Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-023-02273-7

BRIEF REPORT

Baseline pupil diameter does not correlate with fluid intelligence

Matthew K. Robison[1] · **Stephen Campbell**[1]

Accepted: 16 March 2023

Abstract

There has been debate regarding the correlation between baseline/resting state measures of pupil diameter and cognitive abilities such as working memory capacity and fluid intelligence. A positive correlation between baseline pupil diameter and cognitive ability has been cited as evidence for a role of the locus coeruleus-norepinephrine (LC-NE) and its functional connection with cortical networks as a reason for individual differences in fluid intelligence (Tsukahara & Engle, *Proceedings of the National Academy of Sciences*, *118*(46), e2110630118, 2021a). Several recent attempts to replicate this correlation have failed. The current studies make another attempt and find substantial evidence against a positive correlation between pupil diameter and intelligence. Given the data from the current studies in combination with other recent failures to replicate, we conclude that individual differences in baseline pupil diameter should not be used as evidence for a role of the LC-NE system in goal-directed cognitive activity.

Keywords Pupil · Intelligence · Working memory

Introduction

Working memory and fluid intelligence are two hallmarks of the human cognitive system. Working memory is the system that allows people to maintain access to and manipulate multiple pieces of information, including task goals, often when other information competes for attention, whereas fluid intelligence is the ability to reason and solve novel, abstract problems. There are robust individual differences in both working memory capacity and fluid intelligence, even among healthy adults. Therefore, attempts have been made to understand the neurobiological basis for these individual differences. Two recent studies have demonstrated moderate and robust correlations among resting pupil diameter, fluid intelligence, and working memory capacity (Tsukahara et al., 2016; Tsukahara & Engle, 2021b; see also Heitz et al., 2008). Given the link between the locus coeruleus-norepinephrine (LC-NE) system and pupil diameter (Alnæs et al., 2014; Joshi et al., 2016; Joshi & Gold, 2020; Murphy et al., 2014; Reimer et al., 2016), Tsukahara and Engle (2021a) propose that there are individual differences in the functional organization of the resting state brain, particularly the connection between the LC and the higher-level cortical networks that implement goal-directed cognition. According to the theory, better functional organization leads to greater resting pupil diameter, better attention control, and higher fluid intelligence. The authors have publicized this finding with a piece in *Scientific American* magazine (Tsukahara et al., 2021), and it has received independent media coverage in *Discover Magazine* (Learn, 2021).

However, several recent studies have failed to replicate these findings. Unsworth et al. (2021) published a meta-analysis finding a near-zero correlation between working memory capacity and resting pupil diameter across 21 studies. In a sample of over 4,500 individuals, Coors et al. (2022) found near-zero correlations between resting pupil size and working memory, episodic memory, and executive function. However, they did find a significant positive correlation between processing speed and resting pupil diameter. Robison and Brewer (2022) measured resting pupil diameter and cognitive abilities (working memory, attention control, and fluid intelligence) in a sample of 252 young adults and found null correlations between each of the abilities and pupil diameter. Finally, Robison, Coyne, et al. (2022a) measured cognitive ability and resting pupil diameter in a sample of 845 members of the US military

---

✉ Matthew K. Robison
matthew.robison@uta.edu

[1] Department of Psychology, University of Texas at Arlington, Arlington, TX, USA

and found no correlation. Thus, multiple independent attempts to replicate the findings have all failed to do so.

In response, Tsukahara and Engle (2021b) have argued that the failures to replicate a resting pupil diameter-cognition relation suffer from at least two methodological weaknesses. First, they argue that lighting conditions systematically affect both average pupil diameters and the amount of interindividual variability in the sample, creating a range restriction issue. Specifically, under dark conditions the pupil is maximally dilated and the most interindividual variability is observed. Thus, dark conditions are ideal for an examination of individual differences, and therefore the lighting conditions under which resting pupil size has been measured in prior work may be suboptimal for detecting a correlation. Second, much of the prior work has examined individual differences in working memory capacity. Tsukahara et al., (2021) argue that the correlation between fluid intelligence and pupil diameter is particularly robust, and more research should focus on the pupil-intelligence correlation. Indeed, their theory regarding individual differences in the LC system and intelligence uses this correlation as evidence (Tsukahara & Engle, 2021a). Tsukahara et al., (2021) also point to a few additional factors that may produce replication failures, such as an ability-restricted, homogenous sample (i.e., all college students), small sample sizes (<200), and using only single tasks to measure either working memory capacity or fluid intelligence.

Challenging these claims are the recent studies by Coors et al. (2022), Robison, Coyne, et al. (2022a); and Robison and Brewer (2022). Coors et al. (2022)'s sample was large (over 20 times the size recommended by Tsukahara et al., 2021) and age- and ability-diverse. Further, they measured pupil diameter against a black background in a dark room, and cognitive ability was measured with 11 different tasks. Therefore, they met many of the recommendations laid out by Tsukahara et al., (2021). Robison, Coyne, et al. (2022a); Robison, Trost, et al. (2022b) had a large and ability-diverse sample from four different military occupations, sampled specifically because they differ in the cognitive testing scores required for their roles. They also measured pupil diameter in three different external lighting conditions, none of which correlated with cognitive ability. However, Robison, Coyne, et al. (2022a); had only one measure of fluid intelligence in their study. Robison and Brewer (2022) measured fluid intelligence at the factor level with three tasks. But their study was exclusively university participants, and it only measured pupil diameter against a bright gray background. Despite some weaknesses of the three studies cited above, the diversity in sample characteristics and adherence to the general recommendations in these studies challenge the claims made by Tsukahara et al., (2021).

In another attempt to resolve these potential limitations, we recently administered two individual differences investigations of working memory capacity, fluid intelligence, and resting pupil diameter. In both studies, we systematically manipulated lighting conditions to test whether the correlation between pupil diameter and cognitive ability systematically differs based on such conditions.

Study 1

Study 1 was an individual-differences investigation of working memory capacity, working memory precision, and fluid intelligence. As a secondary goal, we collected resting pupil data at the beginning of the session to test for a pupil diameter-intelligence relation.

Method

Participants and procedure

After exclusions, the sample included 190 participants (130 identified as female, 56 as male, three as non-binary, and one as other gender; *M* age = 19 years [*SD* = 2]; 58% of participants identified as White, 42% as Hispanic or Latino, 23% as Black/African American, 15% as Asian, 3% as Native American, and 1% as Native Hawaiian/Pacific Islander; 149 participants indicated that English was their first language). The target sample size was 200 participants, and we used the end of an academic term as the stopping point for data collection. All participants were undergraduate students at the University of Texas at Arlington, participating in exchange for partial course credit. Over the course of a 2-h session, participants completed three change-detection measures of working memory capacity, three continuous-report measures of working memory precision (not analyzed here), and three measures of fluid intelligence. Participants' pupils were measured at the beginning of the session. The experimental protocol was approved by the Institutional Review Board of the University of Texas at Arlington.

Tasks

The color, orientation, and letter change-detection tasks were used as measures of working memory capacity, and the Raven Advanced Progressive Matrices, number series, and letter sets tasks were used as measures of fluid intelligence.

Color change detection (Luck & Vogel, 1997) Each trial started with a 1,000-ms fixation screen with a black fixation cross centered against a gray background. Then, six target items appeared for 250 ms. The target items were colored squares that subtended 3° of visual angle each. The stimuli appeared in six preselected locations spaced equally around the center of the screen. Colors were sampled randomly from a continuous HSV color space, with the requirement that each color was at least 30° apart in the HSV space. After a 1,000-ms blank delay, the items reappeared with one item circled by a black ring. The participants' task was to indicate whether this item was the same color or a different color as the initial presentation using the 'F' and 'J' keys on a keyboard to indicate "same" or "different," respectively. The next trial began after a 500-ms blank intertrial interval. Participants completed six practice trials with accuracy feedback, then 100 experimental trials without accuracy feedback. The dependent variable was the capacity estimate *k* (6 * [hit rate – false alarm rate]; maximum score = 6).

Orientation change-detection (Luck & Vogel, 1997) Each trial started with a 1,000-ms fixation screen with a black fixation cross centered against a gray background. Then, six target items appeared for 250 ms. The target items were black oriented bars that subtended 3° of visual angle each. The stimuli appeared in six preselected locations spaced equally around the center of the screen. Orientations were sampled randomly from a continuous orientation space (0–180°), with the requirement that each orientation was at least 30° apart in the space. After a 1,000-ms blank delay, the items reappeared with one item circled by a white ring. The participants' task was to indicate whether this item was the same orientation or a different orientation as the initial presentation using the 'F' and 'J' keys on a keyboard to indicate "same" or "different," respectively. The next trial began after a 500-ms blank intertrial interval. Participants completed six practice trials with accuracy feedback, then 100 experimental trials without accuracy feedback. The dependent variable was the capacity estimate *k* (6 * [hit rate – false alarm rate]; maximum score = 6).

Letter change-detection (Robison & Brewer, 2020) Each trial started with a 1,000-ms fixation screen with a black fixation cross centered against a gray background. Then, six target items appeared for 250 ms. The target items were letters that subtended 3° of visual angle each. The stimuli appeared in six preselected locations spaced equally around the center of the screen. Letters were sampled randomly without replacement from the set of English consonants. After a 1,000-ms blank delay, the items reappeared with one item surrounded by a black box. The participants' task was to indicate whether this item was the same letter or a different letter as the initial presentation using the 'F' and 'J' keys on a keyboard to indicate "same" or "different," respectively. The next trial began after a 500-ms blank intertrial interval. Participants completed six practice trials with accuracy feedback, then 100 experimental trials without accuracy feedback. The dependent variable was the capacity estimate *k* (6 * [hit rate – false alarm rate]; maximum score = 6).

Raven Advanced Progressive Matrices (Raven et al., 1962) On each trial, a 3 × 3 patterned matrix appeared with the bottom-right piece of the pattern missing. The participants' task was to select from a set of eight possible options the piece that best completed the implicit pattern(s) in the matrix. Participants had a maximum of 10 min to complete as many of the 18 odd-numbered problems as possible (maximum score = 18).

Number series (Thurstone, 1938) On each trial, a sequence of numbers appeared, and the participants' task was to select from a set of five possible options the number that best continued the sequence. Participants had 4.5 min to complete as many trials as possible, with a maximum possible score of 15.

Letter sets (Ekstrom & Harman, 1976) On each trial, a set of four different four-letter sets appeared. Among the sets, three of the four sets followed an implicit rule, and one of the four sets violated this rule. The participants' task was to select the set of letters that violated the rule. Participants had 5 min to complete as many trials as possible, with a maximum possible score of 20.

Baseline pupil measurement

Participants were seated about 60 cm from a computer screen with a Gazepoint GP3HD eye-tracker mounted to the bottom of a 1,920 × 1,080-px monitor. A researcher verified that the eye-tracker was tracking both eyes and that the participant was seated at the appropriate distance from the monitor. Participants' head positions were not fixed with a chinrest. However, participants were instructed to maintain their head position for the measurement interval. For the first measurement, the lights in the room were off, with no light entering the room except for the computer monitors, backlit keyboards, and the experimenter's desktop computer, toward which participants' backs were faced. Participants were administered in a group run room with a maximum of three participants per session. The researcher had the participants start the baselining procedure simultaneously.

The fixation screens showed a black screen (RGB: [0, 0, 0]) for 30 s, followed by a 5-s transition to a gray screen (RGB: [100, 100, 100]) for 30 s, followed by a 5-s transition to a white screen (RGB: [200, 200, 200]). The fixation cross also changed color to provide contrast against the background. For the black screen it was white (RGB: [255, 255, 255]); for the gray screen it was light gray (RGB: [155, 155, 155]); and for the white screen it was dark gray (RGB: [55, 55, 55]). After the first baseline was completed, the researcher turned the lights on, and the procedure repeated. Thus, there were six total measures for each participant (2 lighting conditions × 3 backgrounds fully crossed within subjects). The conditions were motivated by the conditions delivered by Tsukahara and Engle (2021b). We measured the amount of light in each condition using a Sper Scientific Direct Light Meter Lux 840006. These measurements are listed for each condition and study in Table 1. As can be seen in Table 1, both manipulations affected the amount of light in the environment. We also listed the values measured by Tsukahara and Engle (2021b) for comparison.

Data analysis

The data were analyzed in R (R Core Team, 2022) using the *tidyverse* (Wickham et al., 2019), *data.table* (Dowle & Srinivasan, 2020), *cowplot* (Wilke, 2020), *lavaan* (Rosseel, 2012), *lmerTest* (Kuznetsova et al., 2017), and *psych* (Revelle, 2022) packages. The data and analysis scripts are available on the Open Science Framework (https://osf.io/569bt/). Pupil data from the right eye were used. The Gazepoint API automatically flags samples as invalid, and these samples were excluded. We also excluded samples outside a plausible physiological range (<1 mm or >9 mm). We based this criterion on recommendations by Mathôt et al. (2018), who note 2 mm and 8 mm as the lower and upper limits of typical pupil sizes. We extended these bounds to 1 mm and 9 mm, respectively, to account for the fact that we used both very bright (white background) and very dark (black background) measurement conditions. Data were averaged over the last 25 s of each 30-s window to account for the pupillary light reflex.

Table 1 Lux readings by condition and study

Study	Lights	Background	Participant	Screen
Present Study 1	Off	Black	1	1
		Gray	4	33
		White	23	131
	On	Black	97	1
		Gray	105	34
		White	121	130
Present Study 2	Off	Black	1	1
		Gray	7	44
		White	25	123
	On	Black	173	1
		Gray	180	43
		White	192	123
Tsukahara and Engle (2021b)	Off	Black	1	8
Study 2		White	27	208
	On	Black	44	8
		White	83	208

Note. Values were given in lux units. Participant values were measured by placing the spectrometer on the forehead of a researcher, facing the monitor at a ~60-cm distance from the eye-tracker. Screen values were measured by placing the spectrometer directly on the computer monitor. Values from Tsukahara and Engle (2021a, 2021b) are from those listed for their "bright" monitor conditions

To test our questions of interest, we first computed a fluid intelligence *z*-score composite by averaging standardized Raven, number series, and letter sets scores and a working memory *z*-score composite by averaging standardized color *k*, orientation *k*, and letter *k* estimates. Then, we specified a linear mixed effect model with fixed effects of lights (on vs. off), background color (black vs. gray vs. white), the fluid intelligence *z* composite, and their interactions. Intercepts were set to vary randomly across participants.

For the latent variable analysis, the three change-detection tasks were specified to load onto a factor, Raven, number series, and letter sets were set to load onto a factor, and the six pupil measurements were set to load onto a factor. To obtain adequate model fit, the error variances between the lights off/black background and lights off/gray background, between the lights off/gray background and lights off/white background, and between the lights on/gray background and lights on/white background measures were allowed to correlate.

For all measures, reliability was estimated using an odd/even split and applying the Spearman-Brown split-half correction to the correlation between the halves.

Exclusions

For the change-detection tasks, participants were excluded if they had a negative *k* estimate. Negative *k* estimates can be due to either a mismapping of response keys, true guessing, or a failure to understand instructions.[1] Additionally, our scheduling system allowed participants to complete both Study 1 and Study 2. This created an unintended consequence of partially overlapping samples. Therefore, we used the demographic data and time/date stamps to determine which study a participant completed first. Then, we only used data from their first study, as there may be practice effects on the Raven, number series, and letter sets tasks. In some cases, we could not determine which study a participant completed first, and in those cases the participant was excluded listwise.

---

[1] No participants who were included in the analyses for Study 1 had negative *k* values.

Table 2 Descriptive statistics for all measures in Study 1

Measure	*N*	Mean	SD	Skew	Kurtosis	Reliability
Color *k*	182	3.77	0.96	-0.80	0.73	0.78
Orientation *k*	182	2.84	1.18	-0.38	-0.47	0.83
Letter *k*	182	3.03	1.19	-0.48	-0.27	0.82
Raven	182	8.57	3.38	-0.11	-0.56	0.72
Number series	187	7.20	2.84	0.10	-0.44	0.76
Letter sets	171	8.91	3.05	0.05	-0.61	0.79
Pupil - lights off, black	190	6.51	1.03	-0.83	1.92	>0.99
Pupil – lights off, gray	189	4.23	0.82	0.70	1.28	>0.99
Pupil – lights off, white	186	3.46	0.61	1.79	7.13	>0.99
Pupil – lights on, black	189	4.35	0.90	1.14	1.65	>0.99
Pupil – lights on, gray	184	3.86	0.67	1.25	2.92	>0.99
Pupil – lights on, white	184	3.41	0.47	0.72	1.92	>0.99

Note. SD = standard deviation, *k* = capacity estimate

Results

Descriptive statistics are listed in Table 2, and zero-order correlations among measures are listed in Table 3. The first analysis examined whether the correlations among resting pupil diameter, fluid intelligence, and working memory capacity were moderated by the external lighting conditions. The model is summarized in Table 4. As would be expected, pupil diameter was smaller with the lights on and smaller against the gray and white backgrounds compared to the black background. However, there was not a significant main effect of fluid intelligence, nor any significant interactions between the lighting factors and fluid intelligence. Therefore, fluid intelligence did not correlate with pupil diameter overall, and the correlation was not moderated by lighting condition. This pattern of data is plotted in Fig. 1a. The distributions of pupil diameter against each background are plotted in Fig. 2.

Next, we specified the linear mixed effect model with the working memory capacity *z* composite. The model is summarized in Table 4. There was not a significant main effect of working memory capacity, nor any significant interactions between the lighting factors and working memory capacity. Therefore, working memory capacity did not correlate with pupil diameter, and the correlation was not moderated by lighting condition. This pattern of data is plotted in Fig. 1b.

The final analysis specified a confirmatory factor analysis of the relations among working memory capacity, fluid intelligence, and pupil diameter. The model, which is depicted in Fig. 3, fit the data well, $\chi^2(48) = 97.22$, CFI = .96, TLI = .94, RMSEA = .07 90% CI [.05, .09], SRMR = .06.[2] As was expected working memory capacity and fluid intelligence highly correlated, but pupil diameter correlated with neither working memory capacity nor fluid intelligence.

The absence of evidence for a correlation does not necessarily mean evidence against a correlation. Therefore, we compared models using a χ^2 goodness-of-fit comparison and Bayes Factors. First, we specified a "null" model in which the correlations between the working memory capacity and pupil diameter factors and between the fluid intelligence and pupil diameter factors were set to equal zero. Then, we specified a model which freed a correlation between the working memory capacity and pupil diameter factors (keeping the correlations with fluid intelligence set to zero). Comparing this to the null model yielded a non-significant difference, $\Delta\ \chi^2(1) = 0.35$, $p = 0.55$, indicating the null model would be preferred as more parsimonious. Next, we estimated the weight of evidence in favor of the null model (i.e., Bayes Factor) using the BIC of each model (Wagenmakers, 2007). The null model was 11.55 times more likely than the model including the correlation. Next, we specified a third model which freed the correlation between the fluid intelligence and pupil diameter factors (keeping the correlations with working memory capacity set to zero). Compared to the null model, freeing the correlation between pupil diameter and fluid intelligence did not improve the model fit, $\Delta\ \chi^2(1) = 0.05$, $p = 0.82$, indicating the null model would be preferred as more parsimonious. The BIC comparison indicated that the null model was 13.44 times more likely, given the data. Therefore, there was rather strong evidence against a correlation between pupil diameter and fluid intelligence and

---

[2] CFI = comparative fit index, TLI = Tucker-Lewis Index, RMSEA = root mean squared error of approximation, SRMR = standardized root mean residual.

Table 3 Correlations among measures in Study 1

Measure	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Color *k*	--										
2. Orientation *k*	0.50	--									
3. Letter *k*	0.48	0.52	--								
4. Raven	0.29	0.32	0.29	--							
5. Number series	0.24	0.21	0.30	0.36	--						
6. Letter sets	0.17	0.28	0.22	0.21	0.43	--					
7. Pupil - lights off, black	0.04	0.01	-0.07	-0.09	-0.02	0.15	--				
8. Pupil – lights off, gray	-0.03	-0.01	-0.08	-0.09	-0.02	0.18	0.64	--			
9. Pupil – lights off, white	-0.05	0.09	-0.06	0.02	0.05	0.18	0.34	0.75	--		
10. Pupil – lights on, black	-0.02	-0.01	-0.02	-0.06	-0.13	0.16	0.54	0.66	0.59	--	
11. Pupil – lights on, gray	-0.06	-0.03	-0.06	-0.06	-0.11	0.14	0.46	0.74	0.68	0.82	--
12. Pupil – lights on, white	-0.01	0.03	-0.09	-0.04	-0.07	0.12	0.41	0.68	0.77	0.70	0.84

Note. k = capacity estimate

against a correlation between pupil diameter and working memory capacity.

Study 2

Study 2 was similar to Study 1, except it only included measures of fluid intelligence and baseline pupil diameter. However, we used the same lighting manipulations as in Study 1 to test for a correlation between pupil diameter and fluid intelligence, and whether this correlation would be moderated by external lighting conditions.

Method

Participants and procedure

After exclusions, the sample included 172 participants (age *M* = 19 years, age *SD* = 2.5, age range 18–38 years; 122 identified as female, 49 as male, one as non-binary/other gender; 50% identified as white, 45% as Hispanic or Latino, 24% as Black/African American, 19% as Asian, 5% as Native American, and 2% as Native Hawaiian/Pacific Islander). All participants were undergraduate students at the University of Texas at Arlington participating

Table 4 Results of linear mixed model with working memory capacity on pupil diameter in Study 1

Effect	Working memory capacity				Fluid intelligence			
	b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Intercept	6.51	0.06	116.07	<0.001	6.51	0.06	114.47	<0.001
Color (gray)	-2.28	0.05	-44.35	<0.001	-2.28	0.05	-43.93	<0.001
Color (white)	-3.05	0.05	-59.18	<0.001	-3.05	0.05	-58.54	<0.001
Lights (on)	-2.17	0.05	-42.33	<0.001	-2.17	0.05	-41.86	<0.001
Ability	0.00	0.07	-0.05	0.96	-0.02	0.07	-0.27	0.79
Color (grey) × lights	1.81	0.07	24.83	<0.001	1.81	0.07	24.53	<0.001
Color (white) × lights	2.14	0.07	29.30	<0.001	2.13	0.07	28.89	<0.001
Color (gray) × ability	-0.04	0.06	-0.69	0.49	0.02	0.07	0.25	0.80
Color (white) × ability	-0.01	0.06	-0.22	0.83	0.09	0.07	1.30	0.19
Lights (on) × ability	-0.01	0.06	-0.20	0.84	-0.06	0.07	-0.88	0.38
Color (gray) × lights (on) × ability	0.03	0.09	0.37	0.71	0.06	0.09	0.70	0.49
Color (white) × lights (on) × ability	0.04	0.09	0.45	0.66	0.02	0.09	0.24	0.81
ICC	0.58				0.59			
Number of observations	1,122				1,122			
Number of participants	190				190			

Note. b = regression coefficient. *SE* = standard error of regression estimate. For the model labeled Working Memory Capacity, the *z*-score composite of the working memory tasks was entered as the *ability* term. For the modeled labeled Fluid Intelligence, the *z*-score composite of the working memory tasks was entered as the *ability* term

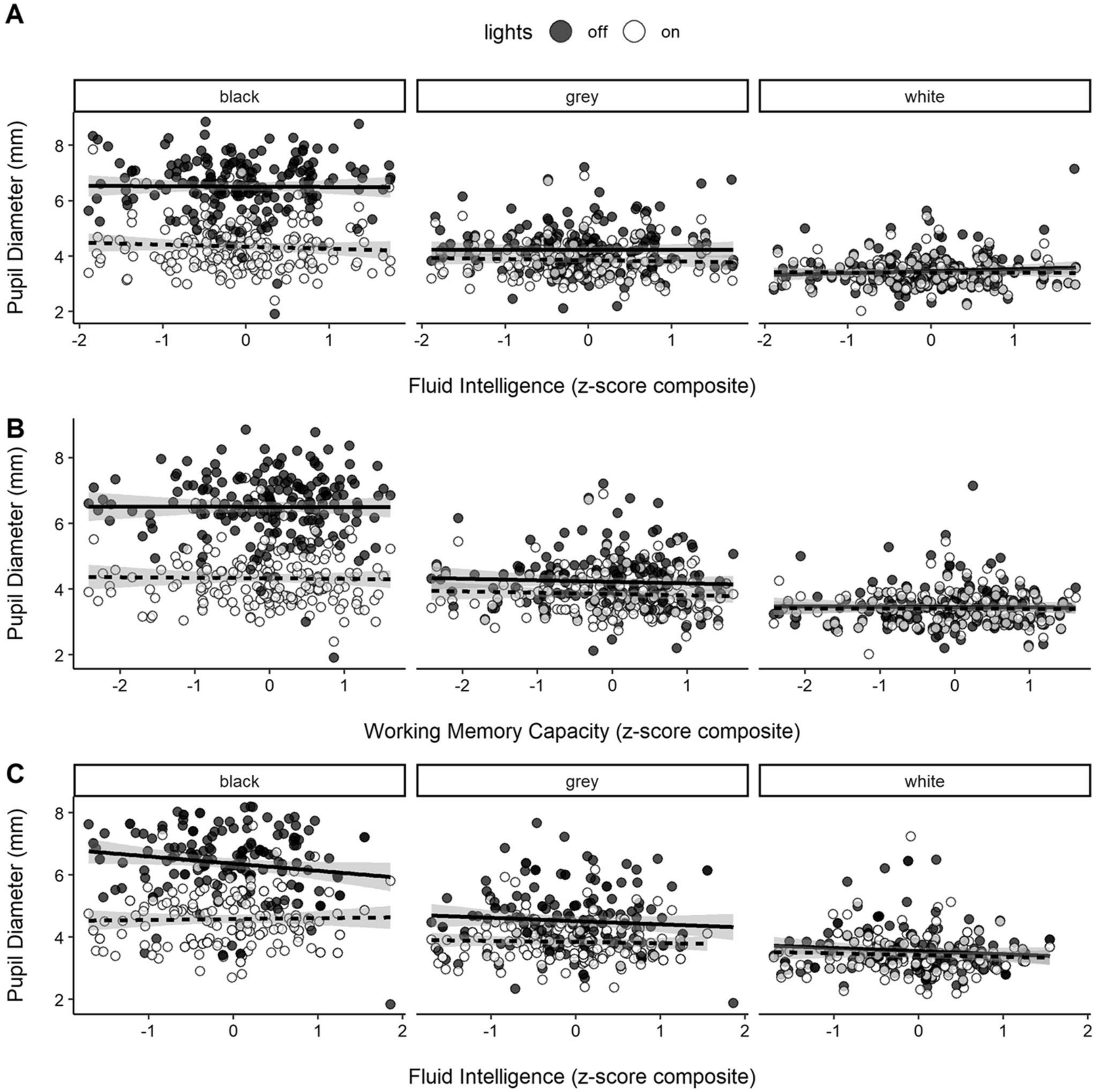

Fig. 1 Scatterplots of pupil diameter in the six lighting conditions by (**a**) fluid intelligence in Study 1, **b** working memory capacity in Study 1, and **c** fluid intelligence in Study 2

in exchange for partial course credit. Over the course of a 1-h session, participants completed three measures of fluid intelligence, a serial reaction time task, and a correlation estimation task. The serial reaction time and correlation estimation tasks are not analyzed here. Participants' pupils were measured at the beginning of the session. At the end of the session, participants were debriefed. The experimental protocol was approved by the Institutional Review Board of the University of Texas at Arlington.

Tasks

The Raven, number series, and letter sets tasks were identical to those used in Study 1.

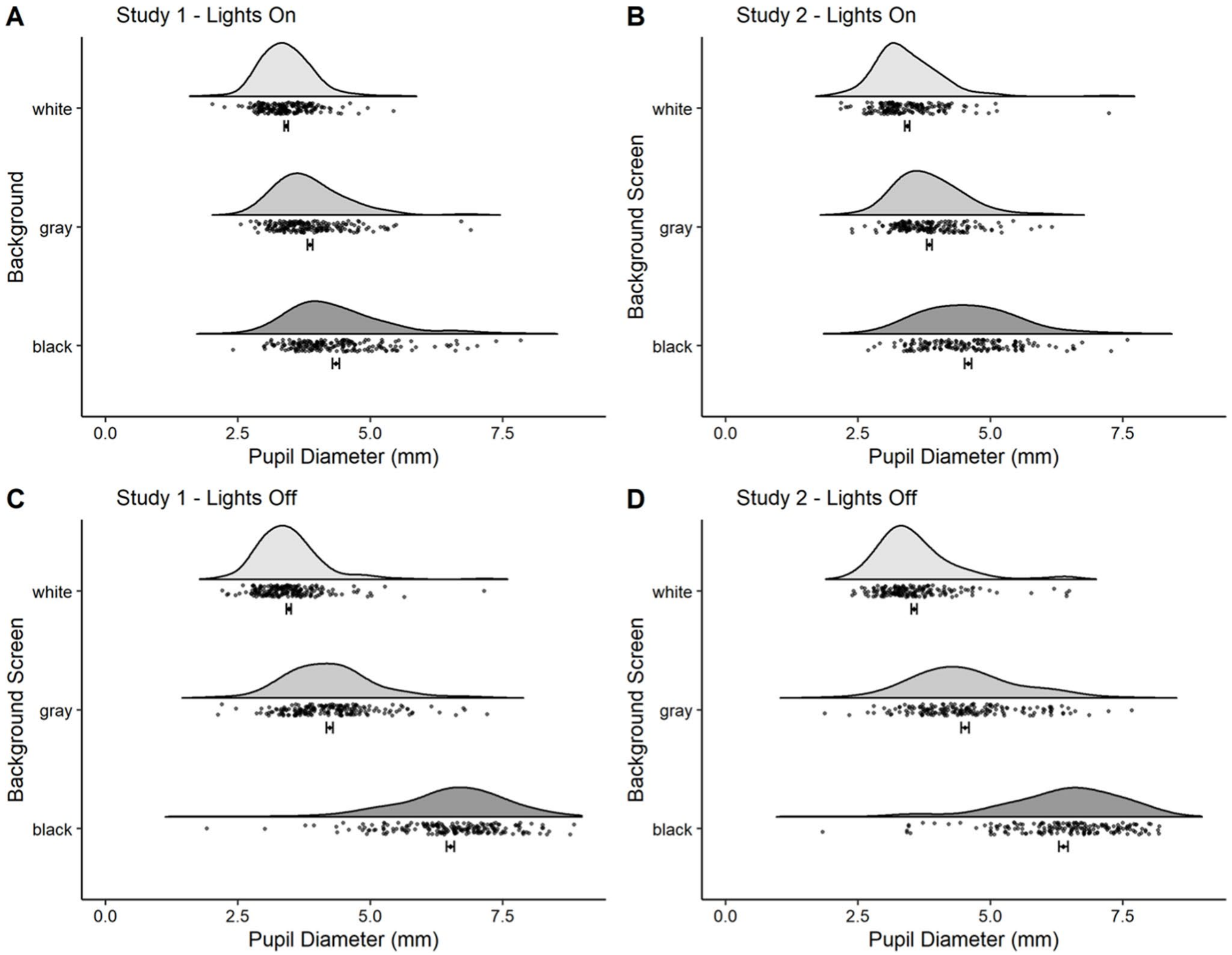

Fig. 2 Distributions of pupil diameter in six lighting conditions for each study. Individual data points below the density curves represent single subjects, and the point with error bars represent the mean +/- one standard error for that condition

Resting pupil measurement

The resting pupil measurement followed a nearly identical procedure to Study 1 with a few exceptions. First, the lights on/lights off order was reversed. Participants viewed the three background screens with lights on first, then lights off. Second, participants were in individual run rooms, rather than the group run room as in Study 1. Therefore, there was no contaminating light from neighboring stations or the experimenter station. Finally, the monitors were slightly smaller (1,600 × 900 px). Otherwise, the eye-tracking setup was the same.

Data analysis

The data analysis procedures were nearly identical to Study 1. All data and analysis code can be found on the Open Science Framework (https://osf.io/569bt/).

Exclusions The same exclusion criteria used in Study 1 were applied in Study 2.

Results

Descriptive statistics are listed in Table 5, and correlations among measures are listed in Table 6. Like Study 1, we first specified a linear mixed effects model with fixed effects for background color (black, gray, white) and lights (on, off) and a *z*-score composite of fluid intelligence (Table 7). The model yielded the same significant environmental effects on pupil diameter: a significant main effect of both the gray and white background (larger pupils compared to gray), and lights (larger pupils with lights off), and significant background color × lights interactions. The effect of room lighting was again larger when the background screen was black compared to when it was gray or white. Interestingly, the model yielded a significant

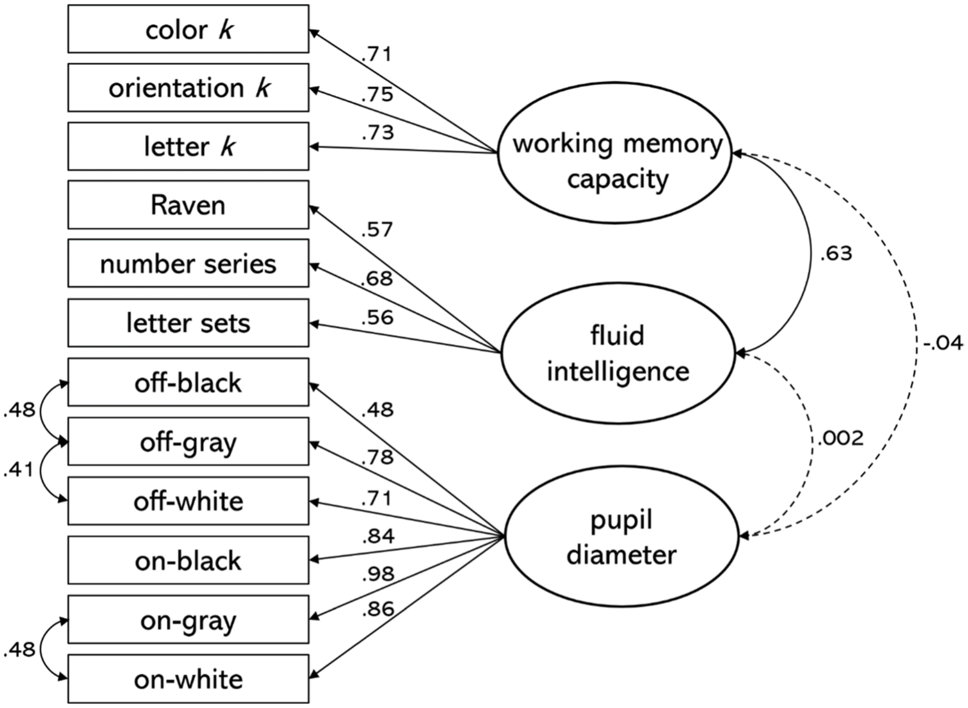

Fig. 3 Confirmatory factor analysis of working memory capacity, fluid intelligence, and resting pupil diameter in Study 1. Solid lines indicate significant paths at $p < .05$. Dashed lines indicate non-significant paths

negative main effect of fluid intelligence on pupil diameter, which is the opposite pattern that Tsukahara and Engle (2021b) showed. The main effect was qualified by a significant interaction with room lighting. In the lights-off conditions, the correlation tended to be more negative than the lights-on conditions (see Fig. 1c).

Next, we specified the same latent confirmatory factor analysis as in Study 1. Additionally, the number series task had a standardized loading above 1, so we specified this loading to equal 1. The final model, depicted in Fig. 4, fit the data well, $\chi^2(24) = 56.23$, CFI = .96, TLI = .93, RMSEA = .09 90% CI [.06, .12], SRMR = .08. The latent correlation between fluid intelligence and pupil diameter was non-significant.

Finally, we performed the same model comparison using χ^2 and BIC. A null model with the correlation between the fluid intelligence and pupil diameter factors set to zero did not fit any worse than a model with a correlation, $\Delta\ \chi^2(1) = 2.12$, $p = 0.14$, and thus would be preferred. The BIC comparison indicated that the null model was 4.54 times more likely. Therefore, there was evidence against the presence of a correlation between fluid intelligence and resting pupil diameter.

Table 5 Descriptive statistics for all measures in Study 2

Measure	*N*	Mean	SD	Skew	Kurtosis	Reliability
Raven	170	8.15	3.22	-0.40	-0.80	0.77
Number series	172	6.72	2.71	0.28	-0.11	0.84
Letter sets	146	7.53	2.89	0.55	0.53	0.80
Pupil - lights off, black	172	6.38	1.09	-1.01	1.62	0.97
Pupil – lights off, gray	168	4.52	0.98	0.48	0.29	>0.99
Pupil – lights off, white	162	3.56	0.72	1.76	4.43	0.99
Pupil – lights on, black	168	4.58	0.87	0.56	0.44	>0.99
Pupil – lights on, gray	169	3.85	0.63	0.80	1.23	>0.99
Pupil – lights on, white	167	3.43	0.62	1.71	7.61	0.99

Note. SD = standard deviation

Table 6 Correlations among measures in Study 2

Measure	1.	2.	3.	4.	5.	6.	7.	8.
1. Raven	--							
2. Number series	0.43	--						
3. Letter sets	0.14	0.40	--					
4. Pupil - lights off, black	-0.16	-0.14	0.01	--				
5. Pupil – lights off, gray	-0.19	-0.13	0.18	0.61	--			
6. Pupil – lights off, white	-0.19	-0.13	0.14	0.22	0.62	--		
7. Pupil – lights on, black	-0.04	0.00	0.05	0.47	0.58	0.52	--	
8. Pupil – lights on, gray	-0.14	-0.04	0.08	0.51	0.62	0.65	0.85	--
9. Pupil – lights on, white	-0.17	-0.07	0.12	0.36	0.50	0.66	0.55	0.77

Table 7 Results of linear mixed model with fluid intelligence on pupil diameter in Study 2

Effect	*b*	*SE*	*t*	*p*
Intercept	6.37	0.07	93.29	<0.001
Color (gray)	-1.85	0.06	-29.39	<0.001
Color (white)	-2.82	0.06	-44.27	<0.001
Lights (on)	-1.78	0.06	-28.20	<0.001
Fluid intelligence	-0.20	0.09	-2.14	0.033
Color (grey) × lights	1.10	0.09	12.32	<0.001
Color (white) × lights	1.66	0.09	18.41	<0.001
Color (gray) × fluid intelligence	0.13	0.09	1.52	0.129
Color (white) × fluid intelligence	0.11	0.09	1.28	0.202
Lights (on) × fluid intelligence	0.26	0.09	3.00	0.003
Color (gray) × lights (on) × fluid intelligence	-0.23	0.12	-1.83	0.068
Color (white) × lights (on) × fluid intelligence	-0.23	0.12	-1.86	0.063
ICC = 0.53				
Number of observations = 1,006				
Number of participants = 141				

Note. *b* = regression coefficient. *SE* = standard error of regression estimate. ICC = 0.53, Number of observations = 1,006, number of participants = 141

Combined analysis

Data were combined across studies to estimate the latent correlation between fluid intelligence and pupil diameter with the largest possible sample. Because there were slight differences in the lighting conditions across studies, and because we reversed the order of lights on/off in Studies 1 and 2, we compared pupil diameters in the 6 conditions across the studies with a 2 (lights on vs. lights off) × 3 (screen background: black vs. gray vs. white) × 2 (Study 1 vs. Study 2) mixed ANOVA. This ANOVA yielded a significant Study × Color × Lights interaction ($F(2, 266) = 25.44$, $p < 0.001$, partial $\eta^2 = 0.07$. We subsequently compared all six conditions with Study as the independent variable with pairwise *t*-tests. No comparisons were significant with an α adjusted to 0.008 (0.05/6).[3] The model, depicted in Fig. 5, fit the data well, $\chi^2(23) = 84.22$, CFI = .96, TLI = .94, RMSEA = .09 90% CI [.07, .11], SRMR = .06. The latent correlation between fluid intelligence and pupil diameter was small and non-significant. Finally, we compared the model to a null model in which the correlation between fluid intelligence and pupil diameter was set to equal zero. The model comparison was not significant, $\Delta\ \chi^2(1) = 1.34$, $p = 0.25$, indicating the null model would be preferred, and the BIC comparison indicated that the null model was 9.75 times more likely.[4] So, although there was a negative main effect of fluid intelligence ($b = -.20$) on pupil diameter, which was the opposite of the hypothesized direction, the latent variable analysis indicated there was considerable evidence against a correlation between fluid intelligence and pupil diameter.

---

[3] See Online Supplemental Materials for full ANOVA and a plot of pupil diameter by condition and study.

[4] We also estimated this model with pupil diameters standardized by study because of the lighting differences and ordering of lights on vs. lights off. The results did not change. The correlation between the fluid intelligence and pupil factors was $r = -0.06$, $p = 0.29$) and the BIC comparisons favored the null model by a factor of 11.03.

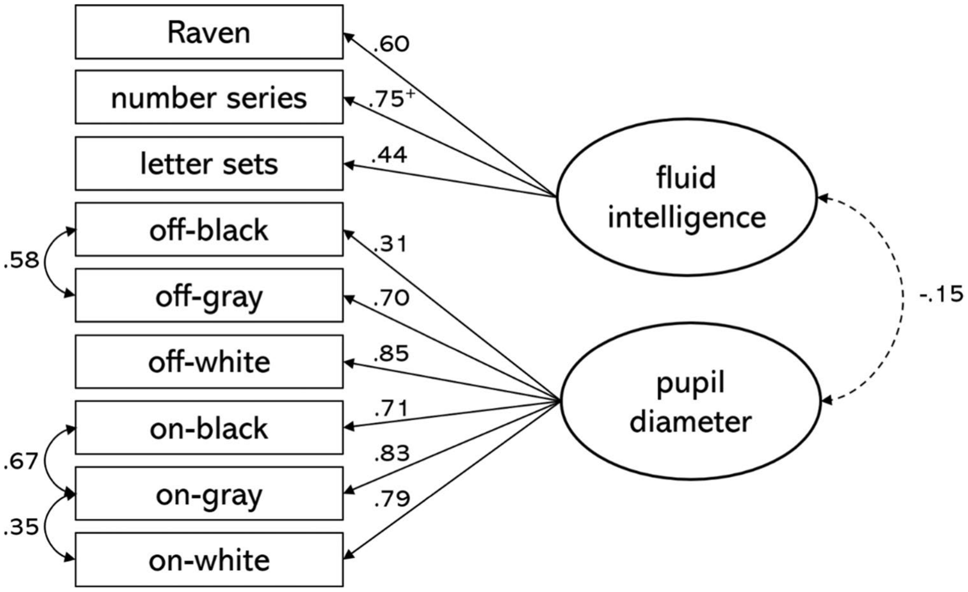

Fig. 4 Confirmatory factor analysis of fluid intelligence and resting pupil diameter in Study 2. Solid lines indicate significant paths at $p < .05$. Dashed lines indicate non-significant paths. $^+$ Factor loading was set to equal 1

General discussion

The current studies examined whether there is a correlation between pupil diameter and cognitive abilities, namely working memory capacity and fluid intelligence. Although two studies have observed rather robust correlations between resting pupil diameter and cognitive ability (Tsukahara et al., 2016; Tsukahara & Engle, 2021b), this finding has yet to be independently replicated. Here, we attempted a close replication of Tsukahara and Engle (2021b), and we found substantial evidence against a correlation between pupil diameter and working memory and between pupil diameter and fluid intelligence.

There has been some debate as to what might be leading Tsukahara and colleagues to find this association and while others do not (see Unsworth et al., 2021, and Tsukahara et al., 2021, for recent commentary). In their most recent report, Tsukahara et al. make several recommendations,

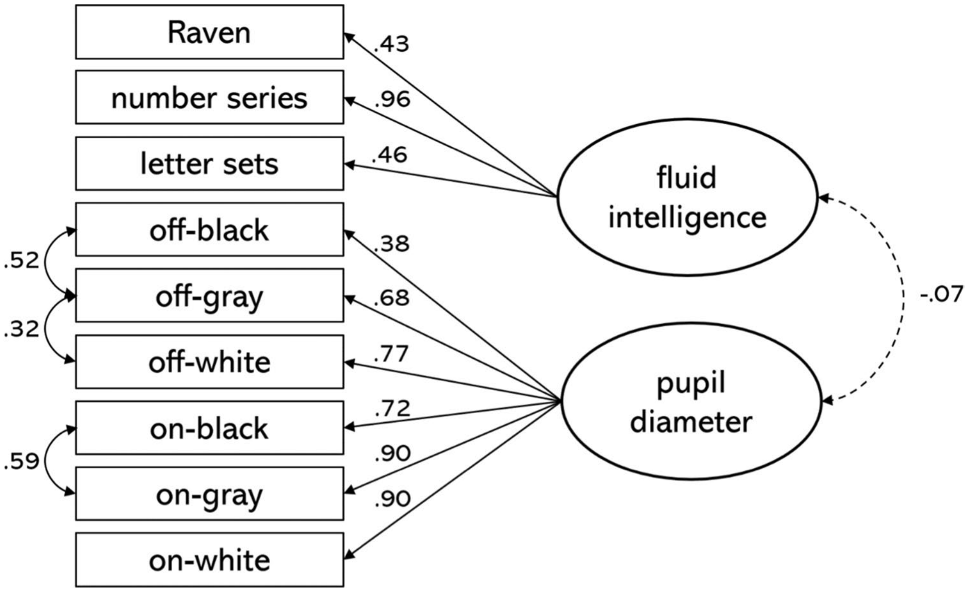

Fig. 5 Confirmatory factor analysis of fluid intelligence and resting pupil diameter combining data across studies. Solid lines indicate significant paths at $p < .05$. Dashed lines indicate non-significant paths

Table 8 Study comparisons

Measure	Tsukahara et al. (2016)	Tsukahara and Engle (2021b)	Current study
Raven	9.9 (3.3)	10.4 (3.4)	8.27 (3.36)
Number series	9.6 (3.2)	9.9 (2.9)	8.14 (3.07)
Lights off - white	3.64 (0.56)	3.07 (0.34)	3.51 (0.67)
Lights off - gray	5.0 (0.82)	--	4.37 (0.91)
Lights off - black	--	5.68 (0.89)	6.45 (1.06)
Lights on - white	--	3.18 (0.35)	3.42 (0.55)
Lights on - gray	--	--	3.85 (0.65)
Lights on - black	--	4.95 (0.80)	4.46 (0.89)

Note. Means are listed with standard deviations in parentheses

including recruiting a diverse sample, using multiple tasks to measure a cognitive construct, ensuring high amounts of interindividual variability in both cognitive scores and pupil diameter, reporting lighting conditions, and asking about potential confounds like caffeine, sleep, and age. We followed these recommendations with one exception: our sample was entirely undergraduate college students. However, we do not believe this sampling difference prevented us from observing the hypothesized correlations. The concern with university participant pools is that they are largely young, educated, high-ability, racially White, and from high socioeconomic backgrounds (Henrich et al., 2010). Most importantly, the homogeneity may create a range restriction issue.

To test whether our sample was range restricted, especially regarding cognitive ability, we compared our distribution of scores on the Raven and number series tests to the distributions observed by Tsukahara et al. in their combined university/community samples.[5] On average, Raven and number series scores in our sample were lower than those observed by Tsukahara et al. with just as much variability in the distribution (see Table 8). Further, the current studies' ranges of pupil diameter had as much variability in the sample as the studies reported by Tsukahara et al. Finally, our sample was racially and ethnically diverse, with 82% of participants identifying with at least one non-White racial/ethnic group. Therefore, despite criticisms of university samples as homogenous and narrow in range, we believe our sample achieved adequate diversity despite comprising exclusively university students. Additionally, Robison, Coyne, et al. (2022a); recruited from a large and diverse sample of US military members, some of whom had college degrees but many of whom did not. Finally, Coors et al. (2022) included from a large and broad sample of adults ranging in age from 30 to 95 years, and they did not find moderation of the pupil-cognition relations by age strata. Therefore, we do not believe replication failures are due to sampling weaknesses.

Conclusion

Overall, we believe there is sufficient evidence to conclude there is not a correlation between resting pupil diameter and cognitive ability. This conclusion is not without theoretical implications. The crux of Tsukahara and Engle's (2021a) theory that the LC-NE system underlies individual differences in goal-directed cognition rests on the presence of such a correlation. We argue that resting pupil diameter cannot not be used as evidence for a role of the LC-NE system in working memory, fluid intelligence, or other related abilities like attention control. That is not to say that the LC-NE system is not important for cognition. In fact, we believe the LC-NE system plays an integral role in goal-directed cognitive activity (Unsworth & Robison, 2017a). However, we would encourage more emphasis on pupillary dynamics, rather than resting state measures. Recently, we and others have presented evidence that pupillary dynamics, measured while people are performing a variety of cognitive tasks, predict a host of cognitive abilities including attention control, working memory capacity, long-term memory, and fluid intelligence (Unsworth & Robison, 2015, 2017a, 2017b, 2018; Madore et al., 2020; Robison & Brewer, 2022; Robison & Unsworth, 2019; Robison, Coyne, et al., 2022a; Robison, Trost, et al., 2022b). Therefore, we would urge future work to focus more on pupillary dynamics, rather than resting state measures, as correlates of cognitive ability.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13423-023-02273-7.

Acknowledgements The authors would like to thank Shaan Jani, Matthew Nguyen, Timber Gustafson, Janine Shuman, Lauren Barrientos, Britney Bright, Aarti Darji, Angie Gonzalez, Jessica Nwanko, and Theresa Ho for their assistance with data collection.

Funding The authors were funded by a cooperative agreement with the U.S. Naval Research Laboratory (Award No. N00173-22-2C006).

References

Alnæs, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of Vision, 14*(4), 1.

Coors, A., Breteler, M. M., & Ettinger, U. (2022). Processing speed, but not working memory or global cognition, is associated with pupil diameter during fixation. *Psychophysiology, 59*(11), e14089.

---

[5] Although the letter sets tasks used in each study were the same, Tsukahara et al. used an extended version with more items and a longer deadline, and therefore means and distributions could not be compared.

Dowle, M., & Srinivasan, A. (2020). data.table: Extension of data.frame. https://CRAN.R-project.org/package=data.table. Accessed 07 Jul 2022.

Ekstrom, R. B., & Harman, H. H. (1976). *Manual for kit of factor-referenced cognitive tests, 1976*. Educational Testing Service.

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*(1), 119–129. https://doi.org/10.1111/j.1469-8986.2007.00605.x

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature, 466*(7302), 29–29.

Joshi, S., & Gold, J. I. (2020). Pupil size as a window on neural substrates of cognition. *Trends in Cognitive Sciences, 24*(6), 466–480.

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron, 89*(1), 221–234.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26.

Learn, J. R. (2021, August 21). Pupil size may be linked to intelligence. *Discover Magazine*. Retrieved from https://www.discovermagazine.com/mind/pupil-size-may-be-linked-to-intelligence. Accessed 07 Jul 2022.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281.

Madore, K. P., Khazenzon, A. M., Backes, C. W., Jiang, J., Uncapher, M. R., Norcia, A. M., & Wagner, A. D. (2020). Memory failure predicted by attention lapsing and media multitasking. *Nature, 587*(7832), 87–91.

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping, 35*(8), 4140–4154.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/. Accessed 07 Jul 2022.

Raven, J. C., Raven, J. C., & Court, J. H. (1962). *Advanced progressive matrices*. HK Lewis.

Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications, 7*(1), 1–7.

Revelle, W. (2022) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA. https://CRAN.Rproject.org/package=psychVersion=2.2.5

Robison, M. K., & Brewer, G. A. (2020). Individual differences in working memory capacity and the regulation of arousal. *Attention, Perception, & Psychophysics, 82*(7), 3273–3290.

Robison, M. K., & Brewer, G. A. (2022). Individual differences in working memory capacity, attention control, fluid intelligence, and pupillary measures of arousal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001125

Robison, M. K., & Unsworth, N. (2019). Pupillometry tracks fluctuations in working memory performance. *Attention, Perception, & Psychophysics, 81*(2), 407–419.

Robison, M. K., Coyne, J. T., Sibley, C., Brown, N. L., Neilson, B., & Foroughi, C. (2022a). An examination of relations between baseline pupil measures and cognitive abilities. *Psychophysiology*. https://doi.org/10.1111/psyp.14124

Robison, M. K., Trost, J. M., Schor, D., Gibson, B. S., & Healey, M. K. (2022b). Pupillary correlates of individual differences in long-term memory. *Psychonomic Bulletin & Review*, 1–12. https://doi.org/10.3758/s13423-022-02081-5

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36.

Thurstone, L. L. (1938). *Primary mental abilities*. University of Chicago Press.

Tsukahara, J. S., & Engle, R. W. (2021a). Fluid intelligence and the locus coeruleus–norepinephrine system. *Proceedings of the National Academy of Sciences, 118*(46), e2110630118.

Tsukahara, J. S., & Engle, R. W. (2021b). Is baseline pupil size related to cognitive ability? Yes (under proper lighting conditions). *Cognition, 211*, 104643.

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

Tsukahara, J. S., Burgoyne, A. P., & Engle, R. W. (2021, June 2). Pupil size is a marker of intelligence. https://www.scientificamerican.com/article/pupil-size-is-a-marker-of-intelligence/. Accessed 07 Jul 2022.

Tsukahara, J. S., Draheim, C., & Engle, R. W. (2021). Baseline pupil size is related to fluid intelligence: A reply to Unsworth. *Cognition, 215*, 104826. https://doi.org/10.1016/j.cognition.2021.104826

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*(3), 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*(4), 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, and Cognition, 43*(12), 1962–1987.

Unsworth, N., & Robison, M. K. (2018). Tracking working memory maintenance with pupillometry. *Attention, Perception, & Psychophysics, 80*(2), 461–484.

Unsworth, N., Miller, A. L., & Robison, M. K. (2021). No consistent correlation between baseline pupil diameter and cognitive abilities after controlling for confounds-A comment on. *Cognition, 215*, 104825. https://doi.org/10.1016/j.cognition.2021.104825

Unsworth, N., Miller, A. L., & Robison, M. K. (2021). Is working memory capacity related to baseline pupil diameter? *Psychonomic Bulletin & Review, 28*(1), 228–237.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14*(5), 779–804.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pederson, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), 1686.

Wilke, C. (2020). cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.1.1. https://CRAN.R-project.org/package=cowplot. Accessed 07 Jul 2022

Open practices statement The data and analysis code are available via the Open Science Framework (https://osf.io/569bt/). Neither study was preregistered.