrative reviews, the advantage is a statistical outcome combining effect sizes in forms of a correlation, thus allowing for conclusions with more reliable statistical evidence. Despite well-known disadvantages such as heterogeneity, the garbage-in garbage-out problem, as well as publication or selection biases, meta-analytic results are more powerful in outlining the scope of the research domain, minimizing data waste, and following focused research compared to traditional narrations (Rosenthal & DiMatteo, 2001).

2.1. Search strategy

To analyze the Index of Cognitive Activity with a meta-analysis, a structured literature review was conducted in March 2020. The research included literature published between January 2000 and April 2021 (including early access publications) using the databases "Web of Science", "Scopus", "Pubmed" and "Google Scholar". The following search term and syntax was defined: ("index of cognitive activity" OR ICA) AND (eye* OR pupil*). Additionally, since Marshall (2000), as a first and fundamental publication, patented the Index of Cognitive Activity, a forward reference search was performed in "Google Scholar". In order to address publication bias, unpublished results were searched for on ResearchGate and asked for in message boards. Moreover, the authors of the considered studies were contacted via email. However, through this procedure, no additional studies could be identified.

2.2. Inclusion criteria

For inclusion in the meta-analysis, publications in journals, conferences, or dissertations were considered. Further, a publication was included in the meta-analysis if: 1) it dealt with empirical experiments investigating the Index of Cognitive Activity as a dependent variable, 2) a cognitive task was conducted with at least two difficulty levels, 3) sufficient information was given to determine effect size, 4) the sample included adults, 5) the sample did not include clinical samples, 6) the language was German or English, 7) the publication date was between 2000 and April 2021, 8) the statistical method applied was based on variance analysis to ensure comparability of the results. This refers to ANOVA calculations since none of the found studies compared only two conditions.

2.3. Data extraction and coding

Relevant articles were initially examined by the first author. Abstract and full paper screening as well as extraction of detailed information and coding from each article was done by two persons. The information included descriptions of the sample, statistical results (*F*-values, *df*, *p*-values, effect sizes), and experimental differences (e.g. study design, task, or response type), which were analyzed for moderating effects. Extracted data including moderators are shown in Tables B.1 to B.2 (workload) and Tables C.1 to C.2 (light).

2.4. Study quality

A scale assessing methodological quality was developed referring to previously reported checklists (Reed et al., 2007; Wells et al., 2011), since no applicable scale was available for the present purpose. Excluding and adapting items concerning treatment and control groups and clinical information required, the scale provides a brief quality overview using eight items (see Table A.1). Quality was estimated by summing weights of the items, scoring between 1 (lowest achievable score) and 10 (highest achievable score) in total. Most items were coded in two weights 1 "yes" and 0 "no" to simplify data extracting. Two items, namely publication type and sample size, were coded in three weights from 0 to 2 in increasing order. However, the scale only gives an orientation concerning quality addressing only the most obvious quality features without going into detail. It does not provide a general methodology, since metrics are simplified and items do not claim completeness or validity.

2.5. Data preparation

For one degree of freedom ($df_1 = 1$), Eq. (1) holds (Rosenthal & DiMatteo, 2001):

$$r = \sqrt{\frac{F}{F + df_2}} \quad (1)$$

where

r = Pearson's r
F = F-value
df_2 = degrees of freedom of error.

For more than one degree of freedom ($df_1 > 1$) the relation between r and F-statistics values can be derived by Eq. (2):

$$r = \sqrt{\frac{1}{1 + \frac{df_2}{F \times df_1}}} \quad (2)$$

where df_1 = degrees of freedom of treatment.

The one-tailed standard normal deviate Z of p was used if no F-statistics were given; Z-scores were determined according to Rosenthal and DiMatteo (2001) if only a range was given for p; Z was assigned with zero with a corresponding r of zero if results were not significant.

2.6. Statistical analysis

Subsequent meta-analytical statistics were calculated using R statistics (version 4.0.3) and the Metafor package (Viechtbauer, 2010) in R Studio following Quintana (2015). Hedges & colleagues' method using Fisher's z-transformation was used for calculation (Hedges & Olkin, 1985; Hedges & Vevea, 1998) proceeding with a random-effects model. The mean ICA of both eyes was prioritized over single eye ICA values when multiple ICA values were given. Otherwise, left eye ICA was designated over right eye ICA. Effect sizes were aggregated by weighted mean according to Hunter and Schmidt (2004) utilizing MAc package (Del Re & Hoyt, 2010) if multiple tasks were conducted using a within-subjects design to correct for measurement error bias in correlation. Between-study heterogeneity was quantified using the Q-value, τ^2, and I^2 statistics. Sensitivity analysis was conducted after Higgins et al. (2021) in order to assess the robustness of the results.

Moderators were analyzed by mixed-effects models running REML-estimator (Corbeil & Searle, 1976) to point out influences of between-study differences on the results. Concerning the workload analysis, task type, study design, i.e. repeated measures, number of simultaneous tasks, determination of time interval, number of factor levels, stimulus and response modalities, data normalization, eye-tracker hardware, and study quality were identified by literature analysis. When proceeding with the analysis of "hardware" as a moderator, the study of Bartels and Marshall (2012) was excluded, since they used four different devices in the evaluation. With regards to the light analysis, it was found to be of interest to investigate if the type of light source and speed of light change, as well as additional cognitive activity during the exposure respective the here used stimulus modality, or study quality influence the results of the meta-analysis.

3. Results

3.1. Study selection

The initial search identified 2022 records (including duplicates), of which 146 remained after title and abstract screening. Twelve records were retrieved for duplication appraisal. The exclusion of references was mainly due to a lack of evaluation of the Index of Cognitive Activity as an eye metric. 134 potentially eligible records were reviewed by full text, of which 120 were excluded due to missing information about data outcome or mismatching inclusion criteria, mostly because of clinical settings or linear regression analysis. Altogether 14 studies were included in the meta-analysis, of which twelve were analyzed concerning workload; four studies provided experiments concerning light sensitivity. Fig. 1 shows a flow chart of the study selection.

3.2. Descriptive statistics

14 studies with a total sample size of $N = 751$ participants were included in the meta-analysis investigating the Index of Cognitive Activity. Sample sizes range between 14 and 150 participants per study (AM = 53.6, SD = 54.2). In total 373 female, 336 male, and 0 non-binary or transgender participants were reported in the studies (k = 13). Mean age of the participants was 24.9 years, SD = 5.0 years (k = 12). Gender and mean age of participants may deviate from the actual N included, since some studies did not correct after excluding data or inclusion was not reported transparently.

3.3. Workload analysis

A total of k = 12 studies ($N = 710$ participants) were included in the meta-analysis investigating the Index of Cognitive Activity concerning workload. Sample sizes ranged between 14 and 150 participants per study (AM = 59.2, SD = 56.9). In total 352 female, 316 male, and 0 non-binary or transgender participants were reported in the studies (k = 11). Mean age reported was 25.0 years, SD = 5.4 years (k = 11). Table B.1 includes descriptive information. Statistical data considered for the analysis are shown in Table 1.

The forest plot in Fig. 2 visualizes the results of the obtained data set. Correlations and 95% CIs are reported for each study as well as the summary effect of $r = 0.61$, 95% CI [0.30, 0.80]. Results of four studies (Czerniak et al., 2021; Korbach et al., 2017, 2018; Rerhaye et al., 2018) indicate no significant correlation between workload and the ICA since the confidence intervals of these studies include zero.

Table 1
Statistical data for meta-analysis of the ICA concerning cognitive workload (k = 12).

ID	Author(s) (year)	F	df1	df2	p	r	ragg[a]
1	Bartels and Marshall (2012)	44.211	2	144	<.001***	0.62	–
2	Czerniak et al. (2021)	1.191	2.8	58.81	.32	0	–
3	Dlugosch et al. (2013)	72.07	4.49	67.39	<.001***	0.91	–
4	Fairclough et al. (2009)	7.26	6	8	<.05*	0.92	–
5	Korbach et al. (2017)	<1	n.a.	n.a.	>.05	0	–
6	Korbach et al. (2018)	2.094	2	75	>.05	0	–
7	Marshall et al. (2004)	10.67	1	21	.004**	0.58	–
8	Matthews et al. (2015)	n.a.	n.a.	n.a.	<.01*	0.19	–
9	Platten (2012)	103.23	2	34	0.001*	0.93	–
10	Reinerman-Jones et al. (2014)	n.a.	n.a.	n.a.	<.01*	0.19	–
11	Rerhaye et al. (2018)	n.a.	n.a.	n.a.	.003*	0.73	0.42
		n.a.	n.a.	n.a.	.136	0	
12	Schwalm et al. (2008)	259.17	1	14	<.001*	0.95	–

* $p < .05$, ** $p < .01$, *** $p < .001$.
[a] Aggregated correlation coefficient after Hunter and Schmidt (2004).

3.3.1. Heterogeneity and publication bias

The test for residual heterogeneity was significant $Q(11) = 134.40$, $p < .0001$, $\tau^2 = 0.2357$ (SE = 0.12), Higgins's I^2 = 92,4%, 95% CI [16.28, 75.56], indicating that variation is nearly completely reflected by actual differences in the population mean, indicating that 92.4% of variation

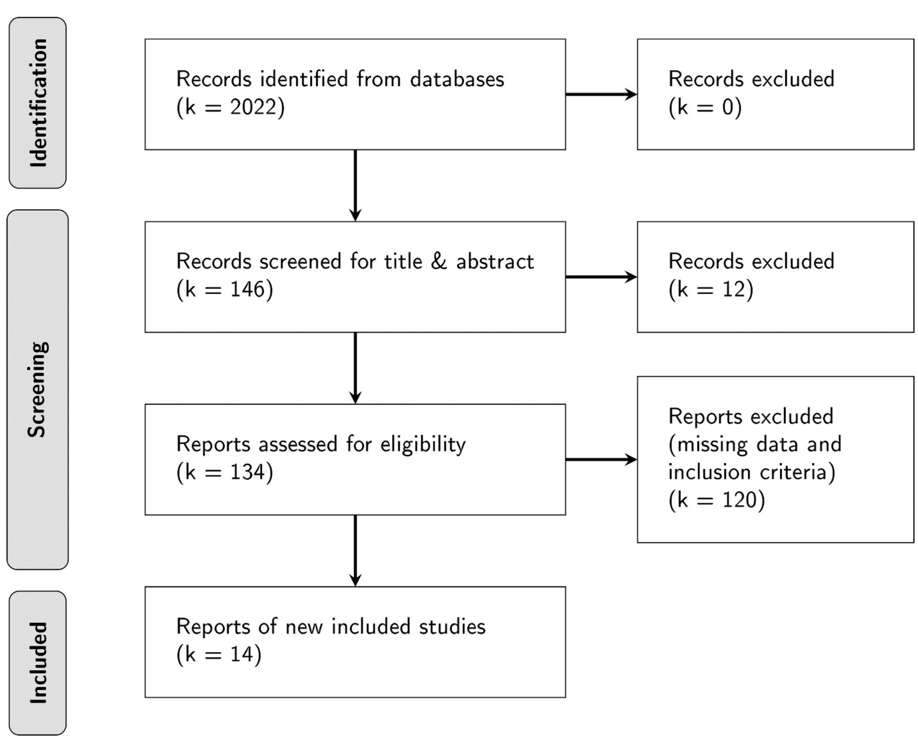

Fig. 1. Identification of relevant studies.

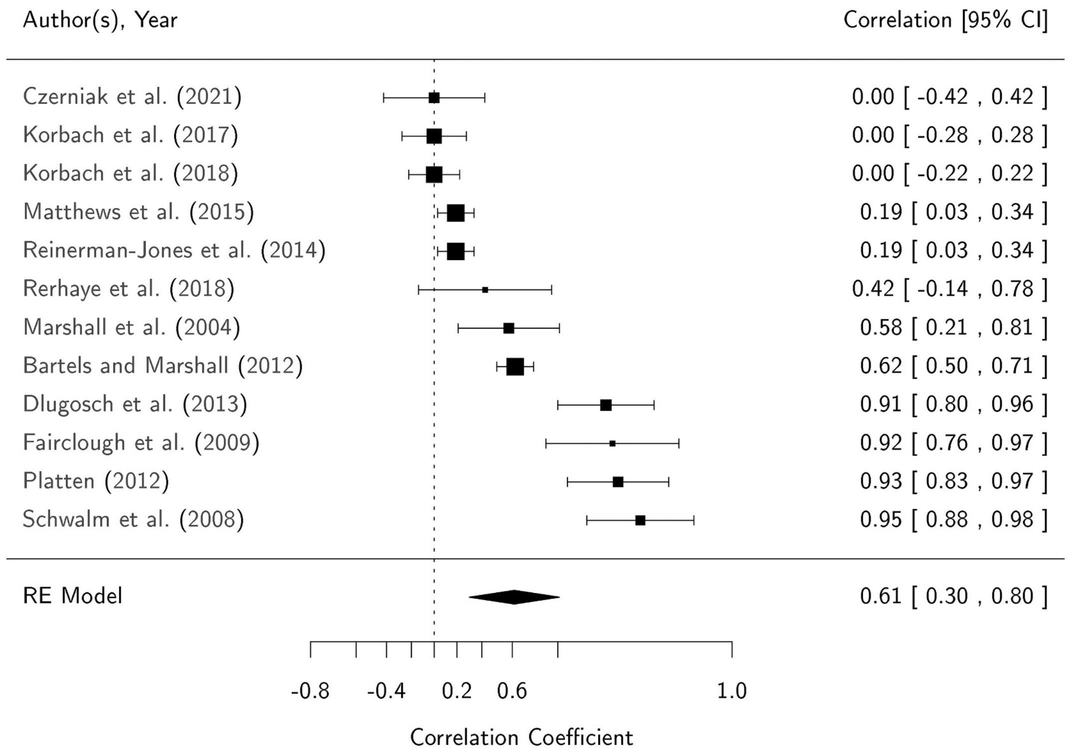

Fig. 2. Forest plot “load” of Pearson *r* correlations concerning the Index of Cognitive Activity (ICA) obtained from a random-effects model.

reflected actual differences in the included studies revealing that there is limited homogeneity in the sample size. A Baujat plot (Baujat et al., 2002) did not show any strong outliers but revealed that the study of Schwalm et al. (2008) mostly influenced heterogeneity. This finding was supported by outlier detection according to Cook (1977). Fig. B.1 shows the corresponding funnel plot. In order to assess the robustness of results, a sensitivity analysis (Higgins et al., 2021) without the study of Schwalm et al. (2008) was conducted. Results confirm robustness of the present main analysis with a significant summary effect of $r = 0.54$, 95% CI [0.23, 0.76].

However, no data were excluded from the meta-analysis. Rank correlation test (Kendall's $\tau = 0.43$, $p = 0.06$) testing for publication asymmetry was not statistically significant, hence there is no evidence of publication bias (Begg & Mazumdar, 1994; Egger et al., 1997).

3.3.2. Moderators

Investigated Moderators and results of the analysis can be derived from Table 2. A description of their characteristics is provided in Table B.2. In order to identify further sources of heterogeneity, moderating effects were investigated. The most obvious difference between the studies was the different types of tasks to be conducted for cognitive workload. Indeed, moderator analysis revealed a significant effect, $Q_M(7) = 151.46, p < .001$, with homogeneous subgroups ($Q_M(4) = 5.06$, $p = .28$). Normalization, $Q_M(2) = 48.29$, $p < .001$, and eye tracker, $Q_M(3) = 11.44$, $p = .01$, moderated significantly, but did not reveal heterogeneous subgroups ($p < .001$). It was further assumed that the pace of presented stimuli had a moderating effect on the ICA due to cognitive arousal and time pressure. However, computing the random effects model reveals that the type of pace did not have a moderating effect. Further, it was investigated whether there was a difference between visual, verbal, and haptic stimulus and response, which showed neither a moderating effect of stimulus nor a moderating effect of response type. Analyzing both the number of factor levels and study quality showed a moderating effect. Additionally, the year of

Table 2
Results of moderator analysis of cognitive workload. Mixed Effects Model (REML).

Moderator	Q_M	df_M	p_M	Q_E	df_E	p_E	I^2 [%]	τ^2
Task	151.46	7	<.001***	5.06	4	.28	0.02	0.00
Study design	2.04	1	.15	155.89	10	<.001***	95.09	0.42
Simultaneous tasks	1.14	1	.29	134.37	10	<.001***	95.57	0.46
Time interval	0.43	1	.51	110.76	8	<.001***	96.30	0.50
Stimulus	2.74	2	.25	105.49	9	<.001***	95.29	0.43
Response	5.28	2	.07	125.69	8	<.001***	94.50	0.38
Factor level	1.60	1	.21	143.54	10	<.001***	95.87	0.44
Normalization	48.29	2	<.001***	44.91	9	<.001***	76.82	0.06
Eye tracker	11.44	3	.01*	35.38	6	<.001***	92.53	0.27
Quality	4.20	1	.04*	103.82	10	<.001***	94.66	0.34
Year	4.05	1	.04*	75.79	9	<.001***	95.23	0.42
Author	0.04	9	.85	99.80	9	<.001***	93.89	0.34

Q_M = Crochan's Q of moderation, df_M = degrees of freedom of moderation, p_M = significance level of moderation, Q_E = Crochan's Q of heterogeneity, df_E = degrees of freedom of heterogeneity, p_E = significance level of heterogeneity, I^2 = Higgins's I^2, τ^2 = Tau2.
* $p < .05$, *** $p < .001$.

publication and Marshall as an author were investigated. Neither revealed moderating effects, $Q_M(1) = 4.05$, $p_M = .04$, $Q_E(9) = 75.79$, $p_E < .001$ and $Q_M(1) = 0.04$, $p_M = .85$, QE(9) = 99.80, $p_E < .001$, respectively.

3.4. Light analysis

The studies included in the meta-analysis (k = 4) involved $N = 77$ participants, ranging between 14 and 22 participants (AM = 18.6, SD = 3.6). In total 32 female, 35 male, and 0 non-binary or transgender participants were reported in the studies (k = 3). The mean age reported was 22.9 years, SD = 3.3 years (k = 3). Descriptive statistics are shown in Table C.1. Statistical data can be derived from Table 3.

The forest plot in Fig. 3 visualizes the obtained data set sorted by effect size. The plot reveals that two studies were not significant showing a confidence interval including zero (Marshall et al., 2004; Rerhaye et al., 2018). Correlations and 95% CIs are reported for each study as well as the summary effect of $r = 0.45$, 95% CI [0.05, 0.73].

3.4.1. Heterogeneity and publication bias

The test for residual heterogeneity was significant ($Q_M(3) = 9.6$, $p < .0001$, $\tau^2 = 0.14$ (SE = 0.17) Higgins's $I^2 = 68.1\%$; 95% CI 2.5, 97.61), indicating that 68.1% of variation reflected actual differences in the included studies revealing that there is limited homogeneity in the sample size. The Baujat plot (Baujat et al., 2002) reveals that the studies of Czerniak et al. (2021) and Marshall et al. (2004) contribute most to heterogeneity. However, since the sample size consisted of a very small number of studies, no sensitivity analysis excluding outliers could be conducted. Fig. C.1 shows the corresponding funnel plot. The rank correlation test (Kendall's $\tau = 0.33$, $p = .75$) was not statistically significant, thus providing no evidence of publication bias (Begg & Mazumdar, 1994; Egger et al., 1997).

3.4.2. Moderators

To explain heterogeneity, moderating effects of light change speed, light source, normalization, and quality were hypothesized. Type, $Q_M(1) = 4.24$, pM = 0.04, $Q_E(2) = 2.96$, $p_E = .23$, and quality, $Q_M(1) = 4.31$, $p_M = .03$, $Q_E(2) = 2.93$, $p_E = .23$, showed significant moderating effects. Investigated Moderators and results of the moderator analysis are shown in Table 4. Characteristics and moderator levels are shown in Table C.2.

4. General discussion

The present meta-analysis investigated the ICA as an informational workload indicator since prior research on that topic revealed contradictory findings. Based on systematic literature research, twelve studies were investigated for this purpose. Pearson's r was applied as an effect size measure due to several advantages over distance measures. Results reveal a significant relationship between informational tasks and the Index of Cognitive Activity and likely indicate its eligibility for task-evoked cognitive workload. Results of the present analysis show that the ICA seems to indicate cognitive workload in informational work tasks, similar to task-evoked pupillary response (TEPR). According to literature, both indicators rely on phasic pupil fluctuations. These are typically derived by baseline-correction of the pupil size in TEPR (Beatty & Lucero-Wagoner, 2000). This procedure does not solve for artifacts of light dilation. In contrast, the ICA algorithm calculates the phasic component through frequency analysis, ignoring the signal amplitude, and thus ignoring light dilation. This was not supported by the present analysis, which revealed an effect of light on the ICA. In summary, both indicators cope with current individual differences in pupil size and lack in providing a zero reference for cognitive workload. However, since there is little insight into the ICA algorithm, a more detailed comparison between the indicators becomes quite difficult and needs to be investigated in future research. Meta-analytic results concerning both, the ICA's sensitivity towards task-evoked cognitive workload and light influences, will be discussed in the following.

Table 3
Statistical data for meta-analysis of the ICA concerning light influences (k = 4).

ID	Author(s) (year)	F	df1	df2	p	r	ragg[a]
1	Czerniak et al. (2021)	27.10	1	17	<.001***	.78	–
2	Kahya et al. (2018)	n.a.	n.a.	n.a.	.15	.51	–
3	Marshall et al. (2004)	3.54	1	21	.07	0	–
4	Rerhaye et al. (2018)	n.a.	n.a.	n.a.	.06	0	0.35
		n.a.	n.a.	n.a.	.012*	.60	

* $p < .05$.
*** $p < .001$.
[a] Aggregated correlation coefficient after Hunter and Schmidt (2004).

4.1. Workload

Results show a high amount of heterogeneity and, thus between-study variability cannot be explained fully by sampling error but also indicates influences of study characteristics. This was supported by moderator analyses, revealing effects of the type of cognitive task, normalization of data, and the type of hardware used. Moderator analysis revealed the following findings: The tasks investigated in the analyzed experiments in particular differed with regard to the type of cognitive resources utilized and task complexity. Therefore, it seemed likely that the ICA's sensitivity depends on task characteristics and was further hypothesized to be influenced by differences in cognitive processing. For this reason, task complexity due to multiple parallel tasks was currently considered as a moderator. Surprisingly, although task type moderated the ICA, no significant influences were found for simultaneously conducted parallel tasks. Based on prior research it was further hypothesized that the ICA is not likely to distinguish between different increasing levels of workload. However, moderator analyses concerning the number of factor levels did not support this assumption. With regards to arousal, it was assumed that, for instance, time pressure is likely to intensify provided attentional resources. Included studies consisted of given time intervals ranging from several seconds to a couple of minutes (paced) and undefined time intervals (self-paced), likewise. It was therefore assumed that the characteristics of the time interval in relation to the task requirements are likely to affect results. However, this hypothesis was not supported by the present moderator analysis.

Moreover, literature on multisensory perception reveals task-dependent differences in information processing effort concerning attention allocation of multiple resources (Chan & Newell, 2008; Wahn & König, 2016). Concerning object-based tasks, the ANS activity was found to rely on separated resources concerning visual and auditory attention (Chan & Newell, 2008; Wahn & König, 2016), whereas visual and tactile modalities partially use similar resources (Dell'Acqua et al., 2001). Following this thought, stimulus and response modalities were likely to moderate the results. However, the present analysis concerning stimuli and response did not indicate any effect of processing modality on the ICA. This leads to the hypothesis that the indicator is not likely affected by perceptual workload.

Despite expected moderating effects caused by differences in information processing by the above discussed cognitive models, several methodological aspects were expected to influence the ICA as a workload indicator, namely: data used for analysis, eye-tracking hardware, and study quality. In particular, only the hardware was found to be homogenously moderating the correlation of these variables significantly. Eye-tracking hardware differs a lot in their specifications, such as device type (remote or glasses), sensor type (infrared or camera), measurement rate (Hertz), or quality. Since the present analysis included studies that used several eye trackers it was expected to find moderating effects by this variable. Indeed, moderator analysis revealed a significant effect of the hardware. However, these findings are contradicting the

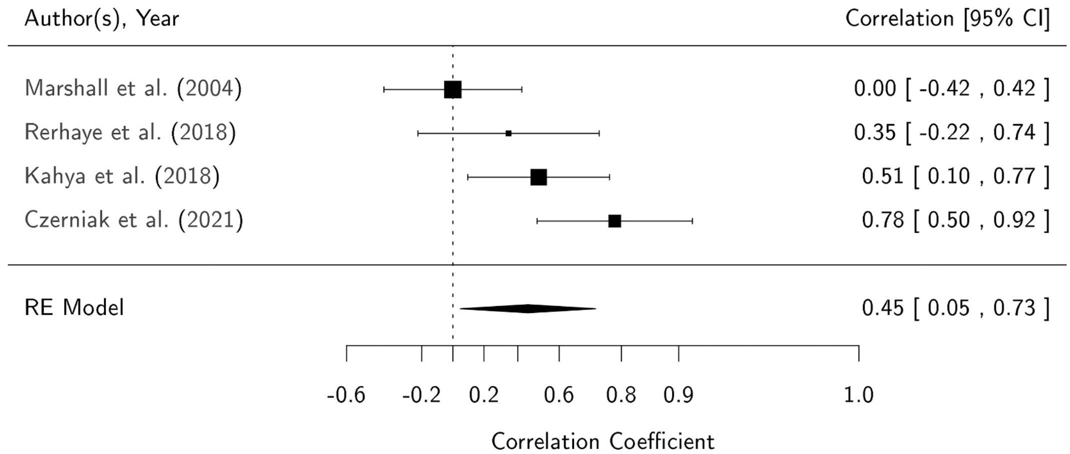

Fig. 3. Forest plot "light" of Pearson *r* correlations concerning the Index of Cognitive Activity (ICA) obtained from a random-effects model.

Table 4
Results of moderator analysis of light influences. Mixed Effects Model (REML).

Moderator	Q_M	df_M	p_M	Q_E	df_E	p_E	I^2 [%]	τ^2
Light source	1.62	2	.45	3.04	1	.08	67.09	0.16
Light change	2.55	2	.28	2.89	1	.09	65.43	0.10
Stimulus	1.06	1	.30	5.93	2	.05*	66.59	0.13
Type	4.24	1	.04*	2.96	2	.23	35.71	0.04
Quality	4.31	1	.03*	2.93	2	.23	32.48	0.03

Q_M = Crochan's Q of moderation, df_M = degrees of freedom of moderation, p_M = significance level of moderation, Q_E = Crochan's Q of heterogeneity, df_E = degrees of freedom of heterogeneity, p_E = significance level of heterogeneity, I^2 = Higgins's I^2, τ^2 = Tau2.

* $p < .05$.

results of Bartels and Marshall (2012), according to which no differences between four investigated eye trackers were shown.

Results reported in the considered studies specifically differed in data preparation. In particular the statistical analysis in most studies refers to original ICA outputs solely adjusted for blinks and measurement errors. Only a little number of studies normalized data by mean and standard deviation (z-transformation) or threshold values (minimum-maximum-normalization). But although a general moderating effect could be observed in this case, results do not indicate a general significant influence through data normalization due to inhomogeneous subgroups. This lack of homogeneity is likely to be related to a high variance in the number of subgroup items per group in the present meta-analysis.

It is well known that the quality of included studies to a meta-analysis affects the meaningfulness of the outcome effect size to a certain amount. Thus, its consideration is mandatory in such analyses. Unfortunately, only few publications were found to be suitable for the present meta-analysis in literature. For this reason, excluding studies was intentionally foregone in the present research. Consequently, it was required to monitor potential effects and to investigate study quality as a moderator. Results did not reveal a significant influence of quality in general, since despite a moderating effect was found at first, subgroups were shown to be inhomogeneous. Thus, it can be assumed that the compilation of the present studies relatively comprehensive and unbiased regarding the analysis. It was further investigated if authorship of the patent holder or publication year, for example due to improved measurement techniques, significantly influenced the results. However, neither Marshall as an author nor the publication year revealed moderating effects.

4.2. Light

The present meta-analysis of light revealed a moderate summary effect on the ICA, indicating that light is likely to affect the ICA in some circumstances despite wavelet separation. This result cannot be explained by the present results. However, it has to be considered that results are likely to be affected contrary due to dependent sample sizes of included studies with repeated measures designs. Consequently, the currently calculated effect size is expected to be slightly overestimated and is likely to be corrected downwards. However, the actual impact of the expected overestimation on the summary effect size cannot be statistically quantified further, at this point. Based on the above stated assumptions, the overall influence of light on the ICA is therefore estimated between a small and moderate effect size. Additionally, and limiting our findings, the study of Kahya et al. (2018) investigated postural demand, but since their results did not reveal a significant main effect on the factor postural control, it was concluded that significant differences were more likely to be related to effects of occluding the eyes than to cognitive postural control.

The findings for heterogeneity again reveal between-study variance and point to moderating variables. However, neither of the hypothesized factors regarding light were found to be significant moderators. Nevertheless, the eye tracker had an influence on the results, revealing that the implementation is likely to differ between hardware platforms. Since heterogeneity was relatively high, it seems likely that even more influences exist that were not considered for the present analysis or were not even reported. In summary, the present analysis could not disprove concerns about the ICA's light sensitivity.

4.3. Limitations

It has to be considered that most of the studies utilized a within-subjects design with multiple factor levels, and thus the summary effect size is likely overestimated (Bortz & Döring, 2002). To avoid further overestimation of the effect size, aggregating multiple effect sizes was prioritized over the common practice of selecting the largest effect size or averaging. Further, setting non-significant results to zero is likely to slightly counteract overestimation.

F-statistics used for this meta-analysis were extracted from reported univariate main effects of task difficulty on the ICA. Hence, interactions between several independent variables were not considered since included studies calculated multiple ANOVAs, not revealing dependencies. Therefore, any effects of other variables cannot be excluded.

Finally, the sample size is relatively small, especially in the light analysis, due to a limited number of publications applying variance analyses. Here, the problem of publication asymmetry has to be considered. When searching for unpublished results, no further studies could be identified. However, present results did not reveal evidence of publication bias. Another aspect to be considered in the light analysis is the use of a very small sample size (k < 10). This has to be outlined with regard to a possible bias of the calculated effect size in a random-effects model (Higgins et al., 2021). However, the present light analysis gives only a brief quantitative overview in advantage over a narrative review but cannot provide statistical evidence as a meta-analysis comparing bigger sample sizes.

5. Conclusion

The presented meta-analytic results on the relationship between task-evoked workload as well as light and the Index of Cognitive Activity extend the current state of knowledge regarding this indicator.

Despite the ICA's reliability for task-evoked workload and sensitivity towards multiple difficulty levels, several ambiguities and influences exist which have to be considered for the analysis. The presented findings reveal that certain cognitive processes are more likely to be represented by the indicator than others so that it seems likely that the indicator may not be an objective measure. It seems likely that the variance is due to the differences in information processing. To get a deeper insight into cognitive processes, it has to be further investigated which tasks are most suitable for workload analyses with the Index of Cognitive Activity. Moreover, the eye-tracking hardware may also play an important role in producing effect sizes for both workload and light, which cannot be explained at this point. In contrast, it does not seem to be important to consider the modality of stimulus and response, since the indicator is not likely to reflect these kinds of cognitive differences. Due to the moderating effect of normalization, a z-normalization of results can be recommended, according to the present results.

In summary, the Index of Cognitive Activity is likely to be a reliable indicator to evaluate task-evoked workload and to distinguish different workload levels, although the scale metrics and influencing factors are still ambiguous and require further research. In this regard, one problem is the lack of transparency in the calculation of the ICA algorithm. The observed moderation effect of normalization supports this assumption. In order to address this problem, another pupillary-based indicator, the Index of Pupillary Activity (IPA) has been developed by Duchowski et al. (2018), which has been shown to be sensitive to changes in cognitive workload. The calculation is based on the ICA algorithm, but uses a different wavelet analysis and a different threshold calculation, which is described in detail. Furthermore, moderators reveal that the ICA is more suitable for assessing certain types of cognitive workload rather than others, and beyond that is likely to be dependent on the hardware. However, advantages over pupillary dilation concerning light influences could not be shown through the present meta-analysis. Thus, the Index of Cognitive Activity is not exclusively preferable over pupillary measurements, when it comes to field investigations with changing light conditions.

Funding

This work was supported by 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.

Declaration of competing interests

None.

Appendix A

Table A.1
Study Quality Assessment Scale.

	Item & category	Weight
1	Publication	
	Not published	0
	Conference proceedings	1
	Peer-reviewed journal + dissertation	2
2	Method explicitly described	
	No, data missing	0
	Yes	1
3	Randomization	
	No	0
	Yes	1
4	Independent samples	
	No (within subjects)	0
	Yes (between subjects)	1
5	Study design appropriate	
	No	0
	Yes[a]	1
6	Sample size	
7	$N \leq 30$	
8	$30 < N < 100$	
9	$100 \leq N$	
10	Results explicitly described	
	No, data missing	0
	Yes, all data reported	1
11	Statistical analysis appropriate	
	No, not appropriate to study design	0
	Yes, appropriate to study design	1

[a] Study was designed either full factorial or semi factorial with relevant factors.

Appendix B. Analysis of cognitive workload

Table B.1
Descriptive statistics of included studies concerning cognitive workload (k = 12).

ID	Author(s) (year)	N	Age (AM)	Age (SD)	Age span	Male	Female
1	Bartels and Marshall (2012)	146	36	n.a.	22-55	36	75
2	Czerniak et al. (2021)	22	28	9	20-56	8	14
3	Dlugosch et al. (2013)	24	24	4	18-30	10	8
4	Fairclough et al. (2009)	14	n.a.	n.a.	19-39	11	3
5	Korbach et al. (2017)	50	22	2	n.a.	n.a.	n.a.
6	Korbach et al. (2018)	78	23	3	n.a.	n.a.	n.a.
7	Marshall et al. (2004)	22	20	n.a.	n.a.	n.a.	n.a.
8	Matthews et al. (2015)	150	20	3	n.a.	85	65
9	Platten (2012)	22	31	7	24-26	n.a.	n.a.
10	Reinerman-Jones et al. (2014)	150	20	3	n.a.	85	65
11	Rerhaye et al. (2018)	14	n.a.	n.a.	n.a.	n.a.	n.a.
			n.a.	n.a.	n.a.	n.a.	n.a.
12	Schwalm et al. (2008)	20	27	n.a.	23-33	10	10

Table B.2
Moderators of cognitive workload.

ID	Task	Design	Sim.[a]	Time interval	Stimulus	Resp.	Factor level	Norm.[b]	Eye tracker	Quality
1	MAC	B	3	P	V, A	V	3	–	EII, F5, TX, SMI	8
2	MA	W	1	S	V	V	5	–	FO	6
3	DD	W	2	P	V	T	10	z	EII	4
4	NB	W	1	S	A	T	6	Min	F5	5
5	L	B	1	P	V	–	2	–	TX	8
6	L	B	1	P	V	–	3	–	TX	8
7	MA	W	1	P	A	V	2	–	EII	6
8	DD	W	2	S	V	T	4	–	F5	9
9	CS	W	2	n.a.	V	T	3	z	EII	7
10	DD	W	2	S	V	T	4	–	F5	8
11	MR	W	1	n.a.	V	n.a.	2	–	DI	6
	S	W	1	n.a.	V	n.a.	2	–	DI	
12	FD	W	2	P	V, A	T	3	z	EII	7

Coding: Task: mental arithmetics (MA), mental arithmetics combination (MAC), driving dual task (DD), n-back (NB), learning (L), detection dual task (DD), CTT + SURT (CS), mental rotation (MR), Stroop (S); study design: between subjects (B), within subjects (W); time interval: paced (P), self-paced (S); stimulus: visual (V), auditory (A); response: verbal (V), tactile (T); eye tracker: Eye-Link II (EII), Facelab5 (F5), TobiTX 300 (TX), SMI RED 250 (SMI), Fovio (FO), Dikablis (DI).
[a] Number of simultaneous tasks.
[b] Normalization (z = z-transformation, min = Min/Max-normalization).

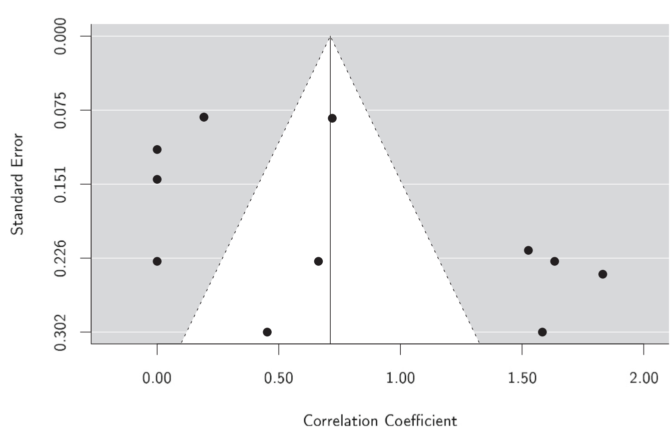

Fig. B.1. Funnel plot "load".

Appendix C. Analysis of light influences

Table C.1
Descriptive statistics of included studies concerning light influences (k = 4).

ID	Author(s) (year)	N	Age (AM)	Age (SD)	Age span	Male	Female
1	Czerniak et al. (2021)	20	26	4	19–34	11	9
2	Kahya et al. (2018)	21	n.a.	n.a.	18–29	n.a.	n.a.

(continued on next page)

Table C.1 (*continued*)

ID	Author(s) (year)	N	Age (AM)	Age (SD)	Age span	Male	Female
3	Marshall et al. (2004)	24	24	4	18–30	10	8
4	Rerhaye et al. (2018)	14	n.a.	n.a.	n.a.	n.a.	n.a.
			n.a.	n.a.	n.a.	n.a.	n.a.

Table C.2
Moderators of light influences.

ID	Light source	Light change	Stimulus	Type	Quality
1	MS	S	V	P	8
2	E	S	A	C	6
3	E	S	A	C	6
4	MS	F	V	C	6

Coding: Light source: monitor screen (MS), environment (E); light change: slowly (S), fast (F), stimulus (of task): visual (V), auditive (A); type: cognitive (C), perceptive (P).

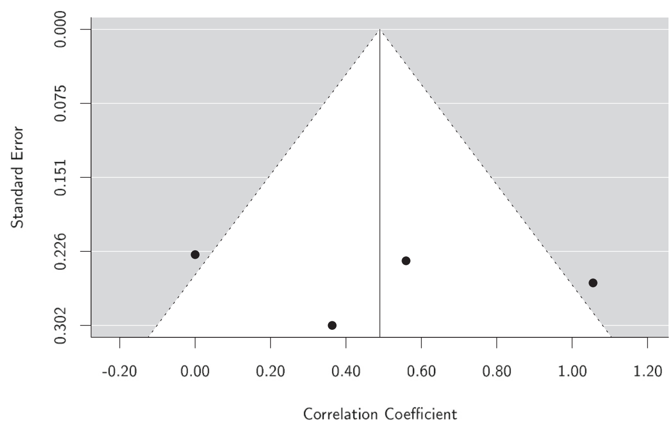

Fig. C.1. Funnel plot “light”.

Appendix D. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2021.103402.

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. *Frontiers in Psychology, 9*, 2387. https://doi.org/10.3389/fpsyg.2018.02387

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

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

Baujat, B., Mahé, C., Pignon, J. P., & Hill, C. (2002). A graphical method for exploring heterogeneity in meta-analyses: Application to a meta-analysis of 65 trials: Graphical method for exploring heterogeneity in meta-analyses. URL: *Statistics in Medicine, 21*, 2641–2652. https://doi.org/10.1002/sim.1221.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 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 Science, 5*, 371–372. https://doi.org/10.3758/BF03328444

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

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088–1101. https://doi.org/10.2307/2533446

Bortz, J., & Döring, N. (2002). *Forschungsmethoden und evaluation*. Berlin, Heidelberg: Springer-Lehrbuch, Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-07299-8

Chan, J. S., & Newell, F. N. (2008). Behavioral evidence for taskdependent "what" versus "where" processing within and across modalities. *Perception & Psychophysics, 70*, 36–49. https://doi.org/10.3758/pp.70.1.36

Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S. Z., Khawaji, A., & Conway, D. (2016). *Robust multimodal cognitive load measurement*. Cham: Springer. https://doi.org/10.1007/978-3-319-31700-7

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics, 19*, 15–18. https://doi.org/10.1080/00401706.1977.10489493

Corbeil, R. R., & Searle, S. R. (1976). Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics, 18*, 31. https://doi.org/10.2307/1267913

Czerniak, J. N., Schierhorst, N., Villani, V., Sabattini, L., Brandl, C., Mertens, A., Schwalm, M., & Nitsch, V. (2021). The index of cognitive activity - eligibility for task-evoked informational strain and robustness towards visual influences. *Applied Ergonomics, 92*, Article 103342. https://doi.org/10.1016/j.apergo.2020.103342

Debue, N., & van de Leemput, C. (2014). What does germane load mean? An empirical contribution to the cognitive load theory. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.01099

Del Re, A., & Hoyt, W. T. (2010). MAc: Meta-analysis with correlations. URL: https://CRAN.R-project.org/package=MAc.

Dell'Acqua, R., Turatto, M., & Jolicoeur, P. (2001). Cross-modal attentional deficits in processing tactile stimulation. *Perception & Psychophysics, 63*, 777–789. https://doi.org/10.3758/bf03194437

Demberg, V. (2013). Pupillometry: The index of cognitive activity in a dual-task study. In *Proceedings of the annual meeting of the cognitive science society* (p. 2159).

Demberg, V., & Sayeed, A. (2016). The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PLOS ONE, 11*, Article e0146194. https://doi.org/10.1371/journal.pone.0146194

Demberg, V., Sayeed, A., Mahr, A., & Müller, C. (2013). Measuring linguistically-induced cognitive load during driving using the ConTRe task. URL:. In *Proceedings of the 5th international conference on automotive user interfaces and interactive vehicular* *applications - AutomotiveUI '13* (pp. 176–183) Eindhoven, Netherlands: ACM Press. https://doi.org/10.1145/2516540.2516546.

Dlugosch, C., Conti, A. S., & Bengler, K. (2013). Driver distraction through conversation measured with pupillometry. In *Proceedings of the 7th international driving symposium on human factors in driver assessment, training, and vehicle design : Driving assessment 2013* (pp. 198–204). Bolton Landing, New York, USA: University of Iowa. https://doi.org/10.17077/drivingassessment.1488.

Duchowski, A. T., Kreijz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., Raubal, M., & Giannopoulos, I. (2018). The Index of Pupillary Activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In R. Mandryk, & M. Hancock (Eds.), *CHI '18: proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–13). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3173574.3173856.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*, 629–634. https://doi.org/10.1136/bmj.315.7109.629

Fairclough, S. H., Moores, L. J., Ewing, K. C., & Roberts, J. (2009). Measuring task engagement as an input to physiological computing. In *2009 3rd international conference on affective computing and intelligent interaction and workshops* (pp. 1–9). Amsterdam: IEEE. https://doi.org/10.1109/ACII.2009.5349483.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta- analysis*. Elsevier. https://doi.org/10.1016/C2009-0-03396-0

Hedges, L. V., & Vevea, J. L. (1998). In *Fixed- and random-effects models in meta-analysis* (p. 19).

Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (Eds.). (2021). *Cochrane handbook for systematic reviews of interventions*. Cochrane. Version 6.2 (updated February 2021). Available from www.training.cochrane.org/handbook.

Hunter, J., & Schmidt, F. (2004). *Methods of meta-analysis* (2nd ed.). https://doi.org/10.4135/9781412985031 (Thousand Oaks, California).

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. *Frontiers in Human Neuroscience, 12*, 288. https://doi.org/10.3389/fnhum.2018.00288

Korbach, A., Brünken, R., & Park, B. (2017). Measurement of cognitive load in multimedia learning: A comparison of different objective measures. *Instructional Science, 45*, 515–536. https://doi.org/10.1007/s11251-017-9413-5

Korbach, A., Brünken, R., & Park, B. (2018). Differentiating different types of cognitive load: A comparison of different measures. *Educational Psychology Review, 30*, 503–529. https://doi.org/10.1007/s10648-017-9404-8

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Marshall, S. P. (2000). Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity. URL: https://patents.google.com/patent/US6090051A/en.

Marshall, S. P. (2002). The index of cognitive activity: Measuring cognitive workload. In J. J. Persensky, B. Hallbert, & H. Blackman (Eds.), *Proceedings of the IEEE 7th conference on human factors and power plants* (pp. 5–9). IEEE. https://doi.org/10.1109/HFPP.2002.1042860.

Marshall, S. P. (2007). Identifying cognitive state from eye metrics. *Aviation, Space, and Environmental Medicine, 78*, Article B165–B175.

Marshall, S. P., Diego, S., Davis, C. L., & Knust, S. R. (2004). *The Index of Cognitive Activity: Estimating cognitive effort from pupil dilation* (p. 10).

Marshall, S. P., Pleydell-Pearce, C. W., & Dickson, B. T. (2003). Integrating psychophysiological measures of cognitive workload and eye movements to detect strategy shifts. In R. H. Sprague (Ed.), *Proceedings of the 36th Annual Hawaii International Conference on System Sciences HICSS03*. IEEE. https://doi.org/10.1109/HICSS.2003.1174298.

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition, 1*, 16. https://doi.org/10.5334/joc.18

Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2015). The psychometrics of mental workload: Multiple measures are sensitive but divergent. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 57*, 125–143. https://doi.org/10.1177/0018720814539505

Matton, N., Paubel, P. V., & Puma, S. (2020). Toward the use of pupillary responses for pilot selection. *Human Factors*. https://doi.org/10.1177/0018720820945163

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and metaanalyses: The PRISMA statement. PLoS medicine 6, 6. Oliva, M., 2019. Pupil size and search performance in low and high perceptual load. *PLoS medicine, 6*, 6. https://doi.org/10.3758/s13415-018-00677-w

Oliva, M. (2019). Pupil size and search performance in low and high perceptual load. *Cognitive, Affective & Behavioral Neuroscience, 19*, 366–376. https://doi.org/10.3758/s13415-018-00677-w

Orlandi, L., & Brooks, B. (2018). Measuring mental workload and physiological reactions in marine pilots: Building bridges towards redlines of performance. *Applied Ergonomics, 69*, 74–92. https://doi.org/10.1016/j.apergo.2018.01.005

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. (2012). *Analysis of mental workload and operating behavior in secondary tasks while driving*. Chemnitz, Deutschland: Technische Universität Chemnitz. Dissertation.

Quintana, D. S. (2015). From pre-registration to publication: A non-technical primer for conducting a meta-analysis to synthesize correlational data. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01549

Reed, D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E., & Wright, S. M. (2007). Association between funding and quality of published medical education research. *JAMA, 298*, 1002. https://doi.org/10.1001/jama.298.9.1002

Reinerman-Jones, L. E., Matthews, G., Barber, D. J., & Abich, J. (2014). Psychophysiological metrics for workload are demandsensitive but multifactorial. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58*, 974–978. https://doi.org/10.1177/1541931214581204

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 S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th congress of the international ergonomics association (IEA 2018)* (pp. 350–359). Cham: Springer. https://doi.org/10.1007/978-3-319-96059-3_38.

Richstone, L., Schwartz, M. J., Seideman, C., Cadeddu, J., Marshall, S., & Kavoussi, L. R. (2010). Eye metrics as an objective assessment of surgical skill. *Annals of Surgery, 252*, 177–182. https://doi.org/10.1097/SLA.0b013e3181e464fb

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52*, 59–82. https://doi.org/10.1146/annurev.psych.52.1.59

Schwalm, M. (2009). *Pupillometrie als Methode zur Erfassung mentaler Beanspruchungen im automotiven Kontext*. Saarbrücken: Universität des Saarlandes. https://doi.org/10.22028/D291-23297. Dissertation.

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.

Sekicki, M., & Staudte, M. (2018). Eye'll help you out! how the gaze cue reduces the cognitive load required for reference processing. *Cognitive Science, 42*, 2418–2458. https://doi.org/10.1111/cogs.12682

Shannon, C. E., & Weaver, W. (1949). In *A mathematical theory of communication* (p. 55).

Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., & Zhang, T. (2019). A systematic review of physiological measures of mental workload. *International Journal of Environmental Research and Public Health, 16*, 2716. https://doi.org/10.3390/ijerph16152716

Tourtouri, E. N., Delogu, F., Sikos, L., & Crocker, M. W. (2019). Rational over-specification in visually-situated comprehension and production. *Journal of Cultural Cognitive Science, 3*, 175–202. https://doi.org/10.1007/s41809-019-00032-6

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review, 25*, 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Viechtbauer, W. (2010). Conducting meta-analyses in with the metafor package. *Journal of Statistical Software, 36*. https://doi.org/10.18637/jss.v036.i03

Vogels, J., Demberg, V., & Kray, J. (2018). The index of cognitive activity as a measure of cognitive processing load in dual task settings. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.02276

Wahn, B., & König, P. (2016). Attentional resource allocation in visuotactile processing depends on the task, but optimal visuotactile integration does not depend on attentional resources. *Frontiers in Integrative Neuroscience, 10*, 13. https://doi.org/10.3389/fnint.2016.00013

Wells, G., Shea, B., O'Connell, D., Peterson, J., & Welch, V. (2011). The Newcastle-Ottawa Scale (NOS) for assessing the quality of case-control studies in meta-analyses. *European Journal of Epidemiology, 25*, 603–605.

Wu, C., Cha, J., Sulek, J., Zhou, T., Sundaram, C. P., Wachs, J., & Yu, D. (2019). Eye-tracking metrics predict perceived workload in robotic surgical skills training. *Human Factors, 62*, 1365–1386. https://doi.org/10.1177/0018720819874544