Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-020-01817-5

BRIEF REPORT

Is working memory capacity related to baseline pupil diameter?

Nash Unsworth[1] · **Ashley L. Miller**[1] · **Matthew K. Robison**[2]

Accepted: 14 September 2020

Abstract

The relation between working memory capacity (WMC) and baseline pupil diameter was examined. Participants (N = 341) performed several WMC tasks and baseline pupil diameter was measured in a dark room with a black background screen. The results indicated a weak and non-significant correlation between WMC and baseline pupil diameter consistent with some prior research. A meta-analysis of available studies (k = 26; N = 4356) similarly indicated a weak and non-significant correlation between WMC and baseline pupil diameter. Moderator analyses indicated that the primary moderator responsible for heterogeneity across studies was where the study was conducted. Studies from one laboratory tend to demonstrate a significant positive correlation, whereas other laboratories have yet to demonstrate the correlation. Broadly, the results suggest that the correlation between WMC and baseline pupil diameter is weak and not particularly robust.

Keywords Working memory · Individual differences in memory capacity

Introduction

Working memory, our ability to actively maintain and use representations for ongoing processing, is a vital component of the broader cognitive system. Variation in working memory capacity (WMC) is related to a number of other cognitive domains (e.g., Engle & Kane, 2004; Unsworth, 2016). A prominent theory of individual differences in WMC suggests that this variation is due to individual differences in attention control (or executive attention) abilities (Engle & Kane, 2004; Kane & Engle, 2002; Unsworth & Engle, 2007). Recently we extended the attention control view of WMC by suggesting that individual differences in WMC and attention control are partially driven by differences in fluctuations of attention control regulated by the locus coeruleus (LC)-norepinephrine (NE) system (LC-NE) (Unsworth & Robison, 2017a, 2017b). The LC is a brainstem neuromodulatory nucleus that is responsible for most of the NE released in the brain, and it has widespread projections throughout the neocortex including frontal-parietal areas (Berridge & Waterhouse, 2003; Szabadi, 2013). The LC-NE system seems to be particularly important for regulating arousal and alertness, which are critical for sustained attention (Aston-Jones & Cohen, 2005; Berridge & Waterhouse, 2003; Szabadi, 2013). As such, we suggested that individual differences in WMC and attention control were partially due to variation in LC-NE functioning.

To examine these issues we and others have relied on pupillometry based on prior research that has suggested that pupillary responses provide an indirect index of LC-NE functioning (Aston-Jones & Cohen, 2005; Gilzenrat et al., 2010; Joshi et al., 2016). Specifically, it is assumed that when tonic LC activity is low (hypoactive mode), individuals are inattentive and nonalert, leading to poor behavioral performance and small baseline pupils. As tonic LC activity increases to an intermediate range (phasic mode), attention becomes more focused, behavioral performance increases, and baseline pupil diameter is at intermediate levels. However, as tonic LC activity increases further, the individual experiences a more distractible attentional state, leading to task disengagement, a reduction in behavioral performance, and an increase in baseline pupil diameter.

If variation in WMC is related to LC-NE functioning, then we might expect there to be a relation between baseline pupil diameter and WMC, with low WMC individuals having either smaller or larger baseline pupil diameters than high WMC individuals (e.g., Unsworth & Robison, 2017a).

✉ Nash Unsworth
nashu@uoregon.edu

[1] Department of Psychology, University of Oregon, Eugene, OR 97403, USA

[2] Department of Psychology, Arizona State University, Phoenix, AZ, USA

Unfortunately, prior research is decidedly mixed on whether there is a relation between WMC and baseline pupil diameter. For example, Heitz et al. (2008) found a positive relation between WMC and baseline pupil diameter measured both before the task (pre-task baseline) and during the task (pre-trial baseline). Heitz et al. suggested that the ability to control attention was likely related to overall arousal levels. More recently, Tsukahara et al. (2016) replicated these results, finding a positive relation between WMC and baseline (both pre-task and pre-trial) pupil diameter. Tsukahara et al. suggested that this relation was likely due to variation in LC-NE functioning. While these studies found evidence for a positive relation, other studies have found different results. For example, Unsworth and Robison (2017b) found a negative relation between WMC and pre-trial baseline pupil diameter measured in two attention control tasks. Furthermore, Unsworth et al. (2019) recently found a weak and non-significant relation between WMC and pre-task baseline pupil diameter. Likewise, Aminihajibashi et al. (2019) recently found a weak and non-significant relation between WMC and pre-task baseline pupil diameter (although they did find a relation with variability in baseline). Thus, while some studies suggest some evidence for a positive relation between WMC and baseline pupil diameter, other studies suggest no relation between the two.

Given the theoretical importance of a possible relation, the current study examined whether WMC is related to baseline pupil diameter. To address this issue, we (1) conducted a new high-powered study, and (2) conducted a meta-analysis of prior studies to get a better sense of the relation between WMC and baseline pupil diameter. The new study served to replicate and extend prior research by addressing some limitations from our prior study (Unsworth et al., 2019). Specifically, in our prior study we measured baseline pupil diameter for 5 min by having participants stare at a black square on a grey screen in a dimly lit room. One issue with measuring baseline pupil diameter in a dimly lit room with a light grey screen is that our baseline pupil diameters were much smaller (*M* = 3.21 mm, *SD* = .49) than prior studies that found a positive relation (e.g., *M* = 5.92 mm, *SD* = 1.09 in Tsukahara et al., 2016). Thus, it is possible that differences in luminance across the studies resulted in an inability to find a relation. Therefore, in the current study baseline pupil diameter was measured by having participants stare at a white cross on a black screen in a dark room, which should result in overall larger pupil diameters and potentially resulting in a positive correlation with WMC. Furthermore, in our prior study (Unsworth et al., 2019) we argued that it is important to assess what participants are thinking about during the baseline measure and whether this influences relations between baseline pupil diameter and WMC. Finally, in our prior study we suggested that future research is needed to assess whether individual differences in motivation and interest in the current tasks are related to baseline pupil diameter and potentially influence the relation with WMC. To examine these issues, participants were asked to indicate what they were thinking about during the baseline eye measure immediately following the baseline measure. Following all of the tasks, participants were asked to indicate their overall levels of motivation, alertness, effort, and interest in the experimental session.

Method

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in our study.

Participants

A total of 341 participants were recruited from the subject-pool at the University of Oregon, a comprehensive state university. Participants were 68% female, between the ages of 18 and 34 years (*M* = 19.32, *SD* = 1.80), and received course credit for their participation. Each participant was tested individually in a laboratory session lasting approximately 2 h. We aimed to have a minimum sample size of 300 participants. With this sample size we have power of .80 (alpha set at .05 two tailed) to detect correlations of $r = .16$ and we have power of .99 to detect correlations of $r = .25$. Participants were not specifically screened for history of psychiatric/neurological disorders, medication, or substance use. Participants were allowed to wear glasses or contacts. Data will be made available on the Open Science Framework.

Materials and procedure

After signing informed consent, all participants completed the operation span, symmetry span, reading span, baseline eye measure, an attention control measure, and several long-term memory tasks, and then filled out the post-experimental questionnaire. The attention control and long-term memory measures were part of another project and are not discussed here.

Baseline eye measure

Participants saw a white cross on a black background in the center of the screen. Participants were instructed to simply stare at the cross. Specifically, participants were told "Please just stare at the white +. Please do not avert your eyes from the screen and do not close your eyes. Although you may blink normally." The task lasted for 30 s. Pupil diameter and eye gaze were continuously recorded binocularly at 120 Hz using a Tobii T120 eye-tracker, integrated in a 17-in. TFT monitor. Participants were seated approximately 60 cm from the monitor with the aid of chinrest in a dark room (illuminance = 1 lux). Missing data points due to blinks, off-screen fixations, and/or eyetracker malfunction were removed and not included in the pupil averages.

Baseline eye-measure questionnaire

Immediately following the baseline eye measure, participants completed a brief questionnaire asking what they were thinking about during the eye task. Specifically, participants were asked to characterize what they were thinking about on the baseline measure by pressing one of six keys. Participants saw:

Please press a number on the keyboard.

1. I was totally focused on the task
2. I was thinking about the task
3. I was distracted by sights/sounds/temperature or by physical sensations (hungry/thirsty)
4. I was daydreaming/my mind was wandering about things unrelated to the task
5. My mind was blank
6. I was drowsy and not very alert

Due to a programming error responses for the first 172 participants were not recorded. Thus, data for this measure are only available for the final 169 participants.

Working memory capacity (WMC) tasks

Operation span Participants solved a series of math operations while trying to remember a set of unrelated letters (see Redick et al., 2012; Unsworth et al., 2005). Participants were required to solve a math operation, and after solving the operation, they were presented with a letter for 1 s. Immediately after the letter was presented the next operation was presented. At recall participants were asked to recall letters from the current set in the correct order by clicking on the appropriate letters. For all of the span measures, items were scored correct if the item was recalled correctly from the current list. Participants were given practice on the operations and letter-recall tasks only, as well as two practice lists of the complex, combined task. List length varied randomly from three to seven items, and there were two lists of each list length for a maximum possible score of 50. The score was total number of correctly recalled items.

Symmetry span Participants recalled sequences of red squares within a matrix while performing a symmetry-judgment task (see Redick et al., 2012; Unsworth et al., 2009). In the symmetry-judgment task, participants were shown an 8 × 8 matrix with some squares filled in black. Participants decided whether the design was symmetrical about its vertical axis. The pattern was symmetrical half of the time. Immediately after determining whether the pattern was symmetrical, participants were presented with a 4 × 4 matrix with one of the cells filled in red for 650 ms. At recall, participants recalled the sequence of red-square locations by clicking on the cells of an empty matrix. Participants were given practice on the symmetry-judgment and square-recall task as well as two practice lists of the combined task. List length varied randomly from two to five items, and there were two lists of each list length for a maximum possible score of 28. We used the same scoring procedure as we used in the operation span task.

Reading span While trying to remember an unrelated set of letters, participants were required to read a sentence and indicated whether or not it made sense (see Redick et al., 2012; Unsworth et al., 2009). Half of the sentences made sense (e.g., “Spring is her favorite time of year because flowers begin to bloom”), while the other half did not (“Even though she was in trouble, she managed to go to the dice and shop”). Nonsense sentences were created by changing one word in an otherwise normal sentence. After participants gave their response, they were presented with a letter for 1 s. At recall, participants were asked to recall letters from the current set in the correct order by clicking on the appropriate letters. Participants were given practice on the sentence judgment task and the letter recall task, as well as two practice lists of the combined task. List length varied randomly from three to seven items, and there were two lists of each list length for a maximum possible score of 50. We used the same scoring procedure as we used in the operation span and symmetry span tasks.

Post-experimental questionnaire

At the end of the experimental session participants completed a brief questionnaire asking about their general level of motivation, alertness, effort, and interest during the entire experimental session. Specifically participants were asked: “How motivated were you, in general, to perform well on the tasks administered during this experimental session?”, “How alert, overall, were you during the tasks administered during this experimental session?”, How much effort, in general, did you put into your performance on the tasks administered during this experimental session?”, and “How interested, in general, were you in the tasks administered during this experimental session?” Participants responded on a 1–6 scale.

Results and discussion

Table 1 shows the descriptive statistics for all measures. Consistent with prior research, we created a WMC

Table 1. Descriptive statistics for all measures

Measure	M	SD	Skew	Kurtosis	Reliability	N
Baseline Pupil	4.89	.54	-.81	.91	.96	328
Baseline PupilSD	.27	.11	.37	-.24	.95	328
WMC	00.00	.89	-.88	.71	.70	331
Focused	.38	.49	.51	-1.77	--	169
TRI	.37	.48	.56	-1.71	--	169
ED	.05	.21	4.30	16.70	--	169
MW	.08	.27	3.20	8.36	--	169
MB	.10	.30	2.68	5.24	--	169
Drowsy	.01	.11	9.11	81.95	--	169
Motivation	4.66	1.02	-.51	-.29	--	310
Alertness	3.83	1.04	-.08	-.24	--	310
Effort	4.77	.94	-.49	-.17	--	310
Interest	3.90	1.23	-.26	-.45	--	310

Baseline Pupil mean baseline pupil diameter, *Baseline PupilSD* standard deviation of baseline pupil diameter, *WMC* working memory capacity factor composite, *Focused* focused on baseline eye task, *TRI* task-related interference on baseline eye task, *ED* external distraction on baseline eye task, *MW* mind-wandering on baseline eye task, *MB* mind-blanking on baseline eye task, *Drowsy* drowsy on baseline eye task, *Motivation* motivated during experimental session, *Alertness* alertness during experimental session, *Effort* effort during experimental session, *Interest* interest during experimental session

composite given that the three working memory span measures were correlated (Operation span – Symmetry span r = .36; Operation span – Reading span r = .63; Symmetry span – Reading span r = .33). The composite WMC score was computed for each participant using principal axis factoring and allowing the three tasks to load onto a single factor. The resulting factor loadings for Operation span, Symmetry span, and Reading span were .82, .43, and .77, respectively. As can be seen in Table 1, the measures had generally acceptable values of internal consistency and most of the measures were approximately normally distributed. The exceptions were the various forms of off-task thoughts in the baseline eye questionnaire due to low response rates for many of the categories. Mean and standard deviation of baseline pupil diameter were similar to several prior reports (Aminihajibashi et al., 2019; Bornemann et al., 2010; van der Meer et al., 2010), especially under similar luminance conditions (e.g., Aminihajibashi et al., 2019; Winn et al., 1994). Importantly, overall mean pupil diameter was larger in the current data than in Unsworth et al. (2019), $t(526) = 35.68$, $p < .001$, $d = 3.26$. Thus, testing participants in a dark room with a black background served to increase overall baseline pupil diameter.

Examining responses to the questionnaires suggested that during the 30-s pre-task baseline measure participants were generally focused on staring at the fixation cross or thinking about the overall task (task-related interference). Fewer participants reported various off-task thoughts such as being distracted, mind-wandering, mind-blanking, or being drowsy. Examining the post-experimental questionnaire suggested participants were reasonably motivated, alert, put effort into the tasks, and were interested in the tasks during the experimental session.

Correlations among the measures are presented in Table 2. The full correlation table is presented for completeness. As can be seen, the only significant correlation with mean baseline pupil diameter was standard deviation of baseline pupil diameter (r = -.21). Critically, WMC and baseline pupil

Table 2 Correlations among the measures

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Baseline Pupil	--												
2. Baseline PupilSD	**-0.21**	--											
3. WMC	0.06	0.05	--										
4. Focused	-0.13	0.10	-0.09	--									
5. TRI	0.11	0.02	0.14	**-0.59**	--								
6. ED	0.00	-0.10	-0.05	**-0.17**	**-0.17**	--							
7. MW	0.05	-0.04	0.11	**-0.23**	**-0.22**	-0.06	--						
8. MB	0.05	-0.09	-0.10	**-0.26**	**-0.26**	-0.08	-0.10	--					
9. Drowsy	-0.15	-0.05	-0.02	-0.09	-0.08	-0.02	-0.03	-0.04	--				
10. Motivation	-0.11	0.04	0.11	0.11	-0.01	0.12	**-0.17**	-0.09	-0.03	--			
11. Alertness	-0.06	0.10	0.06	**0.19**	-0.02	-0.04	-0.12	-0.12	-0.06	**0.57**	--		
12. Effort	-0.07	0.02	0.10	0.15	-0.06	0.04	-0.11	-0.10	0.02	**0.66**	**0.54**	--	
13. Interest	-0.09	0.04	-0.04	0.06	-0.05	0.06	0.05	**-0.21**	-0.06	**0.56**	**0.47**	**0.48**	--

Correlations significant at the $p < .05$ level are bolded diameter were not correlated (r = .06).[1] In fact, computing the Bayes factor for this relation suggested that the evidence was more consistent with the null (BF_{01} = 8.01). Additionally, unlike Aminihajibashi et al. (2019), but consistent with Unsworth et al. (2019), WMC was unrelated to variability in baseline pupil diameter (r = .05; BF_{01} = 10.40). Furthermore, baseline pupil diameter was unrelated to what participants were thinking about during the baseline eye measure and was unrelated to self-reports of overall motivation, alertness, effort, and interest during the experimental session.

The results of the current study were straightforward. Measuring baseline pupil diameter in a dark room with a black background resulted in an increase in pupil diameter compared to our prior work (Unsworth et al., 2019), but this did not result in a positive correlation with WMC. Thus, it seems unlikely that the lack of a relation in Unsworth et al. (2019) was due to luminance conditions for the baseline eye measure. Overall, the current results are consistent with prior studies suggesting no relation between WMC and baseline pupil diameter (Aminihajibashi et al., 2019; Unsworth et al., 2019), but are inconsistent with studies suggesting a positive relation (Heitz et al., 2008; Tsukahara et al., 2016).

Meta-analysis

To get a better sense of the relation between WMC and baseline pupil diameter we next conducted a meta-analysis of available studies.

Method

Study selection

We identified studies by searching through PsycINFO and Google Scholar databases using the keywords “working memory,” “baseline pupil diameter,” “pupil,” and “short-term memory.” Other studies were identified by searching through the references of prior studies examining WMC and pupil diameter, as well as recent research by individuals who we were aware of conducting similar analyses. Several studies were identified in which both WMC and pupil diameter were examined, but the correlations were not reported. In these cases, the authors were contacted and asked to provide the specific correlations, if possible. Our criteria for inclusion in the study were (1) the study had to measure WMC, (2) the study had to measure baseline pupil diameter (either pre-task or pre-trial), (3) the study had to primarily sample young adults (e.g., age < 36 years) given that pupil is related to age (e.g., Birren et al., 1950; Winn et al., 1994), and (4) the study had to assess the full range of participants (i.e., not just examine extreme groups). Extreme-groups designs can be problematic for a number of reasons (Preacher, Rucker, MacCallum, & Nicewander, 2005). For example, when only the top and bottom portions of the distribution are examined, a great deal of information is lost, as the entire middle of the distribution has been excluded. Additionally, although extreme-groups designs are known to increase the ability to detect an effect, these designs can also lead to an increased likelihood of making a Type I error as a result of overestimated effect sizes (Conway et al., 2005; Preacher et al., 2005; Unsworth et al., 2015). Given these issues and given that there were a number of studies that utilized a full range of participants, the extreme groups studies were excluded from the main meta-analysis.[2] If anyone has relevant data, please contact the authors so that we can update the meta-analysis.

Design and analyses

The main analyses of interest were meant to specify the magnitude of the correlation between WMC and baseline pupil diameter. For studies with multiple measures of WMC and available data, a WMC factor composite was computed similar to the current study. When data were not available, the correlations were averaged together to deal with sample dependence issues. We meta-analyzed the studies using a random-effects analysis (which assumes there are meaningful differences across studies) to estimate the mean-weighted correlation coefficients along with the 95% confidence interval for the mean weighted correlations. We also examined heterogeneity of the correlations and examined whether the heterogeneity could be accounted for by the moderator variable using mixed-effects meta-analysis modeling. All analyses were conducted with the Major package in jamovi (Hamilton, 2018).

---

[1] The quadratic effect was also not significant (β = .06, p = .51).

[2] Because two of the initial studies that argued for a relation between baseline pupil diameter and WMC used extreme groups designs (Heitz et al., 2008; Tsukahara et al., 2016) we thought it was important for completeness to rerun the analyses with these studies included as well as data from Unsworth et al. (2004) which was previously reported in Unsworth et al. (2019). Descriptive information for all studies is provided in Table 3. With the extreme groups designs added, the number of samples increased to 30 with a total of 4822 participants. The overall meta-analytic average correlation was .03, 95% CI [-.01, .08], p = .15, indicating that the meta-analytic correlation was not significantly different from zero. Only 6 of the 30 (20%) effect sizes were significantly different from zero and one was in the opposite direction. Thus, even with the extreme groups studies included, there was still little evidence for a relation between baseline pupil diameter and WMC. The I^2 statistic was large (59.73%) and Q was also large and significant, $Q(29)$ = 74.55, p < .001. Similar to the primary meta-analysis we conducted a post-hoc moderator analysis comparing correlations from Georgia Tech vs. everywhere else. The meta-analytic correlation for Georgia Tech was .19, 95% CI [.13, .26] and the meta-analytic correlation for everywhere else was -.01, 95% CI [-.05, .02], and this difference was significant, p < .001. Additionally, when including Georgia Tech vs. everywhere else in the model as a moderator resulted in I^2 = 0.01% and $Q(29)$ = 33.53, p = .22, suggesting that this moderator accounted for the heterogeneity across studies.

Moderator variables

In an effort to examine sources of heterogeneity in the literature, we conducted moderator analyses using the following variables:

Pre-task versus pre-trial We examined whether the type of baseline measurement influenced the relation.

Complex span versus other WMC task We examined whether the type of WMC tasks (complex span vs. storage only tasks) influenced the relation.

Light versus dark room We examined whether the lighting conditions of the room influenced the relation.

White, grey, or black background color We examined whether using different background screen colors (white, grey, or black) influenced the relation.

Results and discussion

Descriptive information for all studies is provided in Table 3. As can be seen, there was considerable variability in the correlations, in baseline pupil diameter, in the measures used to represent WMC, and in the overall lighting conditions. Twenty-six samples met our inclusion criteria consisting of 4,356 participants. The overall meta-analytic average correlation was .01, 95% CI [-.03, .06], p = .63, indicating that the meta-analytic correlation was not significantly different from zero. Figure 1 shows the forest plot of the studies. As can be seen, only three of the 26 effect sizes (12%) were significantly different from zero and one was in the opposite direction. Thus, there was little evidence for a relation between baseline pupil diameter and WMC. Figure 2 shows the funnel plot of the studies. Egger's linear regression was not significant (p = .082), suggesting that there was not significant publication bias.

Examining heterogeneity across the effect sizes suggested that there was quite a bit of heterogeneity. Specifically, the I^2 statistic, which indicates the percentage of between-study variability in the effect sizes due to heterogeneity and not random error, was large (51.04%). The Q statistic (which also gives an indication of heterogeneity) was similarly large and significant, $Q(25) = 57.70$, $p < .001$. To examine this heterogeneity we conducted moderator analyses.

First, we examined whether measuring baseline pupil diameter pre-task versus pre-trial mattered. The meta-analytic correlation for Pre-task baselines (k = 11) was .04, 95% CI [-.04, .12] and the meta-analytic correlation for Pre-trial baselines (k = 15) was -.02, 95% CI [-.06, .03], and this difference was not significant, p = .20. Next, we examined whether the type of WMC measure (complex span vs. storage only) influenced the relation. The meta-analytic correlation for complex span WMC measures (k = 22) was .01, 95% CI [-.04, .05], and the meta-analytic correlation for storage only WMC measures (k = 4) was .04, 95% CI [-.07, .15], and this differences was not significant, p = .56. Next, we examined whether the lighting conditions of the room (light vs. dark room) influenced the relation. Note, Tsukahara and Engle (2020) manipulated lighting conditions within participants in their Experiment 2, so this moderation analysis included both the average correlation for the light and the average correlation for the dark room conditions. Thus, this study accounts for two effect sizes that are actually dependent. Including only the light condition or including or the dark room condition resulted in nearly identical results. The meta-analytic correlation for baselines measured in a light room (k = 14) was .03, 95% CI [-.05, .10], and the meta-analytic correlation for baselines measured in a dark room (k = 11) was .03, 95% CI [-.02, .07], and this difference was not significant, p = .92. Next, we examined whether using different background screen colors (white, grey, or black) influenced the relation. Similar to the light versus dark room analysis, these analyses are complicated by the fact that in both their Experiment 1 and Experiment 2, Tsukahara and Engle (2020) manipulated monitor background conditions within participants. Therefore, we included the correlations for the white, grey, and black conditions for each experiment. Thus, this study accounts for two effect sizes per experiment that are actually dependent. The meta-analytic correlation for baselines measured with a white background (k = 5) was .00, 95% CI [-.10, .11], the meta-analytic correlation for the grey background (k = 15) was .02, 95% CI [-.03, .08], the meta-analytic correlation for the black background (k = 8) was .02, 95% CI [-.09, .12], and these differences was not significant, p = .33.

In analyzing the data it became clear that much of the heterogeneity across studies was likely due to where the study was conducted. That is, studies conducted by Tsukahara, Engle, and colleagues at Georgia Tech tended to demonstrate larger relations than the other studies. Thus, we conducted a post hoc moderator analysis comparing correlations from Georgia Tech versus everywhere else. The meta-analytic correlation for Georgia Tech (k = 3) was .18, 95% CI [.11, .26], and the meta-analytic correlation for everywhere else (k = 23) was -.02, 95% CI [-.05, .02], and this difference was significant, p < .001. Thus, correlations obtained from Georgia Tech tended to be positive and significant, whereas correlations obtained from other laboratories were near zero and not significant. It should, however, be noted that only two of the effect sizes were significant for Georgia Tech. Additionally,

Table 3 Descriptive information for each study

Study	*N*	*R*	*M (SD)*	Tasks	Room	Screen	Pupil measure
*Unsworth et al. (2004)	145	.06	6.08(1.16)	O	Dark	Black, white cross	Pre-task
*Heitz et al. (2008)	167	.24		O	Dark	Black, white cross	Pre-task
Bornemann et al. (2010)	34	.16	4.53(.54)	D	Dim	Grey	Pre-task
Unsworth & Robison (2015)	70	.12	2.79(.31)	VC	Dim	Grey, black cross	Pre-trial
*Tsukahara et al. (2016) E1	40	.49	6.2 est	O, S, R	Dim	Black, white cross	Pre-task
*Tsukahara et al. (2016) E2	114	.28	6 est	O, S, Rot	Dim	Black, grey cross	Pre-task
Tsukahara et al. (2016) E3	337	.24	5.92(1.09)	O, S, Rot	Dim	Black, grey cross	Pre-task
Unsworth & Robison (2017b)	143	-.15	2.59(.28)	O, S, R	Dim	White, black cross	Pre-trial
Sibley et al. (2018)	79	-.28	5 est	O		Black	Pre-task
Unsworth & Robison (2018)	124	-.07	2.75(.32)	VC	Dim	Grey, black cross	Pre-trial
Unsworth et al. (2019)	204	.01	3.21(.49)	O, S, R	Dim	Grey, black cross	Pre-task
Miller et al. (2019) E1	138	.09	3.33(.38)	O, S, R	Dim	Grey, black cross	Pre-trial
Miller et al. (2019) E2	128	-.13	3.52(.49)	O, S, R	Dim	Grey, black cross	Pre-trial
Aminihajibashi et al. (2019)	212	-.06	4.4(.69)	Let Num		Grey, black cross	Pre-task
Robison & Unsworth (2019)	107	.08	3.1(.46)	VC	Dark	Grey, black cross	Pre-trial
Unsworth et al. (in press) E2	140	-.01	2.72(.32)	O, S, R	Dark	White, black cross	Pre-trial
Miller & Unsworth (in press) E1	122	-.10	3.5(.49)	O, S, R	Dark	Grey, black cross	Pre-trial
Miller & Unsworth (in press) E2	134	-.02	3.31(.47)	O, S, R	Dark	Grey, black cross	Pre-trial
Hutchinson et al. (2020) E1	108	.03	4.05(.53)	O	Light	Black, white cross	Pre-trial
Hutchinson et al. (2020) E2	83	-.15	4.23(.52)	O	Light	Black, white cross	Pre-trial
Current Study	328	.05	4.89(.54)	O, S, R	Dark	Black, white cross	Pre-task
Miller & Unsworth (2020) E2	146	.09	3.33(.47)	O, S, R	Dark	Grey, black cross	Pre-trial
Unsworth et al. (2020) E1	151	.02	4.95(.67)	O, S, R	Dark	Black, white cross	Pre-trial
Unsworth et al. (2020) E2	149	-.01	4.74(.60)	O, S, R	Dark	Black, white cross	Pre-trial
Ralph et al. (2020)	231	-.03	3.63(.51)	O, S, Rot	Dim	Grey, black asterisks	Pre-task
Christopher (2019)	226	-.08	1024[a](351)	O, S	Light	White	Pre-task
Robison & Brewer (2020a)	252	.05	4.4(.79)	O, S, R	Dark	Grey, black cross	Pre-task
Robison & Brewer (2020b)	204	-.01	4.64(.67)	O, S, R	Dark	Grey, black cross	Pre-trial
Tsukahara & Engle (2020) E1	310	.17[b]	4.3(.67)	O, S, Rot	Dim	Mixed	Pre-task
Tsukahara & Engle (2020) E2	196	.10[c]	4.64(.71)	O, S, Rot	Mixed	Mixed	Pre-task

*These studies used an extreme-groups methodology and are not included in the main meta-analysis. See Footnote 1

[a] Pupil diameter is based on arbitrary units

[b] This is the average correlation across two conditions

[c] This is the average correlation across eight conditions

O operation span, *S* symmetry span, *R* reading span, *D* digit span, *Rot* rotation span, *VC* visual arrays color, *Let Num* letter–number sequencing task

including Georgia Tech versus everywhere else in the model as a moderator resulted in I^2 = 0.03% and $Q(25)$ = 25.87, p = .36, suggesting that this moderator accounted for the heterogeneity across studies. Because quite a bit of the effect sizes came from our laboratory at the University of Oregon, we also ran a moderator analysis contrasting effects from University of Oregon, Georgia Tech, and everywhere else. The meta-analytic correlation for University of Oregon (k = 14) was .00, 95% CI [-.04, .04], the meta-analytic correlation for Georgia Tech (k = 3) was .18, 95% CI [.11, .26], and the meta-analytic correlation for everywhere else (k = 9) was -.04, 95% CI [-.09, .02], and these differences were significant, p = .02. Contrasting the correlations suggested that the difference between University of Oregon and Georgia Tech was significant, p < .001, as was the difference between Georgia Tech and everywhere else, p < .001. The contrast between University of Oregon and everywhere else was not significant, p = .28. Therefore, heterogeneity across the studies was due to the fact that some studies conducted at Georgia Tech were associated with a significant positive correlation, whereas studies conducted at other laboratories were associated with essentially no correlation.

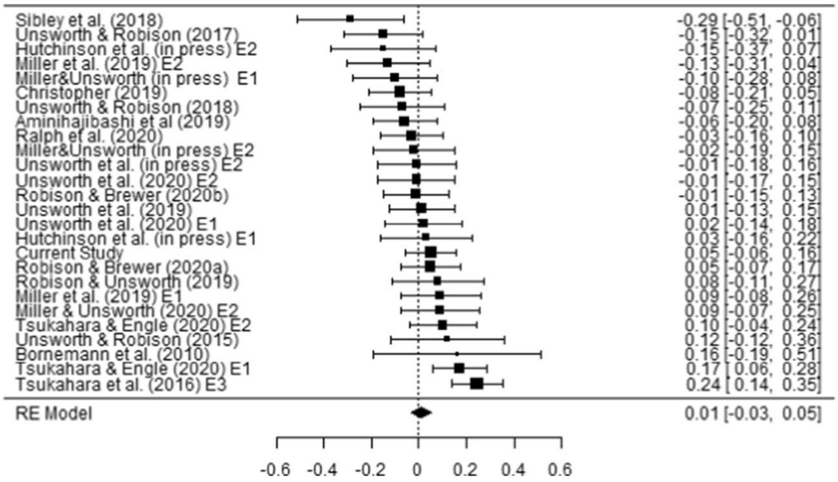

Fig. 1 Forest plot depicting correlations for all of the studies included in the meta-analysis. Note the squares represent the correlations and the lines represent 95% confidence intervals. The diamond at the bottom represents the average meta-analytic correlation. *RE* random-effects model

General discussion

The current study examined whether there is a correlation between WMC and baseline pupil diameter. The overall results were relatively straightforward. A new study where baseline pupil diameter was measured in the dark with a black background screen resulted in a larger baseline pupil diameter than our previous report (Unsworth et al., 2019), but the correlation between WMC and baseline pupil diameter was small and non-significant. Conducting a meta-analysis of available studies resulted in 26 effect sizes with over 4,000 participants that met our inclusion criteria. The meta-analytic correlation between WMC and baseline pupil diameter was small and non-significant. Including studies that relied on an extreme groups design resulted in similar results. A number of moderator analyses suggested that type of baseline measurement, type of WMC task, room lighting, and background screen color did not moderate the relation. However, the moderator analyses suggested that heterogeneity across studies was largely due to where the study was conducted. Studies conducted at Georgia Tech tended to demonstrate significant correlations (two of three studies), whereas none of the studies conducted at other laboratories found a significant positive relation. These results suggest that, broadly, there is little to no relation between baseline pupil diameter and WMC. But, there are somewhat consistent findings from studies conducted at Georgia Tech suggesting a positive relation between WMC and baseline pupil diameter. What are we to make of these discrepancies?

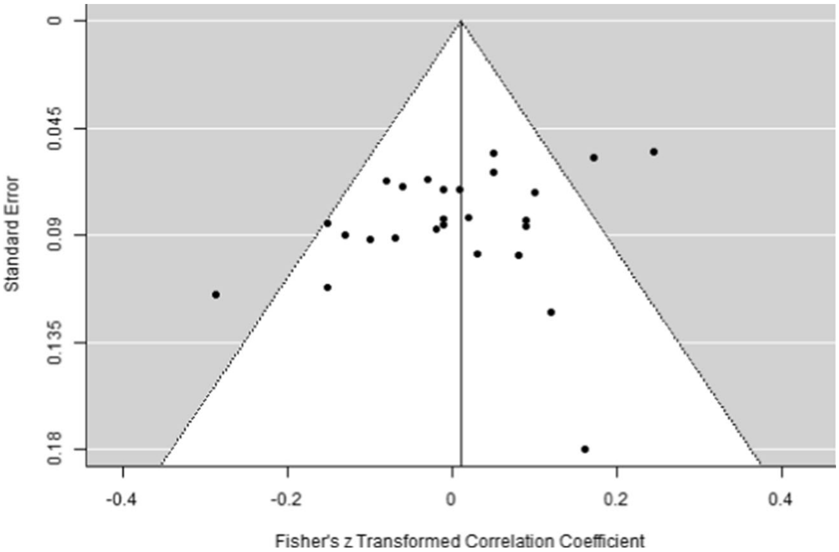

Fig. 2 Funnel plot for correlations between working memory capacity and baseline pupil diameter. Correlations are plotted against standard error. The vertical line represents the population effect size estimate and the diagonal lines represent 95% confidence intervals

First, we note that although the meta-analytic correlation for Georgia Tech was significant, samples from this laboratory do not always find significant relations. For example, in their recent study examining how luminance might impact relations with baseline pupil diameter, Tsukahara and Engle (2020) found that only two of ten possible correlations between baseline pupil diameter and WMC were significant, and only one of these correlations was greater than .15. Thus, even studies from the Georgia Tech laboratory suggest that the correlation is not always robust. Second, Heitz et al. (2008) noted that their correlation was partially due to the fact that age was associated with both baseline pupil diameter and WMC, and when age was partialled out the correlation was reduced (although still significant). Similar reductions in correlations when partialling out age are found with Tsukahara et al. (2016; see Unsworth et al., 2019) and Tsukahara and Engle (2020; although the partial correlations are still significant). In the current study age was correlated with baseline pupil diameter (-.17), but not with WMC (-.05; see also Unsworth et al., 2019). Thus, at least some (but not all) of the variance in the relation seen for the Georgia Tech samples seems to be due to shared variance with age. It is currently unclear what other factors may be accounting for relations seen with the Georgia Tech samples versus everywhere else. Given the overall small (near zero) meta-analytic correlation, future research on this topic should ensure that a very large number of participants are tested to ensure there is sufficient power to detect such a small effect.

While the current study focused on the relation between WMC and baseline pupil diameter, we note that Tsukahara, Engle, and colleagues (Tsukahara et al., 2016; Tsukahara & Engle, 2020) have suggested that there is a stronger and more consistent relation between fluid intelligence and baseline pupil diameter. As such, Tsukahara, Engle, and colleagues have suggested that the baseline pupil diameter to fluid intelligence relation is more important and should be studied. While Tsukahara, Engle, and colleagues have consistently found a relation between baseline pupil diameter and measures of fluid intelligence, other studies have not found such a relation. For example, of the studies listed in Table 3, seven of the studies also had measures of fluid intelligence. In addition to Tsukahara et al. (2016) and Tsukahara and Engle (2020), Bornemann et al., (2010) also found a relation between fluid intelligence and baseline pupil diameter. However, three studies found no relation between fluid intelligence and baseline pupil diameter (Ralph et al., 2020; Robison & Brewer, 2020a; Unsworth & Robison, 2017b). Thus, similar to the relation between WMC and baseline pupil diameter, the relation between fluid intelligence and baseline pupil diameter is not always consistently demonstrated. To get a better sense of this relation, we computed the meta-analytic correlation between fluid intelligence and baseline pupil diameter in these studies along with some unpublished data (Diede & Bugg, 2020). The overall meta-analytic average correlation ($k = 8$; $N = 1585$) was .14, 95% CI [.002, .273], $p = .047$, indicating that the meta-analytic correlation was significantly different from zero. This is by no means a comprehensive meta-analysis, but it does suggest that there is some weak evidence indicating a small positive relation between baseline pupil diameter and fluid intelligence consistent with Tsukahara, Engle, and colleague's claims. However, we note that this relation is not always demonstrated even with fairly large sample sizes and the meta-analytic correlation was barely significant. Thus, more research is needed to examine the robustness of this potentially important relation.

The current results have implications for theories suggesting a role of the LC-NE system in individual differences in WMC and attention control. In particular, the lack of a relation between WMC and baseline pupil diameter suggests that it is unlikely the case that low WMC individuals simply are under-aroused or have lower tonic LC activity than high WMC individuals (Unsworth & Robison, 2017a). Rather, the extent to which baseline pupil diameter provides an indirect index of LC tonic activity suggests a weak and near-zero relation between tonic activity and WMC. A more fruitful line of inquiry may be to examine task-evoked pupillary responses (Unsworth & Robison, 2017a; Unsworth et al., in press).

Overall, the current results are very much in line with conclusions from Unsworth et al. (2019) suggesting that the correlation between WMC and baseline pupil diameter is weak and not particularly robust.

References

References marked with an asterisk indicate studies included in the primary meta-analysis.

*Aminihajibashi, S., Hagen, T., Foldal, M. D., Laeng, B., & Espeseth, T. (2019). Individual differences in resting-state pupil size: Evidence for association between working memory capacity and pupil size variability. *International Journal of Psychophysiology, 140,* 1–7.

Aston-Jones G., & Cohen J.D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience, 28,* 403-450.

Berridge C.W., Waterhouse B.D. (2003) The locus coeruleus-noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews, 42*, 33-84.

Birren, J. E., Casperson, R. C., & Botwinick, J. (1950). Age changes in pupil size. *Journal of Gerontology, 5,* 216-221.

*Bornemann, B., Foth, M., Horn, J., Ries, J., Warmuth, E., Wartenburger, I., & van der Meer, E., (2010). Mathematical cognition – Individual differences in resource allocation. *The International Journal of Mathematics Education,42,* 555-567.

*Christopher, E.A. (2019). *Using pupillometry to observe covert mental activity during prospective memory tasks.* Unpublished Doctoral Dissertation.

Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., & Engle, R.W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12,* 769-786.

Diede, N.T., & Bugg, J.M. (2020). Unpublished data.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.). *The psychology of learning and motivation* (Vol. 44, pp. 145-199). NY: Elsevier.

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience, 10,* 252–269.

Hamilton, W.K. (2018). *MAJOR: Meta Analysis JamOvi R*. For the jamovi project.

Heitz, R. P., Schrock, J. C., Payne, T. W., & Engle, R. W. (2008). Effects of incentive on working memory capacity: Behavioral and pupillometric data. *Psychophysiology, 45,* 119–129.

*Hutchison, K.A., Moffitt, C.C., Hart, K., Hood, A.V.B., Watson, J.M., & Marchak, F.M. (2020). Measuring task set preparation versus mind wandering using pupillometry. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 46,* 280-295.

Joshi, S., Li, Y., Kalwani, R.M., & Gold, J.I. (2016). Relationship between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron, 89,* 221-234.

Kane, M.J. & Engle, R.W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual differences perspective. *Psychonomic Bulletin & Review, 9,* 637 – 671.

*Miller, A.L., & Unsworth, N. (2020). *Attending to encode: The role of consistency and intensity of attention in learning ability.* Manuscript submitted for publication.

Miller, A. L., Gross, M. P., & Unsworth, N. (2019). *Individual differences in working memory capacity and long-term memory: The influence of intensity of attention to items at encoding as measured by pupil dilation*. Journal of Memory and Language, 104, 25–42.

*Miller, A.L., & Unsworth, N. (in press). Variation in attention at encoding: Insights from pupillometry and eye gaze fixations. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Preacher, K.J., Rucker, D.D., MacCallum, R.C., & Nicewander, W.A. (2005). Use of extreme groups approach. A critical reexamination and new recommendations. *Psychological Methods, 10,* 178-192.

*Ralph, K.J., Gibson, B.S., & Gondoli, D.M. (2020). *The Contribution of Goal Maintenance and Alertness to Attentional Control and Their Relation to Working Memory, Fluid Intelligence, and Lapses.* Manuscript submitted for publication.

Redick, T.S., Broadway, J.M., Meier, M.E., Kuriakose, P.S., Unsworth, N., Kane, M.J., & Engle, R.W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28,* 164-171.

*Robison, M. K., & Brewer, G. A. (2020a). *Individual differences in arousal, working memory capacity, attention control, and fluid intelligence.* Manuscript in preparation.

*Robison, M. K., & Brewer, G. A. (2020b). *Individual differences in working memory capacity and the regulation of arousal.* Manuscript submitted for publication.

*Robison, M.K., & Unsworth, N. (2019). Pupillometry tracks fluctuations in working memory performance. *Attention, Perception, & Psychophysics, 81,* 407-419.

*Sibley, C., Foroughi, C., Brown, N., & Coyne, J. T. (2018). Low Cost Eye Tracking: Ready for Individual Differences Research? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 741–745.

Szabadi, E. (2013). Functional neuroanatomy of the central noradrenergic system. *Journal of Psychopharmacology, 27,* 659-693.

*Tsukahara, J. S., & Engle, R. W. (2020). *The locus coeruleus-norepinephrine system and fluid intelligence.* Manuscript submitted for publication.

*Tsukahara, J.S., Harrison, T.L., & Engle, R.W. (2016). The relationship between baseline pupil size and intelligence. *Cognitive Psychology, 91,* 109-123.

Unsworth, N. (2016). The many facets of individual differences in working memory capacity. In B. Ross (Ed.). *The psychology of learning and motivation, 65, 1-46.*

Unsworth, N., & Engle, R.W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114,* 104-132.

Unsworth, N., Heitz, R.P., Schrock, J.C., & Engle, R.W. (2005) An automated version of the operation span task. *Behavior Research Methods*, *37,* 498-505.

*Unsworth, N., Miller, A.L., & Robison, M.K. (2020). *Oculometric indicators of individual differences in preparatory control during the antisaccade.* Manuscript in preparation.

*Unsworth, N., Miller, A.L., & Robison, M.K. (in press). Individual differences in lapses of sustained attention: Oculometric indicators of intrinsic alertness. *Journal of Experimental Psychology: Human Perception and Performance.*

Unsworth, N., Redick, T.R., McMillan, B.D., Hambrick, D.Z., Kane, M.J., & Engle, R.W. (2015). Is playing videogames related to cognitive abilities? *Psychological Science, 26,* 759-774.

Unsworth, N., Redick, T.S., Heitz, R.P., Broadway, J., & Engle, R.W. (2009). Complex working memory span tasks and higher-order cognition: A latent variable analysis of the relationship between processing and storage. *Memory, 17,* 635-654.

*Unsworth, N. & Robison, M.K. (2015). Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. *Psychonomic Bulletin & Review, 22,* 757-765.

Unsworth, N., & Robison, M.K. (2017a). A Locus Coeruleus-Norepinephrine account of individual differences in working memory capacity and attention control. *Psychonomic Bulletin & Review, 24,* 1282-1311.

*Unsworth, N., & Robison, M.K. (2017b). The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 43,* 1962-1987.

*Unsworth, N. & Robison, M.K. (2018). Tracking working memory maintenance with pupillometry. *Attention, Perception, & Psychophysics, 80, 461-484.*

*Unsworth, N., Robison, M. K., & Miller, A. L. (2019). Individual differences in baseline oculometrics: Examining variation in baseline pupil diameter, spontaneous eye blink rate, and fixation stability. *Cognitive, Affective, & Behavioral Neuroscience, 19,* 1074-1093.

Unsworth, N., Schrock, J.C., & Engle, R.W. (2004) Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1302-1321.

van der Meer, E., Beyer, R., Horn, J., Foth, M., Bornemann, B., Ries, J., Kramer, J., Warmuth, E., Heekeren, H., & Wartenburger, I. (2010). Resource Allocation and Fluid Intelligence: Insights from Pupillometry. *Psychophysiology, 47,* 158-169.

Winn, B., Whitaker, D., Elliott, D. B., & Phillips, N. J. (1994). Factors affecting light-adapted pupil size in normal human subjects. *Investigative Ophthalmology & Visual Science, 35,* 1132–1137.